Durham E-Theses

Development of coarse-grained models for the simulation of soft matter systems

THOMAS DAVID POTTER

Development of coarse-grained models for the simulation of soft matter systems

Thomas Potter

A Thesis presented for the degree of

Doctor of Philosophy

Department of Chemistry
Durham University
United Kingdom

February 2019

Abstract

This thesis aims to examine the parametrisation of coarse-grained models for the simulation of soft matter systems. The strengths and weaknesses of a range of methods are examined, and suggestions for improvements are made.

Initially, two *bottom-up* methods, iterative Boltzmann inversion (IBI) and hybrid force matching (HFM) are applied to a liquid octane/benzene mixture and compared to a *top-down* model based on a version of statistical associating fluid theory, the SAFT-γ Mie equation of state. These models are tested for their ability to represent the structure and thermodynamics of the underlying atomistic system, as well as their transferability between temperatures and concentrations. Attempts are then made to address the poor transferability of the *bottom-up* models using a variant of IBI, multi-state IBI (MS-IBI). MS-IBI allows concentration transferable potentials to be generated but is not successful in improving temperature transferability. The state-point dependence of pair potentials is identified as the cause of poor temperature transferability, and initial attempts to address this are discussed.

A range of coarse-grained models of the non-ionic liquid crystal TP6EO2M is examined. HFM is able to give a structurally accurate coarse-grained model; however, the difficulty of sampling all relevant configurations within an atomistic reference system appear to cause problems with calculating accurate association free energies. The new MARTINI 3 *top-down* force field is shown to improve upon the structural and thermodynamic properties of MARTINI 2, allowing larger system sizes to be studied. The nematic and hexagonal columnar chromonic phases are observed, and the concentration dependence seen in the experimental phase diagram is reproduced. This represents the first simulations of chromonic liquid crystal phases using systematic coarse graining.

Declaration

The work presented in this thesis has not been submitted for another degree at Durham University or any other institution. Some parts of the work were carried out collaboratively; this has been highlighted at the relevant points in the text.

The copyright of this thesis rests with the author. No quotation from it should be published without the author's prior written consent and information derived from it should be acknowledged.

Acknowledgements

I would like to start by thanking my supervisor, Prof. Mark Wilson. Mark has always been a great source of scientific knowledge, support and encouragement, and it has been a pleasure to work for him.

I would also like to thank everyone from the Wilson group and the other computational chemistry groups for creating a fantastic, and unique, working environment. Tea breaks, pub trips and film nights have all been highlights of my time here, despite occasional disagreements about my choice of bowl. Particular thanks go to Martin Walker for all the help and advice he has provided, and Jos Tasche for doing all of the SAFT calculations.

I would like to thank Durham's Chemistry Department and HPC Facility for providing me with the resources I needed to complete my research. The Hamilton cluster in particular has been essential to carrying out all of the calculations in this thesis.

Finally, I would like to thank my friends and family, and especially my parents, who have been a constant source of support during my PhD.

Contents

Chapter 1

Introduction to coarse-grained modelling

1.1 Levels of computational chemistry

Molecular modelling is a valuable scientific tool for studying chemical systems. Increasingly, it is being used to complement experiment, due to its ability to look at systems from a molecular perspective. However, in computational chemistry, there must always be a balance between the accuracy of the method used and its computational cost. There are many computational chemistry methods available, some of which are able to treat a small system with very high accuracy, others of which can be used on very large systems by applying appropriate levels of approximation.[1] The range of available method types is summarised in Figure 1.1

The study of individual small molecules is often carried out using methods based on quantum mechanics, such as density functional theory,[2] or more accurate (and expensive) *ab initio* methods.[3] These allow the study of the electronic structure of small molecules and, depending on the method used, can allow very accurate calculations of molecular structures and properties. Such methods can also be used to study the dynamics of small systems over very short timescales (usually no more than 1 ns), for example using Car-Parrinello molecular dynamics.[4]

For larger systems, such as polymers or bulk quantities of small molecules, quantum methods are prohibitively expensive, and atomistic methods based entirely on classical mechanics must be employed. These methods treat molecules as a series of spheres (atoms) connected by springs (bonds), with no explicit treatment of the
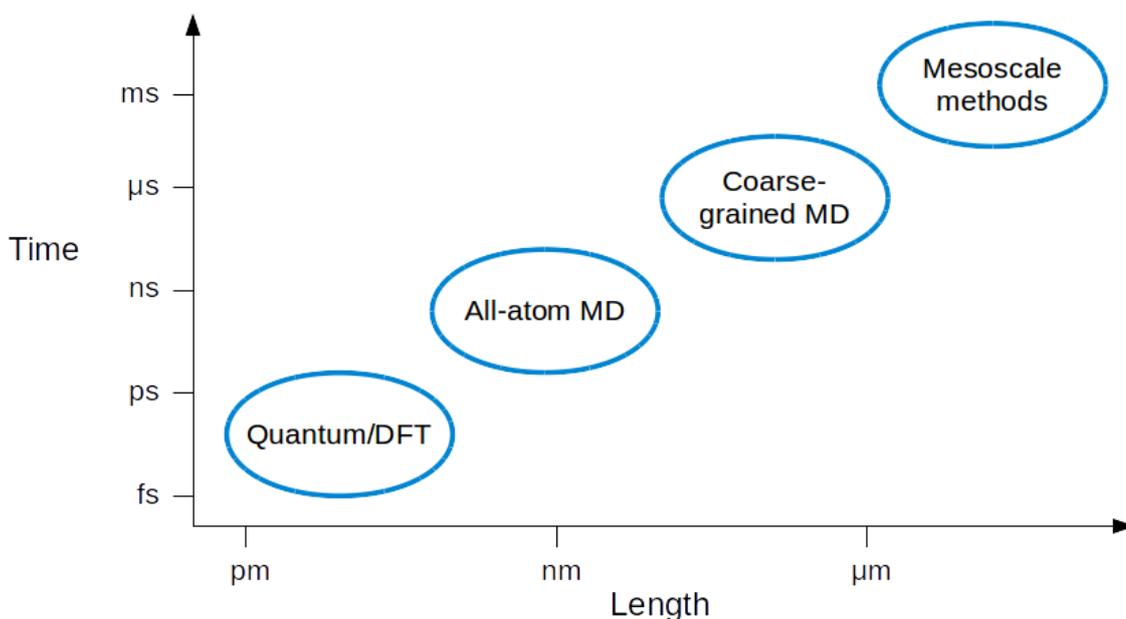
Figure 1.1: General levels at which molecular modelling can be carried out, as a function of the accessible time and length scales at each level.

electrons in the system. They can allow the study of dynamics on the ns - μs scale, and as such are extremely useful for studying the behaviour of larger systems. Force fields are parametrised either from *ab initio* methods or from experimental thermodynamic data. Atomistic force fields have been used to obtain a large amount of useful data, particularly for biomolecular systems where *ab initio* methods are prohibitively expensive in terms of computer time.[5,6]

At the very large end of the spectrum are mesoscale methods; these range from particle-pased methods such as dissipative particle dynamics,[7] to grid-based Lattice Boltzmann simulations,[8] which do not explicitly consider individual particles. These methods can scale to very large systems, up to millimetres or even centimetres on second timescales. However, in going to these scales, they lose the ability to model the underlying chemistry of a system in a detailed way.

There are some processes, like protein dynamics, self-assembly and phase changes, which occur on μs or ms time scales but are driven by chemical interactions. These processes require a middle ground between classical all-atom simulations and mesoscale methods. This has been the driving force for the recent development of coarse-grained molecular dynamics, where the accuracy of atomistic models is retained as far as possible, but at smaller computational cost.[9]

The basic idea behind coarse graining is the removal of unimportant degrees of freedom from atomistic models, by grouping together several atoms as a single

interaction site, often called a coarse-grained bead. This comes with the obvious benefit of reducing the number of force calculations needed at each time step, but there are further benefits. Removing degrees of freedom results in a simpler potential energy surface. There will therefore be fewer local minima for the system to get stuck in, which means that equilibration of a coarse-grained system will be faster than an atomistic one. Finally, many coarse-grained models use softer interaction potentials than atomistic models, allowing for longer time steps. The result of all these factors is that coarse graining facilitates the study of much larger systems over longer time scales than atomistic models.[9]

A large amount of research has been carried out into combining these different scales, so that the area of particular interest can be modelled in detail, and the surroundings at a coarser level. QM/MM, where part of a system is treated at a quantum level and the rest with all-atom classical forcefields, is particularly popular.[10] However, this idea has also been extended to a coarse-grained level. For example, the AdResS method allows the combination of a detailed classical all-atom region with coarse-grained surrounding, by means of an intermediate hybrid region.[11] There have even been examples of a hybrid QM/CGMM system, which would combine a detailed quantum mechanical region with surroundings modelled at the coarse-grained level.[12] The idea of backmapping, in which coarse-grained simulations are used to rapidly equilibrate a system, after which it is converted back to an atomistic representation for accurate data collection, has also been used.[13]

Whether a system is being modelled at a fully coarse-grained level or using a hybrid method, it is important to make the sure the coarse-grained model adequately represents the system. The two main steps to constructing a coarse-grained model are choosing a mapping scheme, which defines how the coarse-grained beads are related to the underlying atomistic structure, and parametrising the interaction potentials. Both of these steps are the subject of significant research efforts.

1.2 Coarse-grained mapping

The choice of coarse-grained mapping is the first decision which must be made when coarse graining a system, and is one of the most important features of the model. Although the removal of degrees of freedom can significantly speed up simulations, it is important that sufficient chemical accuracy is retained. Lower level

degrees of freedom may be crucial, particularly to the behaviour of more complex molecules and polymers. Therefore, the coarse-grained mapping scheme used must be based on a balance between cost and accuracy, which requires detailed knowledge of the system being studied.

There are a number of approaches to selecting a mapping scheme. Several studies have been published detailing automated methods which ensure that the important degrees of freedom are not neglected, and that the coarse-grained model retains the important structural features of the atomistic system.[14,15] However, in most cases, mapping is done by hand, using a knowledge of which groups are chemically important. Regardless of the approach used, the level of coarse graining must first be chosen.

On the boundary between all-atom and coarse grained models are united atom (UA) schemes, where a united atom is defined as 1 heavy atom with all of its associated hydrogens. Since, for most organic molecules, hydrogen is the most abundant element, this removes a significant number of degrees of freedom, and makes calculations more efficient. Prominent examples of force fields which use united atoms are the OPLS-UA[16], Amber United Atom[17] and TraPPE-UA[18] force fields; these were parametrised from a combination of *ab initio* and experimental data. These models compare very well to all-atom models; however the speed-up is relatively modest.

Some studies have used significantly larger coarse grained beads. For example, Dama *et al.* recently proposed a method for constructing ultra coarse grained models, where coarse grained beads could represent tens or even hundreds of heavy atoms.[19] At this resolution, common methods for parametrising coarse grained interactions, such as multiscale coarse graining (MS-CG)[20] and iterative Boltzmann inversion (IBI)[21], are no longer effective, and internal degrees of freedoms must be introduced into the coarse grained beads. The Integral Equation coarse graining method works well on very large beads, which can be up to the size of an entire polymer chain.[22]

In most of the coarse grained models discussed in this thesis, a mapping of 2 - 4 heavy atoms to every coarse grained bead is used. These are based on a number of factors including the separation of the molecules into distinct functional groups, retaining the cyclic nature of molecules like benzene or cycloalkanes, as well as the desire to keep a consistent level of mapping within a system as much as possible. These mappings are able to give a significant speed-up compared to

all-atom models, and work well for simple molecules and polymers. The molecules retain enough internal degrees of freedom to exhibit realistic behaviour; for example, the flexibility of longer chains can still be modelled well. For systems, such as many liquid crystals or lipids, where the molecules cannot be easily split into equally sized beads, a variable level of coarse graining may be used, where different beads include a different number of heavy atoms.[23–25] Alternatively, when using a coarse grained force field with predefined mappings, such as MARTINI (see Section 2.4.1), individual atoms may be missed out to retain a consistent mapping scheme.[26]

1.3   Systematic coarse-graining

There are a number of goals of coarse graining, which influences the type of coarse-grained model which is most appropriate. Often, the objective is gain qualitative information about the properties of a particular type of soft matter system. In these cases, coarse-grained models which capture the shape and relative interaction strengths of the molecules in question are often sufficient. For example, simple shape-based models have frequently been used to model the phase behaviour of liquid crystals.[27] DPD simulations have also been used to efficiently investigate the effects of changing molecular geometries or relative interaction strengths.[28]

Such studies are, however, limited in their ability to represent real systems. In some cases the aim is to obtain more quantitative information about a specific chemical system or systems, but at length and timescales which are not accesible to atomistic models. Systematic coarse graining refers to methods used to include real chemical data in the parametrisation of coarse-grained models, allowing for more accurate and in-depth studies of the effect of chemistry on the properties of soft matter systems.

Parametrising coarse-grained models which accurately represent the chemistry of a system is not straightforward. Several methods for systematically constructing CG potentials have been developed, each with their own advantages and disadvantages, but as yet no single method has been found to consistently produce good results for all systems. The introduction of software packages, such as VOTCA[29,30]and BOCS[31], has allowed CG models to be produced more easily, using some of the more common coarse-graining methods.

Broadly speaking, there are two approaches to systematic coarse-graining. The

first is the *top-down* approach, where a coarse-grained force field is parametrised to reproduce experimental thermodynamic data, in a similar manner to many atomistic force fields. The other is the *bottom-up* approach, where the coarse-grained potential is constructed from an accurate atomistic reference system. Many *bottom-up* methods have been developed; some involve parametrising a potential so that it reproduces certain structural or thermodynamic features of the reference system, while others involve directly calculating interactions from the reference.[9]

Currently, the world of coarse-grained modelling is dominated by models based on the MARTINI forcefield, a *top-down* model originally parametrised for lipid bilayers,[32] but which has been extended to a range of other systems.[26] At the time of writing, the original MARTINI paper had been cited over 2000 times. This popularity can be attributed to MARTINI's ease of use and transferability to a wide range of systems. The forcefield is based on Lennard-Jones potentials, and as such is easily compatible with major molecular dynamics simulation packages such as Gromacs.

While MARTINI is the most widely used coarse grained model, many other methods have been proposed. On the *top-down* side, the SAFT-γ Mie equation of state, a variation of statistical associating fluid theory in which the free energy of a system is written in terms of the parameters of a Mie potential, has been used to develop coarse-grained models.[33] There have been other publications presenting bespoke *top-down* coarse-grained models, using a range of different optimisation algorithms and matching to various thermodynamic properties.[34,35]

When large amounts of experimental data are available, it is possible to construct coarse-grained models which very accurately represent a system over a range of conditions. However, sufficient experimental data may not be available for all systems; in these cases, *bottom-up* models parametrised from atomistic simulation data may be the better option. Two of the earliest, and most widely used, *bottom-up* coarse graining methods are iterative Boltzmann inversion (IBI),[21] based on matching the structure of an atomistic reference, and the multi-scale coarse graining (MS-CG) method,[20] also known as force matching, which focusses on matching the forces in the atomistic reference. The relative entropy method has also been growing in popularity in recent years,[36] and there have been a huge number of other methods presented in addition to these three.

A detailed overview of the methods used in systematic coarse-graining is provided in Sections 2.4–2.6.

1.4   Issues with coarse-graining

Regardless of the method used to parametrise a coarse-grained model, there are two major issues which arise: those of representability and transferability. The former is concerned with the ability of a coarse-grained model to represent physical properties at the thermodynamic state point at which it is parametrised; the latter is its ability to be predictive at different state points, where parametrisation data is not available. Both issues are problematic and exist whether or not a coarse-grained model was developed by *top-down* or *bottom-up* methods.

The problem of representability arises from the fact that a coarse-grained model parametrised to reproduce one observable will not necessarily be able to reproduce other observables of the system at the same state-point.[37,38] Representability is discussed in depth by Wagner *et al.*[39] While it is of course possible to improve the accuracy of a coarse-grained model through the parametrisation of the interactions, the ability of the model to represent a system well is limited by the degree of coarse graining. It is only reasonable for a coarse-grained model to represent observables which are compatible with the resolution of the model, and do not significantly depend on the degrees of freedom which were removed from the all-atom representation. In some cases, the expression for calculating a particular observable in an all-atom representation may not be valid for a coarse-grained representation (see Section 2.2.5). The fact that most coarse-grained potentials are state-dependent means that the calculation of many thermodynamic properties is affected; for example, the standard virial expression for the pressure is invalid if the coarse-grained potentials are volume-dependent. Some representability issues are subtle. For example, coarse-graining through the elimination of degrees of freedom typically changes the balance of enthalpic and entropic contributions to free energy within a model of a molecular system.[40,41] However, the coarse-grained model might still capture the correct phase behaviour, provided free energy changes are well-represented.

Representability and transferability both have similar origins; the state-dependence of effective pair potentials also naturally affects the ability of coarse-grained models to work at multiple state points.[42,43]It is noticeable that transferability differs significantly between different types of coarse-grained models and different types of coarse-grained systems. Typically, *top-down* models parametrised from experimental data measured over a range of conditions have better transferability than

bottom-up models, which are parametrised using reference systems at a much more limited range of conditions (often only one state point). In the early days of systematic coarse-graining from reference atomistic models, it was noted that approaches such as iterative Boltzmann inversion (IBI)[44] could be used to provide coarse-grained models of homopolymers that were (sometimes) transferable across a range of temperatures and molecular weights.[45–47] However, it is now recognised that for mixtures containing many types of interaction sites, where chemical environment often changes with concentration and/or temperature, transferability becomes far more problematic.

A further issue is the connection between the dynamics of a coarse-grained model and the underlying atomistic system. The removal of degrees of freedom results in faster dynamics, and a more computationally efficient model, but the exact relationship between the coarse-grained and atomistic dynamics can be complex. As polymer melts are usually isotropic, it was initially hoped that simple time-scale scaling could be used to link dynamic properties between atomistic and coarse-grained levels.[48] However, for systems with anisotropy or inhomogeneity, or simply with different activation barriers for different processes, there is no guarantee that different dynamic quantities are not accelerated by different amounts in moving from an atomistic to a coarse-grained model. This must be taken into account when interpreting the results of any coarse-grained simulation. Some progress has been made in addressing the dynamics of coarse-grained models in a systematic way, by introducing additional terms which account for the removed degrees of freedom.[49,50] However, such methods do increase the complexity of the model, so may not be necessary in cases where only equilibrium structural or thermodynamic properties are of interest.

From a chemical perspective, two key areas where more accurate coarse-grained models are most needed are in the prediction of local structure and free energy changes, as a function of changing temperature and concentration. The former underpins the use of coarse-grained models to predict complex supramolecular or self-organised structures. This is vital in many areas of biochemical modelling, including prediction of membrane structure and stability[51–53] and protein/nucleic acid interactions.[54] Structure prediction also underpins many important areas of soft matter chemistry, such as micelle formation[55,56] and the formation of the microphase-separated structures seen in lyotropic liquid crystal[28,41,57,58] and block copolymer
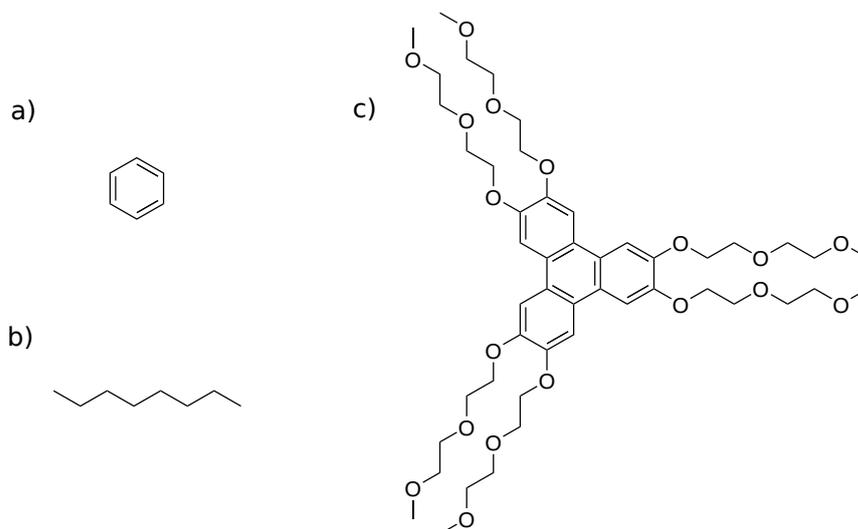
Figure 1.2: Molecular structures of a) benzene, b) octane and c) TP6EO2M.

phase diagrams.[59–61] An accurate representation of free energy changes underpins the prediction of thermotropic liquid crystal phase diagrams,[62,63] liquid-liquid miscibility and the phase boundaries in many industrially important soft matter systems formed from polymers and/or surfactants. Hence, there is considerable interest in methods that can automatically generate a coarse-grained model (either from a *bottom-up* or *top-down* perspective) which is both representative in terms of local structure and free energy changes, and sufficiently transferable to be used over a range of temperatures and concentrations.

1.5 Aims and scope of thesis

There is now a very wide range of methods for constructing coarse-grained models. However, there is currently no coarse graining method that can reproduce all the relevant characteristics of a system to the required degree of accuracy, and with sufficient thermodynamic and chemical transferability; in fact, the question of whether this is possible in a model which also significantly improves upon the computational efficiency of atomistic models is yet to be answered. Therefore, there is still work to be done before the most effective coarse graining method becomes apparent.

The overall aim of this thesis is to investigate a range of methods for parametrising coarse-grained models, both *bottom-up* and *top-down*, for two contrasting systems: an octane-benzene liquid mixture and the chromonic liquid crystal TP6EO2M, the structures of which are shown in Figure 1.2. The octane-benzene mixture was chosen because it forms a miscible mixture across the full composition range, has

both flexible (octane) and rigid (benzene) components, and requires different coarse-grained bead types, resulting in multiple, distinct bead-bead partial radial distribution functions (RDFs). The TP6EO2M system is considerably more challenging to coarse grain. The TP6EO2M molecule consists of distinct hydrophilic and hydrophobic groups which affect its behaviour in aqueous solution, and as a liquid crystal it is structurally more complex than a homogenous liquid. These factors make constructing a coarse-grained model significantly more challenging. In both cases, the representability and transferability of coarse-grained models parametrised using different methods will be examined, and approaches for improving the models investigated.

Chapter 2 will give the theoretical background to this project. It will begin by describing, in general terms, the simulation and analytical methodology used in this thesis. Details of the theory and implementation of a range of coarse-grained methods will then be provided, along with examples of past applications.

Chapters 3 and 4 will deal with the testing of several coarse graining methods on a binary liquid mixture of octane and benzene. Initially, the common *bottom-up* methods, HFM and IBI, will be compared to *top-down* coarse graining via the SAFT-γ Mie equation of state. Methods for improving on the transferability of the *bottom-up* methods will then be investigated, including the use of multiple reference systems with the multi-state IBI (MS-IBI) method and addressing the problem of overfitting using simple Morse potentials. Finally, preliminary work on the use of local density-dependent potentials, which represent a compromise between transferability and the complexity of a coarse-grained model, will be described.

Chapters 5 and 6 will then focus on parametrising coarse-grained models for the non-ionic chromonic liquid crystal TP6EO2M in aqueous solution. The first aim here will be to examine whether it is possible to parametrise an effective *bottom-up* model using HFM for a system such as this, where the atomistic reference simulation is more difficult to perform. A range of *top-down* methods will then be applied, and tested for their ability to produce molecular stacking and liquid crystal phase formation.

Finally, Chapter 7 will provide a summary of the results and conclusions of the thesis, and suggest future research directions which build on what has been done.

Chapter 2

Simulation methods and coarse-graining theory

2.1 Introduction

Throughout this thesis, a range of simulation and analysis methods will be used to test the representability and transferability of coarse-grained models. These coarse-grained models will be parametrised using several coarse graining methods, both *bottom-up* and *top-down*. The aim of this chapter is to provide details of the methods used and other, related, methods and to act as a reference for the rest of the thesis.

The chapter will begin by describing the theory behind molecular dynamics simulations, then the methods used to analyse the structure and thermodynamics of the simulations carried out. Since there is a wide variety of simulation methods available, the advantages of using particular methods will also be discussed. These sections will cover the methods in general terms, with specific details of how they were applied to particular systems given in the relevant results chapters.

The remaining sections will give the theoretical basis behind various coarse graining methods, starting with *top-down* coarse-graining, moving on to *bottom-up* coarse-graining based on matching structure and forces/energies, and then describing more recent work on coarse-grained models extended beyond pair potentials. This will include all of the coarse graining methods used in later chapters; however, other common methods will also be discussed to provide further context for how the different coarse-graining methods are linked.

2.2 Molecular dynamics

2.2.1 Equations of motion

Molecular dynamics (MD) is the simulation of the behaviour of atoms and molecules with time. In classical MD, this is accomplished by solving Newton's equations of motions:

$$\boldsymbol{F} = m\frac{d^2\boldsymbol{x}}{dt^2}, \tag{2.1}$$

where m is the mass of a particle, $\boldsymbol{x}$ is its positions and t is time. The forces, $\boldsymbol{F}$, on each atom are calculated using a force field, as described in Section 2.2.2. There are a number of algorithms for calculating the motions of the particles. The simplest of these is the Verlet algorithm:

$$\boldsymbol{x}_{i+1} = 2\boldsymbol{x}_i - \boldsymbol{x}_{i-1} + \Delta t^2 \boldsymbol{a}_i, \tag{2.2}$$

where $\boldsymbol{x}_i$ is the position of a particle at time step i, $\boldsymbol{a}$ is its acceleration and Δt is the time step.

The Verlet algorithm is time-reversible and provides good energy and momentum conservation. However, the addition of a small number $(\Delta t^2 \boldsymbol{a}_i)$ to a larger number can cause numerical errors which can propagate over longer trajectories. This issue was addressed in the leapfrog[64] and velocity Verlet[65] algorithms, both of which make use of the velocity at the midpoint between two time steps, $\boldsymbol{v}_{i-\frac{1}{2}}$. The leapfrog algorithm calculates the positions and velocities of the particles using:

$$\begin{aligned} \boldsymbol{x}_i &= \boldsymbol{x}_{i-1} + \boldsymbol{v}_{i-\frac{1}{2}}\Delta t \\ \boldsymbol{v}_{i+\frac{1}{2}} &= \boldsymbol{v}_{i-\frac{1}{2}} + \boldsymbol{a}_i\Delta t \end{aligned} \tag{2.3}$$

These equations do not contain a term proportional to Δt^2, and so the numerical errors associated with the Verlet algorithm will not occur in the leapfrog algorithm.

No matter which algorithm is chosen, the choice of Δt is very important to the stability of an MD simulation. All of the motions being modelled will occur on different timescales, and it is crucial that Δt is small enough to capture them all. In general, this means a timestep which is at least 10 times shorter than the period of the highest frequency motion in the simulation. In most all-atom simulations, the

bond vibrations are the limiting factor, and time-steps are on the order of 1 fs.

2.2.2 Force fields

Running a molecular dynamics simulation requires a way to calculate the forces acting upon the particles in the system. In classical molecular dynamics, this is done using a force field, in which the total energy of the system is split up into a number of different contributions which approximate the true energy of the system. In general, the total potential energy in a classical force field is given by:

$$U_{\text{tot}} = U_{\text{vdW}} + U_{\text{elec}} + U_{\text{bond}} + U_{\text{angle}} + U_{\text{dihedral}}. \tag{2.4}$$

These contributions can be split into two categories: bonded interactions (U_{bond}, U_{angle} and U_{torsion}) and non-bonded interactions (U_{vdW} and U_{elec}).

The terms in Equation 2.4 are used in most force fields, but many force fields will use additional terms. The specific terms which are used will be decided based on a trade-off between accuracy and computational cost. Many force fields have been developed for a range of different purposes, particularly for simulating biomolecules[66,67] and small organic molecules[18,68], and these force fields often have poor transferability to different system types. Therefore, the choice of which force field to use will have a big effect on the reliability of a simulation, and depends heavily on the system and properties of interest. In some cases, where existing force fields perform poorly, it may be necessary to either develop new force fields or reparametrise existing ones to accurately describe a system.

All-atom force fields are often parametrised using a combined *top-down/bottom-up* approach, in which parameters are fitted to combination of experimental data and quantum mechanical calculations. For example, in the GAFF force field,[68] the bonded parameters were fitted to reproduce structural data from high level quantum calculations and experimental crystal structures. Non-bonded parameters were taken from the Amber force field,[69] some of which derives from the earlier OPLS[70] force field; these non-bonded parameters were all originally fitted to reproduce experimental thermodynamic data. This reuse of parameters between force fields is relatively common, since it allows more time to be spent on parametrising the interactions which are judged to need improvement.

The functional form of the terms in Equation 2.4 can vary considerably between

force fields. In all-atom force fields, and some coarse-grained models, they usually take simple analytical forms. However, many coarse-grained force fields use numerical potentials, in which U is tabulated and is not described by any analytical function; this allows for much greater flexibility when parametrising a force field.

Bonded interactions

The internal structure of individual molecules has a significant effect on the behaviour and properties of most chemical systems. It is therefore very important that force fields are able to enforce the correct molecular geometry, and this is achieved using bonded interactions.

The bond stretching potential U_{bond} acts between sites which are directly connected, and is usually represented by a harmonic potential:

$$U_{\text{bond}}(r_{ij}) = \frac{K}{2}(r_{ij} - r_0)^2, \tag{2.5}$$

where r_{ij} is the distance between sites i and j, r_0 is the equilibrium bond length, and K is the force constant. This potential form has a minimum at the equilibrium bond length, so enforces the correct bond lengths in the system.

U_{angle} is the angle bending potential, and acts between sites separated by two bonds. Similarly to the bond length, it is often represented by a harmonic potential:

$$U_{\text{angle}}(\theta_{ijk}) = \frac{K_\theta}{2}(\theta_{ijk} - \theta_0)^2, \tag{2.6}$$

where θ_{ijk} and θ_0 are the angle connecting sites i, j and k, and the equilibrium bond angle, respectively.

U_{dihedral} acts on groups of four atoms. If these atoms are consecutively linked by bonds, this is called a proper dihedral; otherwise it is an improper dihedral. In general, these interactions describe the flexibility of a molecule. Proper dihedrals define the potential energy for rotation around a chemical bond. A common choice of functional form is the Ryckaert-Bellemans function:

$$U_{\text{proper}}(\phi_{ijkl}) = \sum_{n=0}^{5} K_n(\cos(\phi_{ijkl}))^n, \tag{2.7}$$

where ϕ_{ijkl} is the proper dihedral connecting sites i, j, k and l (defined as the angle between the ijk and jkl planes), and K_n are a set of coefficients. This functional

form gives a large amount of flexibility in terms of the periodicity and general shape of the dihedral potential, allowing it to be applied to wide range of bond rotations.

Improper dihedrals are commonly used to keep aromatic rings planar. This can be achieved using a simple periodic potential form:

$$U_{\text{improper}}(\xi_{ijkl}) = K(1 + \cos(n\xi_{ijkl} - \xi_0)) \tag{2.8}$$

where ξ_{ijkl} is the improper dihedral connecting sites i, j, k and l, n defines the periodicity of the potential and ξ_0 is the angle at which the potential is at its maximum.

Constraints

While getting the molecular geometry correct is crucial in MD simulations, directly modelling the vibrations of bonds and angles comes at a computational cost. The main reason for this is that these motions limit the timestep which can be used, as discussed in Section 2.2.1. It is possible to replace bonds (and sometimes angles) with constraints. Constraints ensure that the bond lengths are correct, while removing the need to explicitly model bond vibrations, therefore allowing larger timesteps, generally of 2–4 fs, to be used.

The general idea behind constraints is that, at each time step, the coordinates of the particles are updated as usual, and these coordinates are then modified so that the constraints are satisfied. One of the more efficient constraints algorithms is the Linear Constraint Solver (LINCS).[71] In the LINCS algorithm, the modification of the coordinates is done in two steps. Firstly, for each bond constraint, the updated bond is projected onto the same bond from the previous time step. The length of the bond is then corrected to the specified constraint length. The more interconnected the system, the more difficult it becomes to simultaneously satisfy all of the constraints.[71]

Non-bonded interactions

All of the important interactions not covered by the bonded terms in the force field are covered by non-bonded interactions. This includes the interactions between molecules (intermolecular interactions) and between atoms in the same molecule (intramolecular interactions). In most force fields, the non-bonded contribution to

the potential energy is split into two parts: the van der Waals energy (U_{vdW}) and the electrostatic energy (U_{elec}).

U_{vdW} is a combination of a number of different interactions, including steric repulsion, induction and dispersion. In atomistic force fields, it is most commonly described by the Lennard-Jones potential:

$$U_{\mathrm{LJ}}(r) = 4\varepsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^{6} \right], \tag{2.9}$$

where ε is the potential well depth, and σ is the distance where $U_{\mathrm{LJ}} = 0$. The important features of this potential are the repulsive wall, which prevents sites from overlapping, and the attractive well, which defines the distance at which two sites interact the most strongly, and the strength of this interaction. Another important feature is that, at larger distances, the force goes to zero, so the particles do not interact at all. Lennard-Jones potentials are ideally suited for the interactions between atoms. However, in coarse-grained models, softer potentials are often desired. There are a number of other functional forms which can be used this purpose; these include the Mie potential, which is a generalised Lennard-Jones potential where the exponents are not restricted to 12 and 6, and the Morse potential, which is given by:

$$U_{\mathrm{Morse}}(r) = \varepsilon \left(e^{-2\alpha(r-r_0)} - 2e^{-\alpha(r-r_0)} \right), \tag{2.10}$$

where r_0 is the distance where $U_{\mathrm{Morse}}(r) = -\varepsilon$, and α controls the curvature of the potential. The Morse potential shares the key features of the Lennard-Jones potential described above. However it, along with numerical potentials, is useful for coarse-grained models because it allows greater control over the softness or hardness than the Lennard-Jones potential.

Calculating U_{vdW} for each pair of interactions in the system would be extremely computationally expensive. However, since U_{vdW} tends to zero for larger distances, it is possible to employ a potential cut-off. This means that U_{vdW} is only calculated for pairs of sites where r is less than a chosen cut-off, r_{cut}. The value of r_{cut} must be large enough that important interactions are not neglected. This can be implemented by simply shifting the entire potential by a constant such that it goes to zero at r_{cut}. However, this does not guarantee that the forces are smooth at the cut-off, and so often a shifting function is added to the potential, such that both the potential and

the forces are smooth:

$$U(r) = U_{\text{LJ}}(r) + U_{\text{shift}}(r) \tag{2.11}$$

Electrostatics

In most classical models, each atom or site will have a partial charge associated with it. The interactions between these partial charges are described by U_{elec}. The simplest and quickest method is to calculate:

$$U_{\text{elec}} = \frac{q_i q_j}{4\pi\epsilon_0 r}, \tag{2.12}$$

where q_i is the charge on atom i and ϵ_0 is the vacuum permittivity, for all particles within a given cut-off. However, this approach neglects long-range electrostatics, which are important to the behaviour of many systems. While U_{elec} does tend towards 0 as r increases, it does so much more slowly than U_{vdW}, so the cut-off needed to include important long-range interactions would be very large.

A number of methods for including long-range electrostatics have been developed over the years. One of the more commonly used ones is the particle-mesh Ewald (PME) method. In PME, an electrostatic cut-off is defined which splits the electrostatic energy into a short-range and long-range contribution, depending on the distance between two particles. The short-ranged part is calculated as before using Equation 2.12, and the long-ranged part is calculated by a summation in Fourier space;[72,73] this requires that the system is periodic, so is used alongside periodic boundary conditions, as described in Section 2.2.3. This is much cheaper than calculating all of the electrostatics with Equation 2.12, and including the long-range electrostatic contribution gives a significant increase in accuracy.

It is possible to go beyond simple point-charge models, and introduce polarisability to classical force fields. Polarisable models are able to react dynamically to their electrostatic environment, and so are often more transferable than point-charge models. Polarisation can be included by simply adding an extra particle which is attached by a spring to an atom, and interacts only electrostatically with the rest of the system; this is the approach taken in the Drude water model.[74] Alternatively, the AMOEBA force field includes dipoles and quadrupoles, the strength of which can vary according to their environment.[75]

2.2.3 Periodic boundary conditions

The computational cost of running an MD simulation increases with the number of atoms in the system. Therefore, while it is possible in principle to model an extremely large number of atoms, the system size is limited by the computational power available. For example, accurately modelling even a simple bulk liquid would require an unfeasibly large number of molecules; there must be enough molecules that are far enough away from the liquid/vapour phase boundary to capture the behaviour of a liquid. Periodic boundary conditions[76] can dramatically reduce the computational cost, and so are very commonly used for the simulation of a large system like this.

Simulations with periodic boundary conditions involve modelling a small unit cell which is repeated infinitely in three dimensions. This effectively allows one to model an infinite number of atoms in all directions, significantly reducing the computational cost of modelling bulk systems. However, care is still needed when setting up the periodic boundary conditions; the unit cell must be large enough that a molecule cannot interact with its periodic images. Cubic or rectangular unit cells are the most common, and are sufficient for the majority of liquid-state simulations. However, particularly when simulating solid crystalline materials, more complex unit cells can also be used.[77]

2.2.4 Thermodynamic ensembles

For a simulation to have physical meaning, there must be some way of controlling the set of conditions under which they system is being studied. In statistical thermodynamics, these conditions are known as ensembles. The two ensembles most commonly used in molecular dynamics are constant-NVT and constant-NPT, where N is the number of particles, V is the volume of the simulation box, T is temperature and P is pressure. The difference between these two ensembles is that in constant-NVT the volume of the box is fixed, and in constant-NPT the box volume is allowed to vary so that the average pressure remains constant. The best choice of ensemble depends on the properties of interest; for example, the NPT ensemble allows one to study the influence of pressure on other system properties. Modelling a fixed number of particles or box volume is straightforward; however, keeping the temperature or pressure constant is less simple, and in practice this is done using,

respectively, thermostats and barostats.

Thermostats

The temperature of a system can be calculated from its kinetic energy, E_{kin} using:

$$T = \frac{2E_{\text{kin}}}{3Nk_{\text{B}}},$$ (2.13)

where N is the number of atoms.

$$E_{\text{kin}} = \frac{1}{2}\sum_i^N m_i v_i^2.$$ (2.14)

If the temperature is proportional to the velocities of the particles, then a simple way to change the temperature of the system is to alter the velocities; this is how thermostats work.

One of the simplest, and most computationally efficient, thermostats is the Berendsen thermostat[78], which scales the velocities of all particles by the same factor, λ:

$$\lambda = \left[1 + \frac{\Delta t}{\tau_T}\left(\frac{T_0}{T} - 1\right)\right]^{\frac{1}{2}},$$ (2.15)

where τ_T is the temperature coupling constant, T is the temperature, and T_0 is the reference temperature. This method will scale the velocities so that the total kinetic energy, and therefore the temperature, is correct. However, since all velocities are scaled by the same factor, the Berendsen thermostat does not generate a Boltzmann distribution of velocities, so the system will not accurately describe the canonical ensemble. Therefore, it is often used just for equilibration, and thermostats that generate the correct canonical ensemble, such as the Nosé-Hoover thermostat[79,80], are used in production runs.

The Nosé-Hoover thermostat modifies the equations of motion to introduce a friction coefficient, ζ:

$$\frac{d\zeta(t)}{dt} = \frac{1}{Q}\left[\sum m\boldsymbol{v}(t)^2 - (X+1)k_{\text{B}}T\right],$$ (2.16)

where Q is a coupling constant and X is the number of degrees of freedom in the system. Alternatively, a stochastic thermostat can be used, which introduces random fluctuations to the motions of the particles. Both of these methods will generate a

Boltzmann distribution of velocities, and so correctly describe the canonical ensemble.

Barostats

In molecular dynamics simulations pressure is calculated using the virial formula:

$$P_{\text{vir}} = \frac{1}{V} \left(N k_{\text{B}} T + \frac{1}{3} \left\langle \sum_{i}^{N} \boldsymbol{r}_i \cdot \boldsymbol{F}_i \right\rangle \right), \qquad (2.17)$$

where V is the system volume. Pressure is inversely related to volume, and so pressure coupling can be carried out by scaling the size of the simulation box. The Berendsen barostat[78] works in a similar way to the Berendsen thermostat, by scaling the coordinates and box vectors by a factor of μ:

$$\mu = 1 - \frac{\kappa_T \Delta t}{3 \tau_P} \left(P_0 - P \right), \qquad (2.18)$$

where κ_T is the approximate isothermal compressibility of the system, τ_P is a pressure coupling constant, P is the pressure and P_0 is the reference pressure. This barostat generates very large pressure fluctuations. Like the Berendsen thermostat, the Berendsen barostat is usually only used for equilibration, with the more complex Parrinello-Rahman barostat[81] used for production runs. The Parrinello-Rahman barostat has two major advantages: it generates more realistic pressure fluctuations than the Berendsen method; and it allows the cell vectors to change independently, so that the shape as well as the volume of the simulation box can change. It is, however, more expensive than the Berendsen barostat.

2.2.5 State-point dependent potentials

As mentioned in Chapter 1, the use of state-point dependent potential in molecular dynamics simulations will impact the correct expressions for calculating various thermodynamic observables. One important example is the case of calculating pressure. The virial expression for calculating pressure given in Equation 2.17 is only valid in cases where the pair potentials used are not volume/density dependent. In cases where they are volume dependent the correct expression is:

$$P = \frac{1}{V} \left(N k_{\text{B}} T + \frac{1}{3} \left\langle \sum_{i}^{N} \boldsymbol{r}_i \cdot \boldsymbol{F}_i \right\rangle \right) - \left\langle \frac{\partial U}{\partial V} \right\rangle. \qquad (2.19)$$

This is very similar to the virial expression, but includes an additional term describing the volume dependence.[82]

Other observables are also affected by state-point dependent potentials. For example, when the potential depends on temperature, the total energy of a system, E, should be given not simply by the Hamiltonian, H, but by:

$$E = \left\langle H + \frac{\partial H}{\partial \ln \beta} \right\rangle, \qquad (2.20)$$

where $\beta = \frac{1}{k_B T}$.[83]

It should be noted that, throughout this work, the standard expressions for thermodynamic observables have been used, rather than those modified to take account of state-point dependence.

2.3 Simulation analysis

2.3.1 Radial distribution functions

MD simulations are often used to study the structure of chemical systems. This can be done visually, by looking at how the simulation trajectory evolves over time, but it is often useful to have a more quantitative way to look at structure. The radial distribution function, $g(r)$, describes the probability of a particle being present at a distance r from another particle in the simulation box. This is normalised relative to the distribution expected for an ideal gas of the same density:

$$g(r) = \frac{1}{4\pi r^2 N \rho} \sum_i^N \sum_{j \neq i}^N \langle \delta(r - r_{ij}) \rangle, \qquad (2.21)$$

where δ is the Kronecker delta function. The shape of $g(r)$ indicates how ordered a system is, and can therefore be used to distinguish between different states of matter. The shape of the peaks may also be used to look at finer structural features.

It is possible to define partial radial distribution functions for different pairs of particle types. For example, $g_{AB}(r)$ would be the probability of two particles of type A and B being found a distance of r from each other. In this case, the $N\rho$ term in Equation 2.21 is replaced by $N_A N_B / V$

2.3.2 Free energy calculations

The Bennett acceptance ratio (BAR) method is a commonly used method for calculating the free energy difference (ΔF_{BA}) between two states:

$$\Delta F_{\mathrm{BA}} = RT \ln \frac{\langle f(H_A - H_B + C) \rangle_{\mathrm{B}}}{\langle f(H_B - H_A - C) \rangle_{A}} + C, \tag{2.22}$$

where C is a shift constant and $f(x)$ is the Fermi function:

$$f(x) = \frac{1}{1 + e^x}, \tag{2.23}$$

A separate simulation is run for each state, A and B. At regular intervals in the simulation, the Hamiltonian (H) is calculated separately using the force fields for states A and B, and the difference between these two values is determined. The value of C for which $\langle f(H_A - H_B + C) \rangle_B = \langle f(H_B - H_A - C) \rangle_A$ is then calculated numerically. This gives an estimate of the energy shift required such that the probabilities of the forward and reverse processes are equal, and therefore an estimate of the free energy difference as $\Delta F_{\mathrm{BA}} \simeq C$.[84]

The free energy estimate will only be accurate if there is sufficient overlap between the energy distributions of the two states. Therefore, for most real cases, it is necessary to consider a series of alchemical states between A and B, so that each neighbouring state has a large energy overlap. The total free energy difference is then calculated as the sum of all the intermediate free energy differences (see Equation 2.24

$$\Delta F_{\mathrm{BA}} = \sum_{i=1}^{n-1} \Delta F_{i+1,i} \tag{2.24}$$

Linear decoupling of van der Waals interactions can result in singularities close to the end points of the decoupling. When the interaction is almost fully decoupled, U will be close to zero at all distances except very close to $r = 0$, where U jumps to a high value. This problem is solved by using soft-core potentials for the intermediate states:

$$U_{\mathrm{sc}}(r) = (1 - \lambda)U_{\mathrm{A}}(r_{\mathrm{A}}) + \lambda U_{\mathrm{B}}(r_{\mathrm{B}}) \tag{2.25}$$

$$r_{\mathrm{A}} = \left(\alpha \sigma_{\mathrm{A}}^6 \lambda^p + r^6 \right)^{\frac{1}{6}} \tag{2.26}$$

$$r_{\mathrm{B}} = \left(\alpha \sigma_{\mathrm{B}}^6 (1 - \lambda)^p + r^6 \right)^{\frac{1}{6}} \tag{2.27}$$

where λ is the coupling parameter, α is the soft-core parameter and p is a positive integer. α and p are usually chosen to be 0.5 and 1, respectively.

2.3.3 Potentials of mean force

The potential of mean force (U_{PMF}), the potential which reproduces the average forces over a particular reaction coordinate, is often used as a measure of the free energy profile for that coordinate.[85] In the case of a PMF for the separation of two particles, this is the reversible work associated with moving the particles from infinity to a distance of r, which is related to the radial distribution function by:

$$U_{\mathrm{PMF}}(r) = -k_{\mathrm{B}} T \ln g(r). \tag{2.28}$$

PMFs can be calculated by carrying out a series of simulations where the molecule or molecules of interest are constrained at a range of points along the reaction coordinate of interest, s. When s is the separation distance between two molecules, U_{PMF} can be calculated at $s = r$ by integrating the average constraint force between r and r_{max}, where $U_{\mathrm{PMF}} = 0$, as described in Equation 2.29. r_{max} must be large enough that U_{PMF} does not change on increasing r further.

$$U_{\mathrm{PMF}}(r) = \int_r^{r_{\mathrm{max}}} \left[\langle f_{\mathrm{C}} \rangle + \frac{2 k_{\mathrm{B}} T}{s} \right] ds. \tag{2.29}$$

The second term here is an entropic term which accounts for the increased rotational volume which is available at larger separation distances.[86]

2.4 *Top-down* coarse graining

2.4.1 MARTINI force field

The most frequently used *top-down* force field is MARTINI, developed by Marrink *et al* in 2007[32]. This model, based on an earlier *top-down* lipid model by the same authors,[87] consists of a large set of pre-defined coarse grained interaction sites, each of which represents a different type of functional group commonly found in biomolecular systems. The non-bonded interaction potentials for these

sites are represented by 12-6 Lennard-Jones functions, parametrised to reproduce the partitioning of small molecules between a set of organic solvents. Matching to partitioning free energies is advantageous, as it ensures that the overall free energy of a system is reproduced reasonably well, even if the entropy/enthalpy balance is incorrect because of the coarse graining process. The bonded interactions are represented by weak harmonic oscillators, and are parametrised to reproduce molecular structures from atomistic simulations. The particle types and interaction potentials for a particular system are chosen based on the thermodynamic properties of that system, as well as comparison with atomistic simulations.

One clear advantage of the MARTINI force field is its ease of use compared to other coarse-graining methods. Since all of the bead types and their interactions are predefined, one only has to select an appropriate bead type for each coarse grained site. However, this clearly limits the flexibility of the method, and means that it cannot always be fine-tuned for every system.

Because of this ease of use, MARTINI has been extended to many different systems types, including proteins[88,89], carbohydrates[90] and polymers[91,92] and has given some useful qualitative and semi-quantitative results relating to their structure and dynamics which would not have been accessible at the atomistic level.

The use of sharp 12-6 functions for the non-bonded potentials in MARTINI is advantageous because the majority of existing MD codes are capable of handling this type of function. However, the process of coarse graining should result in softer interaction potentials, and the 12-6 functions implemented in MARTINI are generally too hard. This is particularly apparent in the case of MARTINI water, which freezes at around 300 K in the presence of a nucleation site. This issue was addressed by the addition of 'antifreeze' particles, which disturb the lattice packing of the water, and prevent freezing. However, the addition of antifreeze particles lowers the density and self-diffusion constant of bulk water by around 10%, and therefore reduces the accuracy of the model.[32] A more recent version of MARTINI uses a polarisable water, consisting of a central particle interacting via LJ interactions, and two additional oppositely charged beads which give the model orientational polarisability; this model improves many of the properties of MARTINI water, including the poor representation of the melting point.[93]

2.4.2 Statistical Associating Fluid Theory (SAFT)

Statistical associating fluid theory (SAFT) is an equation of state originally developed to calculate the thermodynamic properties and phase behaviour of fluids.[94] It describes the Helmholtz free energy of a system as:

$$A = A^{\text{ideal}} + A^{\text{mono}} + A^{\text{chain}}, \tag{2.30}$$

where A^{ideal} is the ideal Helmholtz energy, A^{mono} is the free energy of the monomers, and A^{chain} is the free energy due to chain formation.

SAFT has recently been used to parametrise coarse grained models for use in molecular dynamics, using a framework called SAFT-γ. In this framework, intermolecular interactions are described by a Mie potential, which is a generalised form of the 12-6 Lennard-Jones potential.[33,95]

$$U^{\text{Mie}} = C\varepsilon\left[\left(\frac{\sigma}{r}\right)^{\lambda_r} - \left(\frac{\sigma}{r}\right)^{\lambda_a}\right], \tag{2.31}$$

where ε is the well depth, σ is the radius of the interaction, r is the inter-particle distance, and λ_r and λ_a are repulsive and attractive exponents. The constant C is given by:

$$\frac{\lambda_r}{\lambda_r - \lambda_a}\left(\frac{\lambda_r}{\lambda_a}\right)^{\frac{\lambda_a}{\lambda_r - \lambda_a}}. \tag{2.32}$$

When λ_r and λ_a are 12 and 6, respectively, C equates to 4, and so the Mie potential is the same as a Lennard-Jones potential. The advantage of a Mie potential over a Lennard-Jones potential is that the variable exponents give the Mie potential a greater degree of flexibility, allowing the hardness or softness of a particle to be tuned depending on its properties.[95]

The SAFT-γ approach has been used to develop coarse grained potentials for a number of systems. These include small molecules such as carbon dioxide[33] and water[96], long chain linear alkanes[97], aromatic compounds[98] and a range of binary and ternary mixtures[99]. The SAFT-γ force field was parametrised to reproduce thermodynamic properties, and as such it performs very well for the calculation of thermodynamic properties such as vapour liquid equilibria, heat capacities and thermal expansion coefficients. The potentials are generally transferable over a large range of temperatures.

2.5 *Bottom-up* coarse graining: structure-based methods

2.5.1 Boltzmann inversion (BI)

One of the simplest methods for creating a coarse-grained interaction potential is by Boltzmann inversion of a probability distribution, as shown in Equation 2.33.[100]

$$U(q) = -k_{\mathrm{B}}T \ln P(q), \tag{2.33}$$

where $U(q)$ is the interaction potential, and $P(q)$ can refer to a coarse-grained bond, angle or dihedral distribution, or a non-bonded radial distribution function (RDF). The main assumption made here is that the interaction in question is not correlated with other interactions, meaning that the effective pair interaction is not affected by the presence of other particles in the system. This often works well for bonded interactions, and so is often the method used for parametrising these potentials. In the case of non-bonded interactions, Boltzmann inversion works well for gas-phase systems, but fails for liquids where there are significant correlations between the interactions.[29,100]

2.5.2 Iterative Boltzmann inversion (IBI)

According to the Henderson uniqueness theorem, if a potential is found which exactly matches the structure of the reference, then that potential is a unique solution to the structure-matching problem.[101] However, for non-bonded interactions in liquids, the potentials from Boltzmann inversion do poorly at matching the atomistic structure. Because these interactions are strongly correlated with other interactions, the potential of mean force will include contributions from multi-body correlation functions which must be considered when parametrising effective pair potentials. One method which implicitly includes these correlations is iterative Boltzmann inversion (IBI), in which an iterative scheme can be used to match the structure.[21]

$$U_{\mathrm{n}+1}(r) = U_{\mathrm{n}}(r) + k_{\mathrm{B}}T \ln \frac{g_{\mathrm{n}}(r)}{g_{\mathrm{target}}(r)} \tag{2.34}$$

An initial guess is made by Boltzmann inversion of the radial distribution function for the interaction, obtained after mapping the atomistic trajectory to a coarse-

grained representation. A short coarse grained simulation is carried out using this potential, and the potential is updated according to the difference between the RDF from this simulation and the target RDF, as shown in Equation 2.34. This procedure is carried out iteratively until the two RDFs match to an acceptable degree.

Additional potential updates can also be applied to match other properties of the reference system. For example, the interaction potentials from IBI do not reproduce the pressure of the atomistic reference, since coarse-grained pair potentials are volume dependent, and this affects the representation of many thermodynamic properties such as pressure. For this reason, pressure correction is usually carried out along with IBI, where the potential is updated according to Equation 2.35, which applies a linear correction.[21,29,102]

$$\Delta U(r) = A\left(1 - \frac{r}{r_{\text{cut}}}\right), \tag{2.35}$$

where

$$A = \text{sgn}(\Delta P)0.1k_{\text{B}}T\min(1, |f\Delta P|). \tag{2.36}$$

This form of A scales the potential update according to the difference in pressure between the coarse-grained and reference systems, ΔP, and a tuning parameter f, which is used to help with the convergence of the iterations.

It should be noted that matching the pressure of the reference means that the compressibility will no longer be accurately reproduced. Isothermal compressibility is obtained by integration of $g(r)$ and so is very sensitive to small errors in $g(r)$ caused by including the additional pressure correction in the IBI procedure. Therefore, one must choose which of the two quantities is most important for the system being studied.[103]

The potentials constructed using IBI have been parametrised specifically to reproduce the radial distribution functions of a system; therefore, this method works very well for studying structural features. There have been several studies in which IBI has been used to investigate structural features of polymers; these features are often inaccessible to atomistic simulations due to the long timescales and large systems involved. This was done in one of the very first IBI studies in 2005, where the experimental dependence of the gyration radius of polystyrene on molecular weight was reproduced extremely well using IBI potentials.[104] The method has also been applied to polymer blends, such as a mixture of polyisoprene and polystyrene; in this

case the aggregation behaviour of the different chain types agreed only qualitatively with experiment.[105].

IBI has also been applied to molecular liquids. It is able to accurately reproduce the bulk density of a range of compounds, including hydrocarbons such as hexane[30] and toluene[106], and polar compounds such as water[107]. In all of these cases, pressure correction was necessary to obtain the correct density.

There are also a number of disadvantages of using IBI. One is that, in systems with many types of interaction site, where many different potentials must be parameterised, it can be difficult achieve convergence of all of the RDFs simultaneously. This is because there will be cross-correlations between the different interactions, therefore updating one interaction potential to improve its RDF may adversely affect the RDFs for other interactions.[29]

A recent extension to IBI has allowed more accurate modelling of the solvation behaviour of materials. This involves adding a linear correction to the potentials, similar to the linear pressure correction described above, so that the Kirkwood-Buff integrals (KBIs), which are related to the solvation free energy, match those of the atomistic reference; this is called KB-IBI. It was shown that KB-IBI reproduces the solvation behaviour of benzene-water and urea-water mixtures better than standard IBI.[108,109]

2.5.3   Inverse Monte Carlo (IMC)

Inverse Monte Carlo (IMC) is similar to IBI in that it involves iteratively updating potentials to match the structures of atomistic and coarse-grained systems. It also begins with an initial guess from Boltzmann inversion of a radial distribution function; however, the potential update is calculated from:

$$\langle S_\alpha \rangle - S_\alpha^{\text{ref}} = A_{\alpha\gamma}\Delta U_\gamma. \tag{2.37}$$

Here, the interaction potential is grid-based, and α and γ are indices which refer to the grid-points (across all pair potentials in a multi-component system). S_α is the number of pairs of particles corresponding to grid-point α, ΔU_γ is the potential update, and $A_{\alpha\gamma}$ is $\frac{\partial \langle S_\alpha \rangle}{\partial U_\gamma}$.

Unlike IBI, IMC explicitly includes cross-correlations in the potential update (within $A_{\alpha\gamma}$). This means that, for systems with a large number of interactions, fewer

iterations are required to achieve convergence. However, in order to calculate these cross-correlations, more statistics are needed than in IBI, and so longer simulations are required for each iteration. IMC is therefore more computationally demanding than IBI.[29]

IMC has been less commonly used than IBI; however, there have been some notable studies. It was used to construct a potential for water from a vapour/liquid reference system, which allowed the RDF and pressure of liquid water to be matched without the need for additional pressure correction.[110] In one study, it was used to model several lipid bilayers, and was used to show events normally inaccessible to atomistic simulations, such as the self-assembly of the bilayer.[111] In another study, it was used to construct a model for a phospholipid/cholesterol bilayer, and was able to predict the presence of certain domains in the system; however, the authors noted that this information was qualitative rather than quantitative.[112]

2.5.4   General-purpose optimisation algorithms

It is also possible to use more general optimisation algorithms to parametrise coarse-grained force fields. The VOTCA package,[113] includes two such methods: the downhill simplex algorithm[114] and the Covariance Matrix Adaptation Evolution Strategy (CMA-ES)[115].

In the simplex method, a simplex with $n + 1$ vertices, where n is the number of parameters to be optimised, is constructed. Each vertex on the simplex represents a point in parameter space. A function can be minimised by performing a series of geometric transformations on the simplex to explore parameter space. In the context of coarse-graining, the parameters will be parameters of the coarse-grained model, and the objective function could be any property measured using that coarse-grained model. An example would be the difference between the coarse-grained and atomistic reference RDFs.

The algorithm is initialised by running coarse-grained simulations using the parameters represented by the $n + 1$ starting vertices, and calculating the objective function for each of these vertices. The vertices are then sorted from best to worst, according to the objective function. A combination of reflections, expansions, contractions and reductions is then carried out to try to find a minimum in parameter space.[114]

CMA-ES performs an optimisation by evaluating the objective function for a set of points in parameter space, randomly chosen to sample a normal distribution. The mean and width of the distribution are then updated according to the relative fitness values of each of the points; the new mean and width are then used to generate a new set of points in parameter space. This method performs better than the simplex method in cases where a large number of parameters must be optimised.[113,115]

2.5.5   Parametrisation from multiple states

In common with other *bottom-up* coarse graining methods, the potentials generated using IBI are highly state-point dependent, meaning that they cannot necessarily be transferred to systems with different temperatures or concentrations to the reference system. A number of methods have been developed to achieve better temperature transferability. For example, in one study, a temperature dependent scaling factor was applied, which allowed the density of toluene to be determined more accurately at different temperatures, although this was only possible over a certain temperature range.[46] An extension of this method was published by Farah *et al.* in which an IBI model for hexane was derived at two reference temperatures, and the temperature dependent coarse-grained potential was then determined by linear interpolation between those two reference temperatures:

$$U(r, T) = \frac{T - T_\mathrm{U}}{T_\mathrm{L} - T_\mathrm{U}} U(r, T_\mathrm{L}) + \frac{T_\mathrm{L} - T}{T_\mathrm{L} - T_\mathrm{U}} U(r, T_\mathrm{U}) \qquad (2.38)$$

The resulting range of models reproduced both the structures and densities of liquid hexane between the reference temperatures studied.[116] However, the method assumes that the potential is linearly dependent on temperature. This will not always be a valid assumption, particularly in cases where there is a phase transition between the two reference temperatures.

Multi-state IBI (MS-IBI) was proposed by Moore *et al.* to address the transferability problem. The idea behind MS-IBI is to use multiple reference simulations to parametrise a single coarse-grained model; this should allow the model to be transferable between the state points used for the reference simulations. This method has been used to model simple linear alkanes[117] as well as more complex lipid bilayers[118].

The potential update in MS-IBI is given by taking the average of the IBI potential updates for each reference system:

$$\Delta U^{\text{MSIBI}} = \frac{1}{N} \sum_{i=1}^{N} \eta_i \Delta U_i^{\text{IBI}}, \tag{2.39}$$

where N is the number of reference systems, and η_i is a scaling factor for reference system i, chosen to aid with the convergence of the MS-IBI scheme.

Moore *et al.* only address the ability of MS-IBI to accurately reproduce structural properties across a range of state points. However, bulk properties such as density are not addressed. It is claimed that the MS-IBI approach could be extended to pressure correction, but this has so far not been tested.[117]

2.6 *Bottom-up* coarse-graining: force- and energy-based methods

2.6.1 Force matching (FM)

Force matching (FM), sometimes known as the multiscale coarse-graining method (MS-CG) is a method for the construction of coarse-grained potentials by mapping the interatomic forces from an atomistic simulation onto a coarse-grained system.

It is based on a variational principle (see Equation 2.40) which states that, by minimising the objective function (which is related to the difference between the potential of mean force (PMF) of the reference and coarse-grained systems), one approaches the true effective PMF for that system.[20] The variational principle was rigorously derived from statistical thermodynamics by Noid *et al.*,[119] and so the method has a much sounder basis in fundamental physics than other methods like IBI.

$$\chi^2 = \frac{1}{3LN} \sum_{l=1}^{L} \sum_{i=1}^{N} \left| \mathbf{F}_{il}^{\text{ref}} - \mathbf{F}_{il}^{p}(x_1, ..., x_M) \right|^2, \tag{2.40}$$

where $\mathbf{F}_{il}^{p}$ and $\mathbf{F}_{il}^{\text{ref}}$ are the total force on bead i in snapshot l, for the coarse-grained and reference systems, respectively, and $x_1, ..., x_M$ are coefficients of the functions to which the coarse-grained forces are fitted. The reference forces on a coarse-grained bead are obtained by summing the forces on each of the atoms which map onto that bead. In practice, there are several algorithms for the minimisation of χ^2, as described by Lu *et al.*[120]

In the block averaging approach, implemented in the VOTCA package, force

matching is done by solving a series of linear equations shown in Equation 2.41. The coarse grained PMF is constructed from a series of spline functions, and the coefficients of these functions are obtained from solving Equation 2.41, with the additional constraint that the first derivatives of the spline functions are continuous. In order to reduce the memory required for this calculation, the trajectory is split into blocks. Force matching is carried out separately for each block, the final result is calculated by averaging the results from all of the blocks. The maximum feasible block size is dictated by the memory available, although it should be large enough that Equation 2.41 is overdetermined.[20,29] It has been suggested that results could be improved by randomising the selection of frames in each block, which could help to prevent potential issues with blocks being correlated.[121]

$$\mathbf{F}_{\mathrm{cg}}(x_1, ..., x_M) = \mathbf{F}_{\mathrm{ref}}, \tag{2.41}$$

Alternatively, the BOCS package implements the normal equation algorithm. This approach significantly reduces the memory requirements of force matching by reducing the set of equations to be solved to Equation 2.42. This allows force matching to be carried out over the whole trajectory at once, without the need for block averaging.[31]

$$\mathbf{F}_{\mathrm{cg}}^T \mathbf{F}_{\mathrm{cg}}(x_1, ..., x_M) = \mathbf{F}_{\mathrm{cg}}^T \mathbf{F}_{\mathrm{ref}} \tag{2.42}$$

A number of different systems have been studied using the MS-CG method. Initially, it was applied to a simple lipid bilayer,[20] as well as to water and methanol,[122] and for both systems the coarse-grained RDFs showed reasonable agreement with the atomistic ones, although not as close as structure-based methods like IBI. Since then, it has been used to model a range of more complex systems, such as mixed lipid bilayers[123] and ionic liquids,[124] and in these cases a similar level of structural accuracy has been obtained. More recently, the MS-CG method was applied, along with a centre-of-charge mapping system, to model a series of polar organic molecules.[125]

Several possible improvements have been made since MS-CG was first introduced. Some of these involved the improvement of the basis functions used in the spline fitting. In one study by Das and Andersen, it was shown that using multiresolution basis functions could reduce statistical noise in the PMF obtained by MS-CG, although this was only tested on a simple two-component Lennard-Jones

fluid.[126] In another, it was shown that by using 3-body terms in the spline fitting, it was possible to achieve a much better match between the coarse-grained and atomistic RDFs of water.[127] This could prove important when coarse-graining aqueous systems, and other systems dominated by hydrogen bonding.

Like IBI, the potentials created using MS-CG do not reproduce the pressure of the atomistic system. Two possible ways have been developed of getting around this problem. Das and Anderson proposed the inclusion of an additional, volume-dependent, constraint during the force matching procedure.[128] This was tested on a simple two-component Lennard-Jones mixture, and gave accurate results for the pressure in the *NVT* ensemble, and the bulk density in the *NPT* ensemble. This volume dependent term is discussed further in Section 2.7.1. Izvekov *et al.* introduced an additional constraint to the force-matching equations so that the virial of the reference was matched.[129] On application of this method to nitromethane, it was found that this resulted in an approximately linear pressure correction term, similar to that used in IBI pressure correction.

Hybrid force matching (HFM) is a method which allows the combination of two different coarse graining methods for one system. In this case, the bonded potentials are obtained through simple Boltzmann inversion, while the non-bonded potentials are obtained as usual through force matching. The use of easily obtainable intramolecular potential functions guarantees sensible molecular geometries and helps eliminate some of the problems that can be associated with a lack of sampling of higher energy conformations in the atomistic reference system. In particular, this poor sampling can cause a poor representation of bonded interactions in standard MS-CG models, which is fixed by the use of bonded interactions from Boltzmann inversion.[29,130] A number of studies using this hybrid approach have found that is able to give good structural accuracy when applied to a range of soft matter systems.[30,124,131]

2.6.2   Generalised Yvon-Born-Green (g-YBG)

The generalised Yvon-Born-Green method (gYBG) is based on the Yvon-Born-Green equation in liquid state theory, and describes the relationship between a pair potential and the two- and three- body correlation functions in a molecular

system.[132] The g-YBG equation can be written as:

$$- k_{\mathrm{B}}T\nabla \ln g(r) = \nabla U(r) + \rho \int \nabla U(r') \left[\frac{G^3(r,r')}{g(r)} - g(r') \right] dr', \qquad (2.43)$$

where $G^3(r,r')$ is a three-body correlation function and ρ is the concentration.

This leads to a system of linear equations which can be solved to give the PMF for a particular interaction. It has been shown that this solution is the same as the optimal solution to the variational principle in MS-CG, so gYBG can in principle be used as a direct alternative to MS-CG, which does not involve force matching.[133]

The gYBG method has been shown to give similar results to MS-CG when applied to molecular liquids such as toluene[134] and heptane[135], although larger atomistic reference simulations were required to achieve this level of accuracy. It has also been used for more complex systems, specifically a poly(ethylene oxide) based ionomer, for which it was able to give a reasonably accurate picture of the solvation structure of the system.[136]

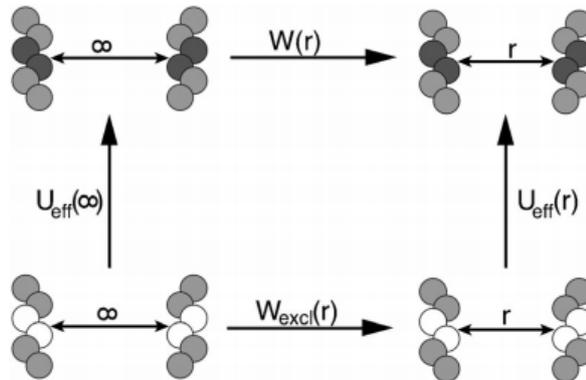### 2.6.3   Conditional Reversible Work (CRW)

Figure 2.1: Thermodynamic cycle used to calculate CRW interaction potentials. Reproduced from Ref. 106 with permission from the PCCP Owner Societies.

Unlike the other *bottom-up* methods discussed so far, where coarse grained models are parametrised based on matching certain properties of the atomistic reference, the conditional reversible work method (CRW) involves directly calculating the interaction potentials. This is done by utilising a thermodynamic cycle, shown in Figure 2.1, and using Equation 2.44. Figure 2.1 shows the procedure for calculating the interaction between two coarse-grained beads, each consisting of the two central carbon atoms of a UA hexane. $W(r)$ is obtained by calculating the PMF for the

separation of two hexane molecules, where r is the distance between the COMs of the two pairs of central atoms. $W_{\text{excl}}(r)$ is the same PMF, but this time calculated with the van der Waals interactions between the central carbons turned off. The difference between $W r)$ and $W_{\text{excl}}(r)$ gives the coarse-grained potential at a distance of r.[106] This method is usually combined with Boltzmann inversion for calculating bonded interactions.

$$U_{\text{eff}}(r) - U_{\text{eff}}(\infty) = W(r) - W_{\text{excl}}(r) \tag{2.44}$$

Since CRW potentials have been directly calculated, it is argued that they have more physical meaning than potentials obtained through other methods, and that this gives them greater transferability. It was shown that CRW potentials more accurately reproduce the atomistic thermal expansion coefficient when modelling liquid toluene than IBI potentials.[106] It should be noted that, for the temperature at which the potentials were derived, IBI was better able to reproduce the bulk density. In another study, it was shown that, in the case of simple linear alkanes, potentials generated for dodecane were able to accurately reproduce the atomistic reference density in coarse grained simulations of decane, octane and hexane.[137]

2.6.4   Relative entropy

The relative entropy for a coarse-grained trajectory, defined in Equation 2.45, can be used as a measure of how the coarse-grained system and the mapped atomistic system differ from each other[36]:

$$S_{\text{rel}} = \sum_{\mathbf{r}} p_{\text{AA}}(\mathbf{r}) \ln \left(\frac{p_{\text{AA}}(\mathbf{r})}{p_{\text{CG}}[M(\mathbf{r})]} \right) + S_{\text{map}}, \tag{2.45}$$

where $p(\mathbf{r})$ are probability distributions as a function of the configurations of the system, M is a mapping operator between a coarse-grained and an atomistic configuration, and S_{map} is the mapping entropy which accounts for the degrees of freedom which are removed upon coarse-graining. S_{map} does not depend on the coarse-grained interaction potential, so is not considered in the relative entropy parametrisation. This can be written in terms of coarse-grained and atomistic potential energy functions, as:

$$S_{\text{rel}} = \beta \left\langle U_{\text{CG}} - U_{\text{AA}} \right\rangle_{\text{AA}} - \beta \left(A_{\text{CG}} - A_{\text{AA}} \right) + S_{\text{map}}, \tag{2.46}$$

where $\beta = k_{\text{B}}T$ and S_{map} is the mapping entropy, which depends only on the coarse-grained mapping used.

The relative entropy can also be thought of as a measure of the quality of a coarse-grained model. Therefore, coarse-grained models can be parametrised by minimising the relative entropy in terms of a set of parameters of the model,[138] λ, using:

$$\frac{1}{\beta} \frac{\partial S_{\text{rel}}}{\partial \lambda} = \left\langle \frac{\partial U_{\text{CG}}}{\partial \lambda} \right\rangle_{\text{AA}} - \left\langle \frac{\partial U_{\text{CG}}}{\partial \lambda} \right\rangle_{\text{CG}}. \tag{2.47}$$

This can be done by running a simulation for a test set of parameters, and then iteratively altering the parameters to minimise S_{rel}. However, to avoid running a separate simulation for each λ, an energy reweighting method was developed which allows the same trajectory to be used for a range of trial coarse-grained models.[138] This reduces the problem to a minimisation of:

$$\Delta S_{\text{rel}} = -\beta \left(\left\langle \Delta U_{\text{CG}} \right\rangle_{\text{AA}} + \left\langle \Delta U_{\text{CG}} \right\rangle_{\text{CG},\lambda^0} \right), \tag{2.48}$$

where λ^0 is the set of parameters used to generate an initial coarse-grained trajectory. New coarse-grained trajectories must be generated periodically, to reduce the errors arising from large differences between λ and λ^0.

2.7 Beyond simple pair potentials

2.7.1 Volume potentials

As mentioned in Section 2.5.2, *bottom-up* coarse-grained models rarely match the pressure of their atomistic reference, requiring some sort of pressure correction term. While this is often a simple correction to the pair potential, other methods have been proposed. In 2010, Das and Andersen introduced a volume-dependent contribution to the coarse-grained potential, giving a non-bonded potential with the form:

$$U_{\text{nb}} = U_{\text{pair}}(\mathbf{r}) + U_V(V), \tag{2.49}$$

where

$$U_V(V) = N \left(\psi_1 \frac{V}{\bar{v}} + \psi_2 \left(\frac{V}{\bar{v}} - 1 \right)^2 \right) \quad (2.50)$$

V is the volume of the simulation box, $\bar{v}$ is the average volume of the atomistic reference and N is the number of coarse-grained beads in the system.[128] psi_1 and psi_2 are coefficients which are parameterised so that the pressure of the coarse-grained system matches that of the reference system. This is done by using force matching to find a non-bonded pair potential, then using a pressure matching procedure to optimise the parameters of the volume potential.

Dunn and Noid introduced an iterative pressure matching algorithm in 2015, which improved the representation of density and pressure fluctuations over Das and Andersen's method.[139] This method was then applied to hexane/toluene mixtures,[140] and it was shown that a concentration-transferable model could be obtained by using the same pair potential across concentrations, and carrying out pressure matching to determine the volume-dependent potential at each concentration of interest.

More recently, Rosenberger and van der Vegt applied pressure matching to obtain temperature-transferable models for a range of liquid alkanes.[141] In their approach, pressure matching was carried out only at the two end points of a temperature range of interest. The model for any temperature within that range can then be obtained by linear interpolation between U_V at the end points, in an analogous way to Equation 2.38.

2.7.2 Local density potentials

Volume-dependent potentials are a useful way of improving the transferability of coarse-grained models over those using using simple pair potentials. However, the volume-dependent contribution to the energy depends only on the global volume of the system. In systems with local variations in density, for example those containing phase boundaries, this may not be sufficient to accurately represent the interactions in the system.

The use of local density potentials in coarse-grained models originated in multi-body DPD[142], in which the interaction between two particles depends on the local particle density around each of the particles. This interaction type was first included in a *bottom-up* coarse-grained model by Sanyal and Shell[143], where the non-bonded

potential is a the sum of a traditional pair potential and a local density dependent potential:

$$U_{\mathrm{nb}} = U_{\mathrm{pair}} + U_{\mathrm{LD}}, \qquad (2.51)$$

where

$$U_{\mathrm{LD}} = \sum_i f(\rho_i). \qquad (2.52)$$

Here, f is a function of the local density, ρ_i, around an interaction site i. This local density is recalculated at every simulation timestep by:

$$\rho_i = \sum_{j \neq i} \psi(r_{ij}), \qquad (2.53)$$

where ψ is an indicator function which determines how much site j contributes to the local density. A number of functional forms for ψ have been used; in general, the function should decay smoothly from 1 to 0 around a given cutoff, ensuring the continuity of forces with r_{ij}. The choice of cutoff for ψ is particularly important for multiphasic systems, where the cutoff should not exceed the width of the phase boundary. This allows the local density to be a clear indicator of which phase a given bead is part of.

An important aspect of this interaction type is that, while U_{LD} is not a simple pair potential, the computational cost of calculating forces scales like a simple pair potential. The local density contribution to the pair force on site i due to a second site j can be written as:

$$\mathbf{F}_i = -\left[\frac{df(\rho_i)}{d\rho} + \frac{df(\rho_j)}{d\rho} \right] \frac{d\psi(r_{ij})}{dr} \frac{\mathbf{r}_i - \mathbf{r}_j}{r_{ij}} \qquad (2.54)$$

In Sanyal and Shell's initial study, the relative entropy method was applied to parametrise an implicit solvent model for a model super-hydrophobic polymer. In this case, U_{LD} contains implicit information about the amount of water surrounding a polymer bead, and so allows the coarse-grained model to capture the hydrophobic effects which lead to the polymer's collapse in aqueous solution.[143]

Since then, a number of studies have explored the use of this interaction type for systems with phase boundaries. In 2017, DeLyser and Noid presented a method for parametrising a volume dependent potential using pressure matching, as described in Section 2.7.1, and then converting this to a local density potential.[144] The resulting

model, parametrised using a homogeneous liquid as reference, was transferable to a liquid-vapour interface. In 2018, Sanyal and Shell[145], and Jin and Voth[146], presented models for liquid-liquid mixtures, parametrised using the RE and UCG methods respectively, which described the phase boundary more accurately than simple pair potentials. These two studies both showed how a local density dependent model for a multicomponent system may be parametrised, and showed that it may be not be necessary to include LD potentials for all pairs of beads.

Chapter 3

Assessing the accuracy and transferability of common coarse-graining methods

3.1 Introduction

As detailed in Chapter 2, a huge number methods for parametrising coarse-grained models have been developed over the years. There have been a number of review articles which have described various methods and their applications.[9,147,148] However, there have only been a few studies which provide a detailed comparison between different coarse graining methods,[106] and particularly few which have critically compared the accuracy of *bottom-up* and *top-down* methodology on the same system and using the same criteria. Such a study would be of great benefit to the coarse-grained modelling community by providing insights into the strengths and weaknesses of different methods, and therefore informing the choice of coarse graining method.

Many publications which showcase coarse graining methods test their models on homogeneous single-component liquids, either small molecules with no internal structure after coarse graining,[20] or simple linear hydrocarbons and polymers.[21,100,137] While such systems provide a good initial test of the effectiveness of a coarse-grained model, to really get an idea of the effectiveness of a method, it is necessary to test it on more complex systems. The octane-benzene system has a number of properties which make it a good choice for this purpose. Firstly, the

two components have very different molecular geometries. This requires any coarse-graining method to be able to deal with cross-interactions, as well as with both flexible and rigid molecules. It also forms a miscible liquid mixture right across the composition range, which avoids complications associated with coarse-graining a system with an interface.

Of the *bottom-up* methods described in Chapter 2, iterative Boltzmann inversion (IBI)[21] and force matching[122] are perhaps the two most widely used. They each parametrise to match an entirely different quantity (radial distribution functions and forces, respectively); it would be useful to study to what extent this causes differences in the properties of the resulting models. The aim of this chapter will be to test the structural and thermodynamic representability, as well as the chemical and temperature transferability, of coarse-grained models produced using these two *bottom-up* methods, and to compare them to a *top-down* model parametrised using the SAFT-γ Mie framework.

The work within this chapter is included in the publication: "Assessing the transferability of common *top-down* and *bottom-up* coarse-grained molecular models for molecular mixtures."[149] It is reproduced from Ref. 138 with permission from the PCCP Owner Societies.

3.2 Computational details

3.2.1 Atomistic simulations

The IBI and MS-CG coarse-graining methods both require a reference atomistic model. The reference employed for the octane-benzene mixture was a modified version of the GAFF[68] force field. The modified force field, GAFF-LCFF, was parametrised by Boyd and Wilson to accurately capture the experimental densities and heats of vaporization of a range of molecules, including medium chain alkanes.[150,151] All simulations were performed using the GROMACS 4.6 package.[152] We used a leap-frog algorithm, with a time step of 2 fs. A Nosé-Hoover[79,80] thermostat was used to keep the temperature constant, at 298 K unless stated otherwise, and a Parrinello-Rahman[81] barostat was used to keep the pressure constant at 1 bar for constant NPT simulations. A particle mesh Ewald[73] (PME) was used to calculate electrostatic interactions, employing a short-range cutoff of 1.2 nm. A 1.2 nm cut-off

was used for Lennard-Jones interactions, and a long-range dispersion correction was applied. All bonds were constrained using the Linear Constraints Solver[71] (LINCS) algorithm within GROMACS.

A series of binary mixtures of octane and benzene were simulated with octane mole fractions (x_{oct}) of 0.0, 0.2, 0.3, 0.5, 0.7, 0.8 and 1.0. For each system, a total of 1600 molecules was simulated. Initially, for each system, a short constant NVT equilibration run was carried out, followed by equilibration at constant NPT to allow the density to reach equilibrium. The reference data for coarse-graining were then obtained from a 2 ns production run at constant NPT, at 298 K and 1 bar.

3.2.2 Parametrisation of coarse-grained models

Coarse-grained mapping and the parametrisation of the IBI and HFM models were carried out using the VOTCA-CSG package, versions 1.2.4[29,30] and 1.3.[113] Full descriptions of these methods can be found in Chapter 2.

Coarse-grained mapping.

Coarse-grained mapping was carried out by assigning two heavy atoms and their associated hydrogen atoms to a coarse-grained bead, and setting the interaction site for that bead as the centre of mass of those atoms.

$$\mathbf{R}_I = \sum_i^n \frac{\mathbf{r}_i m_i}{\sum_i^n m_i}, \tag{3.1}$$

where $\mathbf{R}_I$ is the positions of coarse-grained bead I, $\mathbf{r}_i$ is the position of an atom i which is included in the coarse-grained bead, m_i is the mass of bead i and n is the number of atoms which are mapped to the bead I.

The coarse-grained mapping schemes used are shown in Figure 3.1. For octane, a 4-site representation was used, with an outer bead type (A) and an inner bead type (B). This gives rise to 3 bonds, 2 angles and 1 dihedral. For benzene, a 3-site representation, with only one bead type (R) and three bonds, was used. This gives rise to a number of nonbonded interactions to parametrise: 3 for pure octane (A-A, A-B and B-B), 1 for pure benzene (R-R), and 2 additional cross interactions (A-R and B-R) for the octane/benzene mixtures. It should be noted that for the bottom-up models, no intramolecular 1-3 or 1-4 nonbonded interactions are included. However, they are included in the *top-down* models (see below).
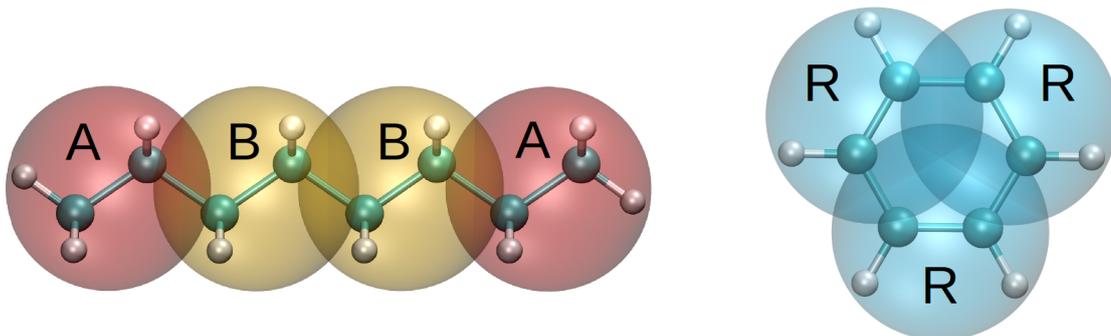
Figure 3.1: Coarse-grained mapping scheme for octane and benzene. Each coarse-grained bead represents two carbon atoms and their associated hydrogens.

Bonded interactions

Bonded potentials, $U(q)$, were obtained from the equilibrated reference atomistic simulations of pure octane and benzene by simple Boltzmann inversion:

$$U(q) = -k_{\mathrm{B}}T \ln P(q), \tag{3.2}$$

where q is a particular coarse-grained degree of freedom (e.g. a distance, angle or dihedral) and $P(q)$ is the normalised probability distribution of q. Final potentials were extrapolated into poorly sampled regions. For these simple systems, the assumption that bonded interactions are not correlated with other interactions is a good one. As the bond stretching potentials calculated for coarse-grained benzene were very steep, the LINCS algorithm was used instead to constrain bonds for that molecule, with the bond length taken as the minimum of the Boltzmann inverted potential at 0.2203 nm. The bonded interactions parametrised from pure octane and benzene were used in coarse-grained simulations for all concentrations. The bonded potentials and distributions are shown in Figures 3.2 and 3.3.

Iterative Boltzmann inversion (IBI).

For the (softer) nonbonded potentials, where multi-body effects are important, the potentials from Boltzmann inversion are unable to reproduce the structure of the reference system. For these interactions, an iterative Boltzmann inversion method was required.[21] Here, for each system, separate reference RDFs were calculated for each non-bonded interaction. The RDFs for pure octane and benzene were calculated from trajectories containing 1000 snapshots. For the mixtures, it was found that
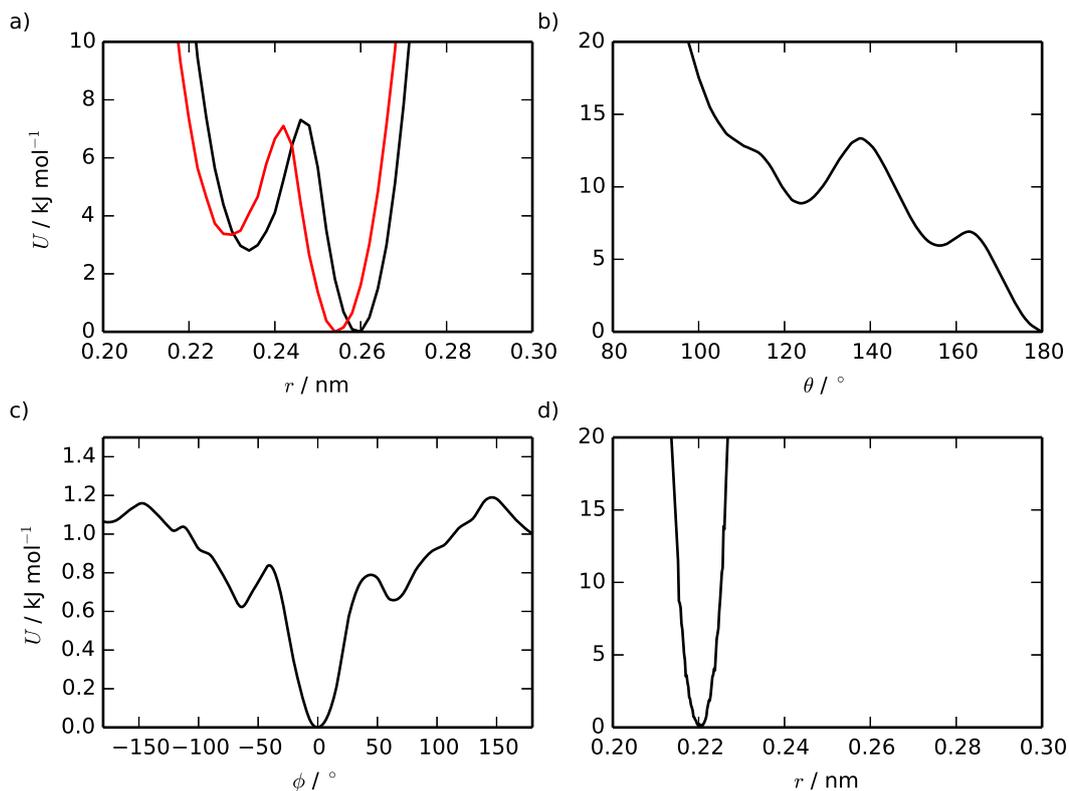
Figure 3.2: Coarse grained bonded potentials from Boltzmann inversion of the pure reference systems at 298 K: a) octane A-B (black) and B-B (red) bonds, b) octane angle, c) octane dihedral and d) benzene bond (not used in simulations).
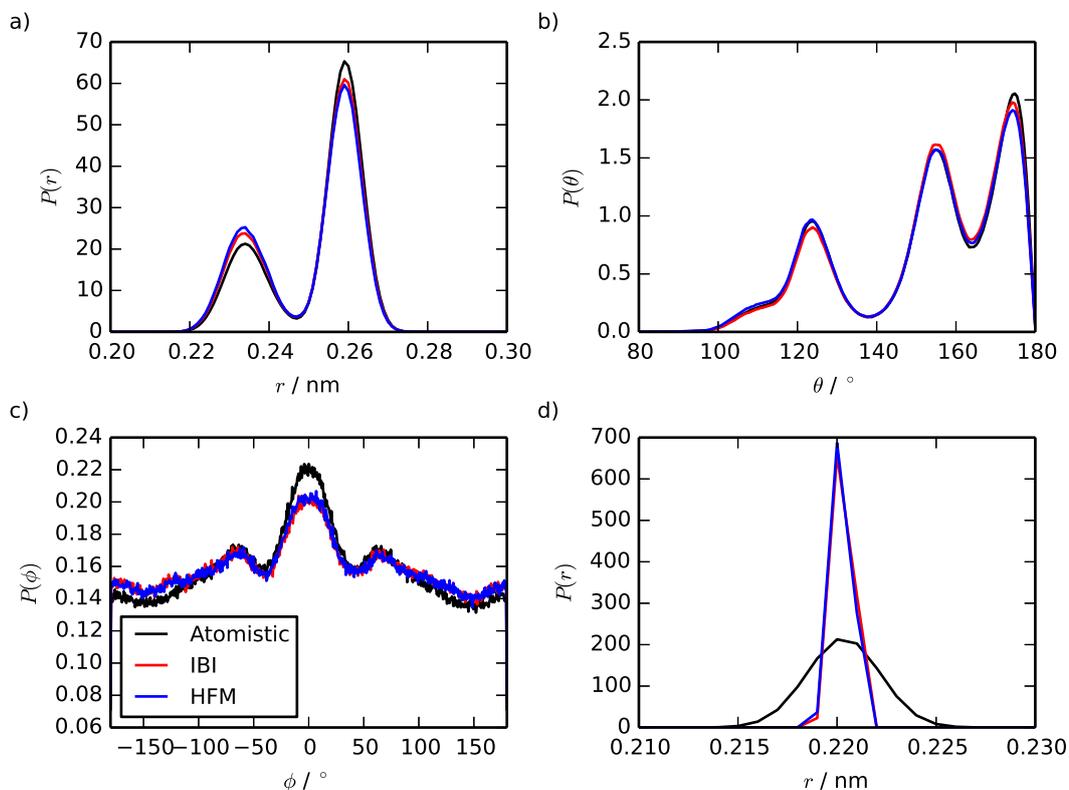

Figure 3.3: Bonded distributions for 50% octane system from atomistic, IBI and HFM simulations at 298 K. a) octane A-B bond, b) octane angle, c) octane dihedral and d) benzene bond distribution.

5000 snapshots were required for a smooth RDF, due to the decreased sampling of each interaction. These RDFs were used as targets for the IBI procedure.

At each stage of IBI, a coarse-grained simulation was run using a set of test potentials. The first stage used potentials from direct Boltzmann inversion. After each simulation, an update was applied to one of the potentials, according to Equation 3.3, with each interaction being updated in turn

$$U_{n+1}(r) = U_n(r) + \alpha k_B T \ln \frac{g_n(r)}{g_{target}(r)}, \tag{3.3}$$

where r is the inter-site distance and $g(r)$ is the radial distribution function. The potential updates from this are often too large, so α is a scaling factor which is chosen to speed up the convergence of the scheme. This procedure was carried out iteratively until the reference and test system RDFs matched to an acceptable degree. Once each interaction had been updated another simulation was run, and a linear pressure correction was applied to all of the potentials simultaneously:

$$\Delta U(r) = A\left(1 - \frac{r}{r_{cut}}\right), \tag{3.4}$$

where

$$A = \mathrm{sgn}(\Delta P)0.1k_B T \min(1, |f\Delta P|). \tag{3.5}$$

In these expressions, r_{cut} is the cut-off distance for the interaction, ΔP is the difference in pressure between the reference and the coarse-grained system and f is a scaling factor, which is chosen to prevent the pressure from oscillating around the desired value. Since the magnitude of the pressure correction depends on ΔP, the correction can be iteratively applied until the pressure of a coarse-grained system is correct.[21,29] The parameters used in the IBI procedure for each of the models are given in Table 3.1

This procedure was repeated until the coarse-grained and target RDFs matched, and the pressure of the system was within 5 bar of the target pressure of 1 bar. For the 70% and 80% octane systems, the R-R (benzene-benzene) interaction was very difficult to converge compared to the other interactions. For these systems, the procedure was adapted to allow several steps in which only the R-R potential was updated, with care taken to ensure that the other interactions and pressure remained converged. The non-bonded potentials parametrised using IBI are plotted

Table 3.1: Scaling factors used, and iterations required, for IBI optimisation and pressure correction for the HFM model. The numbers in parentheses are, respectively, the number of iterations after which only the R-R interaction and pressure were optimised, and the α used for the iterations after this.

x_{oct}	IBI			HFM	
	Iterations	α	f	Iterations	f
0.0	62	0.1	0.0005	20	0.0001
0.2	183	0.4	0.0005	8	0.0005
0.3	248	0.3	0.0005	10	0.0005
0.5	278	0.35	0.0005	12	0.0005
0.7	300 (297)	0.35 (0.15)	0.0005	12	0.0005
0.8	266 (271)	0.35 (0.2)	0.0005	10	0.0005
1.0	56	0.5	0.001	5	0.0005

in Figure 3.4

Hybrid force matching (HFM)

The hybrid force matching (HFM) method was also applied to the octane-benzene mixture. Coarse-grained bonded interactions were modelled using the Boltzmann inverted potentials (or constraints) described above. Non-bonded potentials were obtained using force matching. The reference non-bonded forces, excluding all intramolecular interactions, were calculated by passing through the reference trajectory and outputting only the forces resulting from intermolecular interactions. The trajectories obtained contained 1000 snapshots taken at intervals of 2 ps; force matching was carried out on blocks of 25 frames each, and the coarse-grained force functions were obtained by averaging over all of the blocks. The resulting force functions were integrated and extrapolated to low inter-site distances to give the HFM coarse-grained potentials for that system.

While the atomistic reference simulations were carried out in the constant NPT ensemble, it should be noted that, strictly speaking, the MS-CG method is only valid using a constant NVT reference, because the an reference simulation will include density fluctuations which could not be captured by the resulting effective pair potential. A version of the method consistent with the constant NpT ensemble, which includes a volume dependent part, was suggested by Das and Andersen.[128] This was later extended to a volume matching method, which can act as a pressure correction.[43,144] However, this requires including additional, volume-dependent, terms in
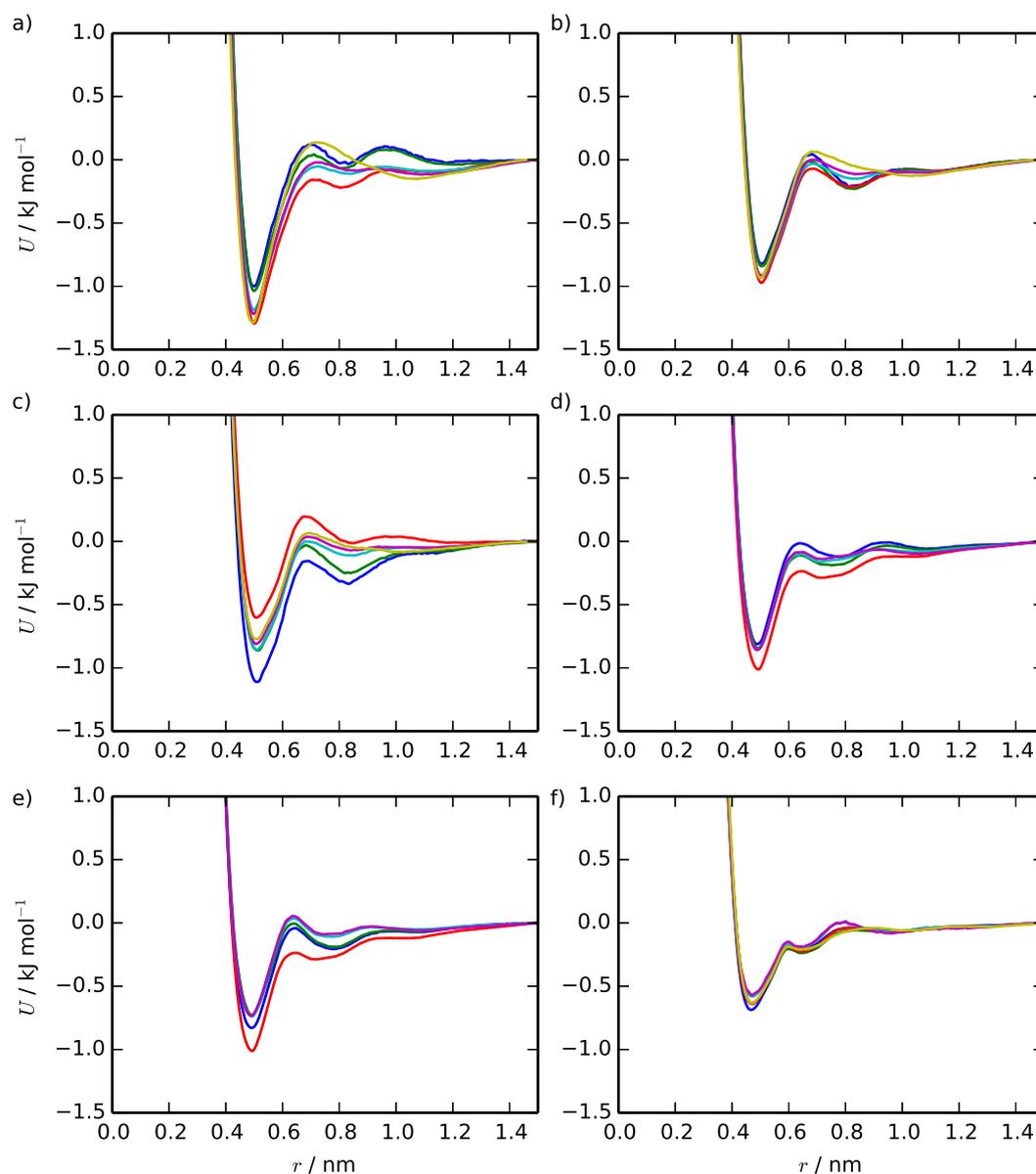
Figure 3.4: Non-bonded potentials parametrised using IBI for the a) A-A, b) A-B, c) B-B, d) A-R, e) B-R, f) R-R interactions. Each potential in each plot parametrised at a different octane concentration: 20% (blue), 30% (green), 50% (red), 70% (cyan), 80% octane (purple) and pure octane/benzene (yellow).

Figure 3.5: Non-bonded potentials parametrised using HFM for the a) A-A, b) A-B, c) B-B, d) A-R, e) B-R, f) R-R interactions. Each potential in each plot parametrised at a different octane concentration: 20% (blue), 30% (green), 50% (red), 70% (cyan), 80% octane (purple) and pure octane/benzene (yellow).
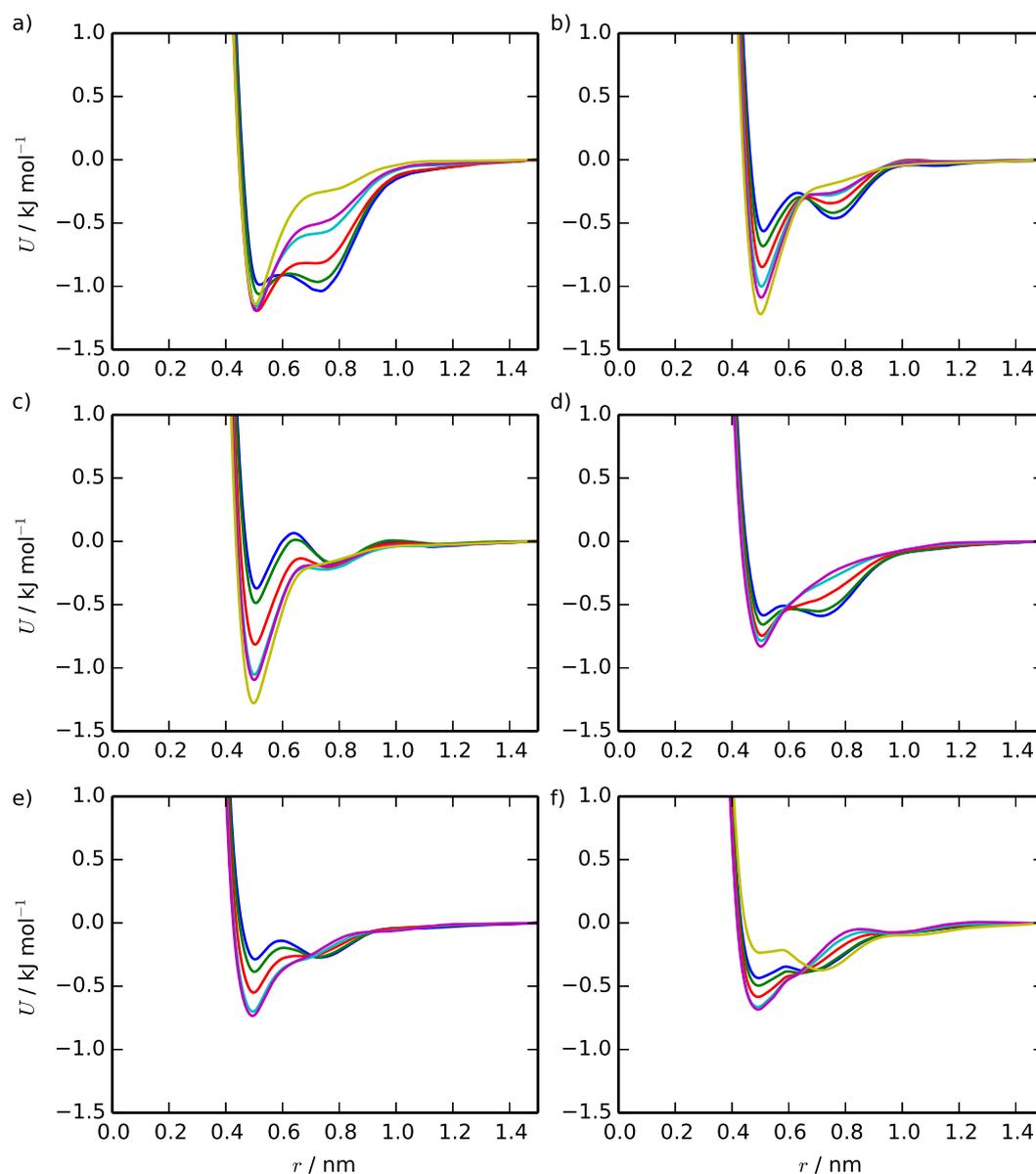
the pair potential. To avoid these additional terms, a pressure correction was applied to the force-matched potentials using the iterative ramp correction described in Equation 2.35 (noting that this does not strictly address the state-point dependence). This was applied to each system to give a pressure of close to 1 bar at the experimental density of the system. At each iteration of the pressure correction, a 250 ps simulation was run, and the pressure of the system was calculated based on the final 200 ps of the trajectory. Each of the non-bonded potentials was updated at each iteration, with a small scaling factor (see Table 3.1) to ensure that the pressure converged. The non-bonded potentials parametrised using HFM with pressure correction are given in Figure 3.5

SAFT-γ Mie model

The SAFT-γ Mie model was provided by Jos Tasche, and details of its parametrisation are given in this section.[149] All simulations using the model were carried out by the author.

Within the SAFT-γ Mie equation of state (EoS), a molecule is represented by a chain of tangentially connected beads interacting with the Mie potential:

$$U^{\mathrm{Mie}} = C\varepsilon\left[\left(\frac{\sigma}{r}\right)^{\lambda_{\mathrm{r}}} - \left(\frac{\sigma}{r}\right)^{\lambda_{\mathrm{a}}}\right], \tag{3.6}$$

where ε and σ are the well depth and the segment diameter, respectively. λ_{r} and λ_{a} are the repulsive and attractive exponents, for which $\lambda_{\mathrm{r}} = 12$ and $\lambda_{\mathrm{a}} = 6$ is the usual Lennard–Jones potential. Fitting of the parameters, ε, σ, λ_{a} and λ_{r} to reproduce the available experimental data (which could include temperature and concentration-dependent data) leads directly to a coarse-grained molecular model.[33,97]

The Mie parameters for the cross-interactions can be obtained from the following mixing rules:

$$\sigma_{ij} = \frac{\sigma_{ii} + \sigma_{jj}}{2}, \tag{3.7}$$

$$\lambda_{ij} - 3 = \sqrt{(\lambda_{ii} - 3)(\lambda_{jj} - 3)}, \tag{3.8}$$

$$\varepsilon_{ij} = (1 - k_{ij})\frac{\sqrt{\sigma_{ii}^3\sigma_{jj}^3}}{\sigma_{ij}^3}\sqrt{\varepsilon_{ii}\varepsilon_{jj}}. \tag{3.9}$$

The k_{ij} parameter allows the well-depth of the cross-interaction, ε_{ij}, to be adjusted (if necessary) to fit experimental mixture data.

Table 3.2: Mie potential parameters obtained for benzene and n-octane using the SAFT-γ Mie framework.

Interaction	σ/nm	ϵ/K	λ_r	λ_a
R–R	0.3490	258.28	11.58	6.0
A/B–A/B	0.3768	255.92	12.70	6.0
A/B–R	0.3629	256.53	12.13	6.0

SAFT-γ Mie includes no parameters for the intramolecular interactions apart from the Mie potential for non-bonded interactions and σ as a bond length. Therefore chains are effectively modelled as *semi-flexible* without angle or dihedrals interactions but including 1-3 and 1-4 nonbonded interactions.

The SAFT calculations were performed with our own implementation of the SAFT-γ Mie expressions, based on the recent publication by Papaioannou *et al.*[153] The SAFT-γ Mie model used the same mapping as the *bottom-up* models, with the one exception that the same interaction potentials were used for beads A and B. The Mie potential parameters of the benzene model were taken from Lafitte *et al.*,[98] who parametrised the Mie potential parameters to match the vapour pressure and liquid density of benzene over a range of temperatures (300–562 K). For octane, Mie potentials were developed using the corresponding state correlation by Mejia *et al.*,[154] in which Mie potential parameters are determined from just three experimental data points: the acentric factor, the critical temperature and the liquid density. The cross-interaction parameters were calculated using Equations 3.7-3.9, with $k_{ij} = 0$. The Mie potential parameters are given in Table 3.2. The interaction potentials were cut at short distances (around 0.5σ, where the value of U_Mie is at least $10000 \times kT$) and extrapolated quadratically to lower distances using the same procedure used for bottom-up potentials. This results in a slightly softer interaction potential than a *pure* Mie potential, which is helpful when carrying out free energy calculations. Because SAFT is based on tangentially bonded spheres, bond lengths were constrained at σ for MD simulations.

3.2.3   Coarse-grained simulations

The equations of motion were integrated using the leap-frog algorithm of GROMACS with a time step of 2 fs. A Nosé-Hoover thermostat was used to keep the temperature constant at 298 K, and, for constant pressure simulations, a Parrinello-

Rahman barostat was used to keep the pressure constant at 1 bar.

For *bottom-up* models, a 1.5 nm cut-off was used for non-bonded interactions. For each coarse-grained model, a 1 ns constant-NpT production run was carried out, with coordinates output every 0.2 ps. At temperatures other than 298 K, a 500 ps constant-NpT equilibration run was carried out before the production run. Starting structures were obtained from atomistic snapshots, after implementing the mapping shown in figure 3.1.

For the SAFT-γ Mie model, a 2.0 nm cut-off was used, and all coarse-grained bonds were constrained using LINCS. Since the molecular geometry (tangentially bonded spheres) in these models is slightly different to the atomistic structure mapped onto a coarse-grained representation, starting structures were constructed by randomly placing 1600 molecules into a box using the Gromacs *genbox* tool, which prevents insertion of particles within a given van der Waals radius of an existing particle. A steepest-descent energy minimisation and a 200 ps equilibration run were carried out. Production runs were then carried out as for the *bottom-up* models.

3.2.4   Free energy calculations

Free energy calculations were carried out using the Bennett Acceptance Ratio (BAR) method. A full description of this method can be found in Section 2.3.2.

The free energies of solvation were calculated for octane and benzene in solvents consisting of octane/benzene mixtures of various concentrations. The solvation free energy is the thermodynamic process of transferring a molecule from an ideal gas into solution, at 298 K and 1 bar. These were calculated by decoupling the intermolecular interactions of one molecule of octane or benzene with the surrounding solvent, while leaving all intramolecular interactions intact; the solvation free energy is the negative value of the free energy calculated from the decoupling process. All simulations using atomistic and pressure corrected coarse-grained models were run in the constant NPT ensemble, so the energies calculated are Gibbs free energies. Simulations of coarse-grained models without pressure correction were run in the constant NVT ensemble, at the equilibrium density of the atomistic system; the energies from these systems are therefore Helmholtz free energies.

For atomistic systems, Coulombic interactions were decoupled linearly, then van

der Waals interactions were decoupled using soft-core potentials with $\alpha = 0.5$ and $p = 1$. Nineteen intermediate λ values were chosen between 0 and 1, with a λ spacing of 0.05.

For coarse-grained systems, van der Waals interactions were decoupled linearly. Between $\lambda = 0$ and $\lambda = 0.9$, a λ spacing of 0.05 was used. In order to prevent singularities just as the interactions were about to disappear, a much higher concentration of λ points was used for $\lambda > 0.9$: typically a spacing of 0.01 up to $\lambda = 0.99$, 0.002 up to $\lambda = 0.998$ and then progressively smaller spacings down to 0.0001 between $\lambda = 0.9999$ and $\lambda = 1.0$. Errors were calculated for each λ spacing to check where additional λ points were required.

3.3 Results

3.3.1 Densities of single-state models

With a pressure correction applied, the density of each atomistic reference system was reproduced well by both *bottom-up* coarse-grained models parametrised for that concentration, with errors of less than 0.6%, as shown in Figure 3.6. The accuracy is highly dependent on applications of the pressure correction. Without pressure correction, constant-NVT simulations of the IBI and HFM models at the correct density give pressures of up to 2000 bar, leading to significant errors in density if the same potentials are run at constant NpT. After the application of pressure correction, all of the models gave pressures of 1 ± 3 bar. Within a force matching methodology, it is usually accepted that the pressure will not be predicted correctly and that these simulations should be run at constant volume, although pressure correction based on a volume-dependent term has been proposed.[43,128] However, Figure 3.6 shows that a simple linear pressure correction can successfully be applied to force matched potentials within the hybrid scheme employed here.

3.3.2 Densities from the SAFT-γ Mie model

The densities obtained from molecular dynamics simulations using the potentials developed via SAFT-γ Mie are shown in Figure 3.7. The results are good with respect to transferability across concentration and temperature ranges, and are indicative of the accuracy of SAFT-γ Mie model as a theory. Here the good quality
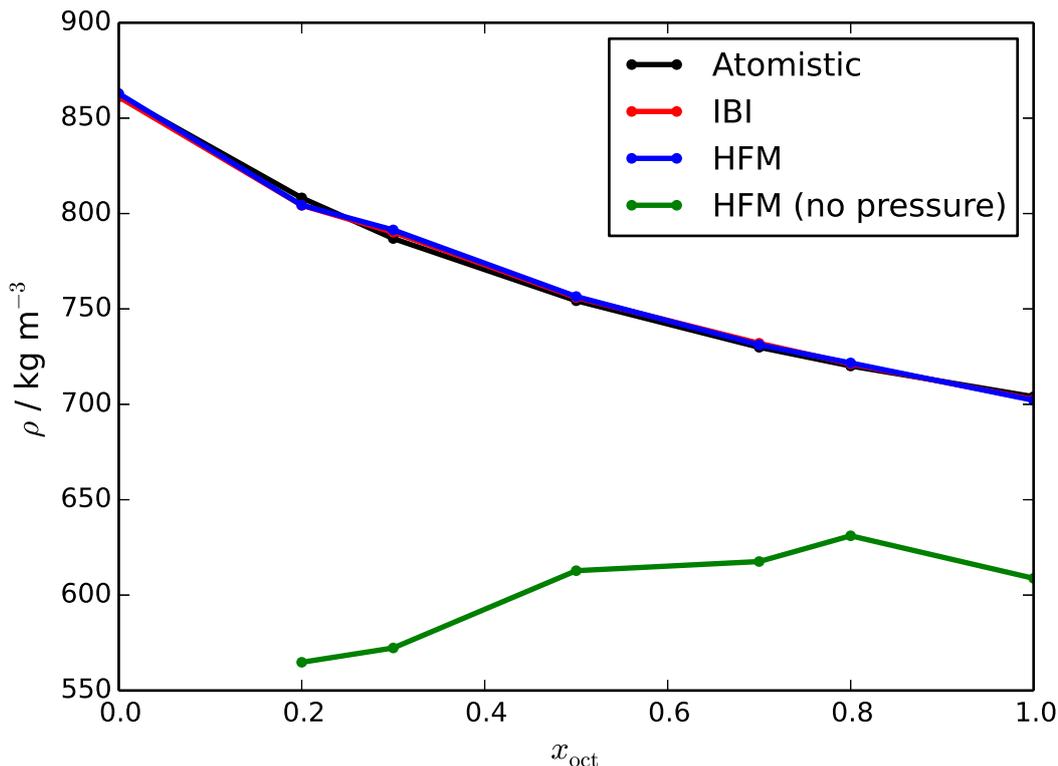
Figure 3.6: Comparison of densities calculated using atomistic and *bottom-up* coarse-grained models. The IBI and HFM models were parametrised and run at the specified concentration, 298 K and 1 bar. The green line shows results from HFM potentials without pressure correction.

predictions across changing concentrations rely on the reliability of mixing rules and the corresponding state correlation of Mejia *et al.*[154] SAFT is very versatile in terms of the nature of experimental data that can be fitted and so could be directly fitted to experimental densities to reproduce density data for mixtures. However, these data will not generally be available for many practical coarse-graining applications.

3.3.3 Structural accuracy of *bottom-up* and *top-down* models

Figure 3.8 gives site-site radial distribution functions for the $x_{oct} = 0.3$ octane system. These results are representative of the accuracy observed for other values of x_{oct}. Unsurprisingly, given the RDF fitting procedure used, the models created using IBI were able to match the radial distribution functions (RDFs) of the atomistic references almost exactly. Although not as good as IBI, HFM is also able to reproduce atomistic RDFs with good accuracy. Similar levels of accuracy have been seen for many systems using full force matching[20,122] (where the bonded interac-
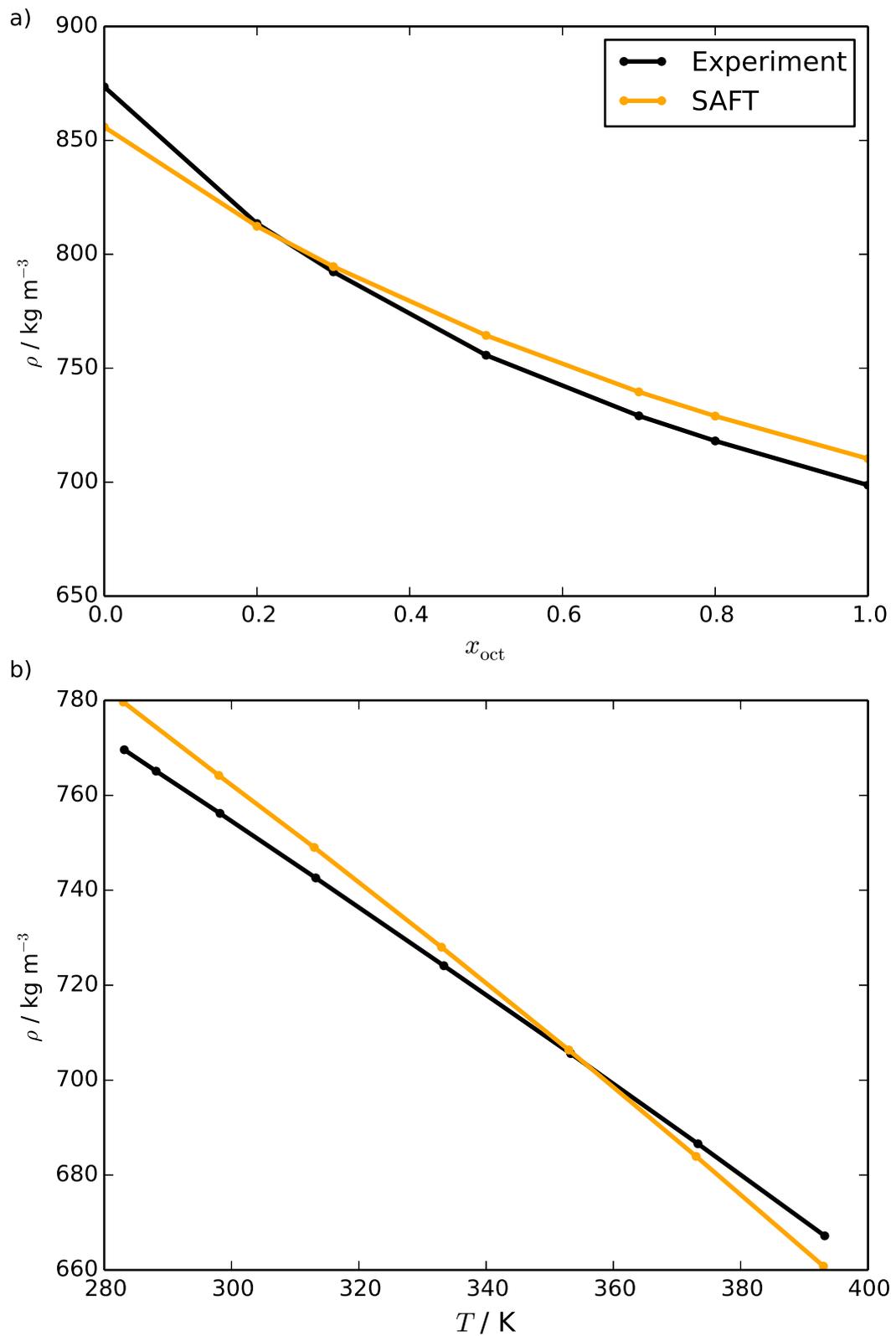
Figure 3.7: a) Concentration dependence of mixture densities at 1 bar and 298 K, b) temperature dependence of density at $x_{oct} = 0.5$ and 1 bar. Results obtained from simulations of SAFT-γ Mie derived potentials and experiment.[155,156]

tions are also determined using force matching), and for pure liquid hexane using HFM.[30] However, the results for octane/benzene mixtures confirm that the separation of bonded and non-bonded terms, which is the basis of HFM is also valid for multicomponent systems with a greater number of interactions.

Interestingly, the use of a pressure correction did not significantly affect the RDFs in either the pure components or the mixtures. The linear pressure correction has the greatest effect on the long-range tail of the potential, which has significant effects on thermodynamic properties such as pressure. While the absolute value of ΔU is larger at smaller r (see Equation 2.35), the attractive well and short-range repulsion are still preserved by the pressure correction. The local packing in liquids is most heavily influenced by short-range interactions; even hard sphere models which include only short-range repulsions are able to capture the structures of liquids.[157] Therefore, it is not surprising that the pressure-corrected models are still able to capture the RDFs.

The SAFT-γ Mie RDFs are particularly poor, with high first and second peaks of the RDFs appearing at distances that are too short. The more complex peak shapes present in the atomistic and *bottom-up* coarse-grained RDFs are also absent in the SAFT-γ Mie RDFs. This is the case for mixtures (as shown in Figure 3.8) and the pure components. Here, the constraint of using tangential spheres leads to longer bond lengths and shorter van der Waals radii, compared to the *bottom-up* models where adjacent beads are allowed to overlap. As will be discussed in detail in Chapter 5, this can cause significant issues in more structurally complex systems such as chromonic liquid crystals,[40] where an incorrect molecular structure can cause unphysical molecular aggregation. Such effects are perhaps less important at higher levels of coarse-graining, where local structure prediction is not expected, and it is more important to capture the average thermodynamic properties of the fluid.

The ability of the models to reproduce three-body structural correlations was also investigated; this is essential to the ability of a coarse-grained model to represent the underlying atomistic model well.[158] The quantity $G^3(r, r')$, given by:

$$G^3(r, r') = \left\langle \sum_i \sum_{j \neq i,k} \sum_{k \neq i,j} (\cos(\hat{\mathbf{u}}_{ij} \cdot \hat{\mathbf{u}}_{ik}) \delta(r_{ij} - r) \delta(r_{ik} - r') \right\rangle, \qquad (3.10)$$

where $\hat{\mathbf{u}}_{ij}$ is the unit vector between sites i and j and r_{ij} is the distance between the two sites, is a useful measure of how well a model describes three-body correlations,
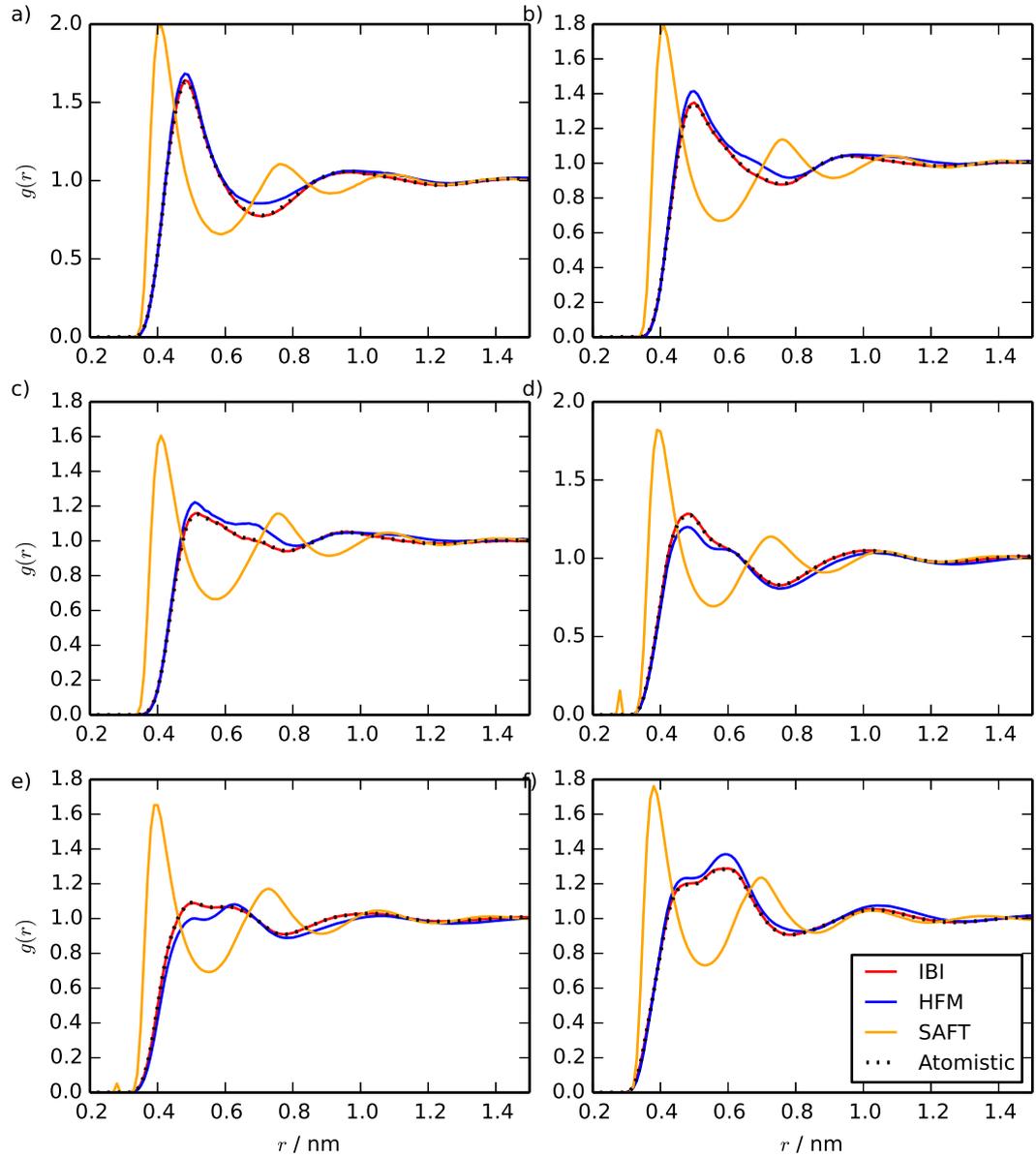
Figure 3.8: Intermolecular radial distribution functions calculated at 298 K, 1 bar and $x_{\mathrm{oct}} = 0.3$, from atomistic reference simulations, IBI and HFM models parametrised at the given concentration and the SAFT-γ Mie model, for the: a) A-A, b) A-B, c) B-B, d) A-R, e) B-R, f) R-R interactions.

as described by Noid *et al.*[158] $G^3(r, r')$ is a measure of the average angle between the vectors $\hat{\mathbf{u}}_{ij}$ and $\hat{\mathbf{u}}_{ik}$, as a function of the distances between the sites. This quantity was calculated for the RRR and RAA triplets from simulations of the $x_{\text{oct}} = 0.5$ mixture, using the IBI, HFM and SAFT-γ Mie models, and also for the atomistic trajectory mapped to a coarse-grained representation. These results are shown in Figure 3.9 for $r' = 0.5$ nm, which corresponds to the first peak in the RDF. The IBI and HFM models both compare well to the atomistic reference in both cases. However, the HFM results are slightly less accurate; this is expected, given that HFM performs slightly worse than IBI in reproducing RDFs. These results are encouraging given that the target of both methods is to match two-body correlations. The negative peaks at around $r = 0.6$ nm indicate the exclusion of particle j from the immediate surroundings of particle k, while the positive peaks at around $r = 0.9$ nm represent the strong probability of finding particle j at a distance corresponding to the second peak on the RDF. The SAFT-γ Mie model does not represent three-body correlations well, which is not surprising given its poor representation of the site-site RDFs.

3.3.4 Transferability

The potentials obtained from both *bottom-up* methods differ somewhat when obtained from different reference system concentrations. This can be clearly seen in Figures 3.4 and 3.5. Here, particularly for the HFM case, there is considerable variation in the overall shape of the effective pair potential across the concentration range. The HFM potentials show a double well structure, which can be explained by the fact that these are two-body potentials which implicitly include three-body correlations in their parametrisation. Each R bead has two bonded neighbours, while B beads have one and A beads have none. Therefore, at higher benzene concentrations the influence of three-body correlations where two of the beads are bonded to each other will increase, leading to the growth of the second peak at higher benzene concentrations. The A-A potential varies most with concentration and there is a smaller change for other pair potentials, particularly the R-R potential.

This immediately suggests that the models will not be particularly transferable, and this turns out to be the case. Figure 3.10 shows a comparison with the A-A partial RDF obtained from atomistic simulations at $x_{\text{oct}} = 0.2$, using HFM models
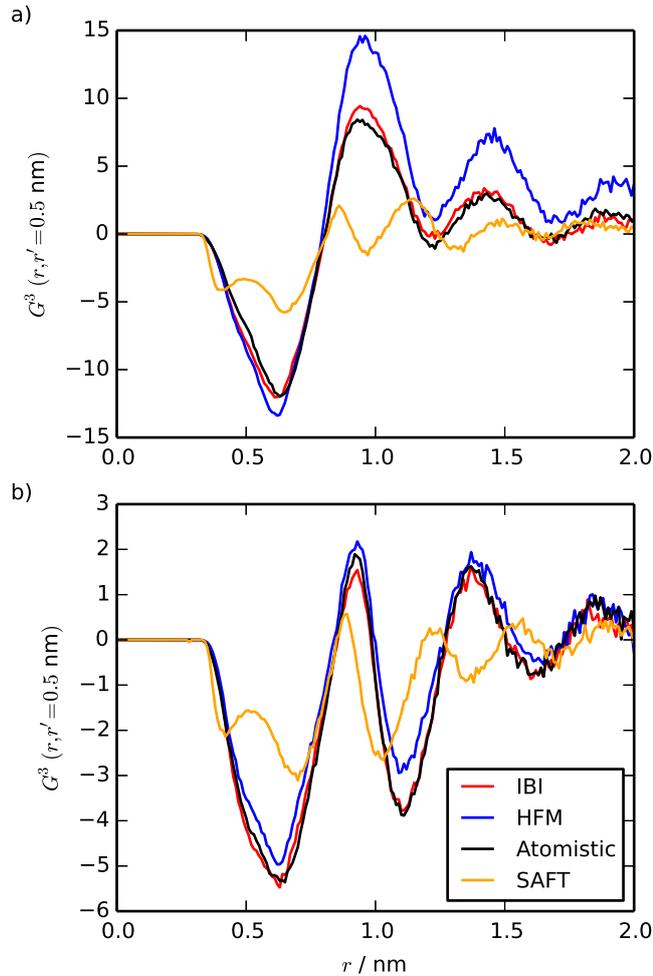
Figure 3.9: Three-body correlations calculated for the $x_{oct} = 0.5$ system, for the a) RRR and b) RAA triplets, calculated using atomistic, IBI, HFM and SAFT-γ Mie models. Note that intramolecular pairs are excluded from these calculations.

parametrised at $x_{oct} = 0.2$ and $x_{oct} = 0.5$. The HFM results are noticeably worse for the model parametrised at the higher concentration. Moreover, the HFM and IBI potentials parametrised at $x_{oct} = 0.5$, are unable to reproduce the correct density for any other concentration, as shown in Figure 3.11a.

Both single-state *bottom-up* models show poor temperature transferability when compared to the atomistic reference system. For pure octane, the densities of the coarse-grained systems diverged from the atomistic density on increasing or decreasing the temperature from 298 K, as shown in Figure 3.11b. The poor temperature transferability of IBI potentials for simple liquids has been shown in the past.[106] These results show that HFM suffers from exactly the same problem. The similarity of the results from the two methods likely comes from the fact that the same method was used for pressure correction in each case. The *ad hoc* nature of the linear pressure correction means that there is no guarantee of transferability to different state
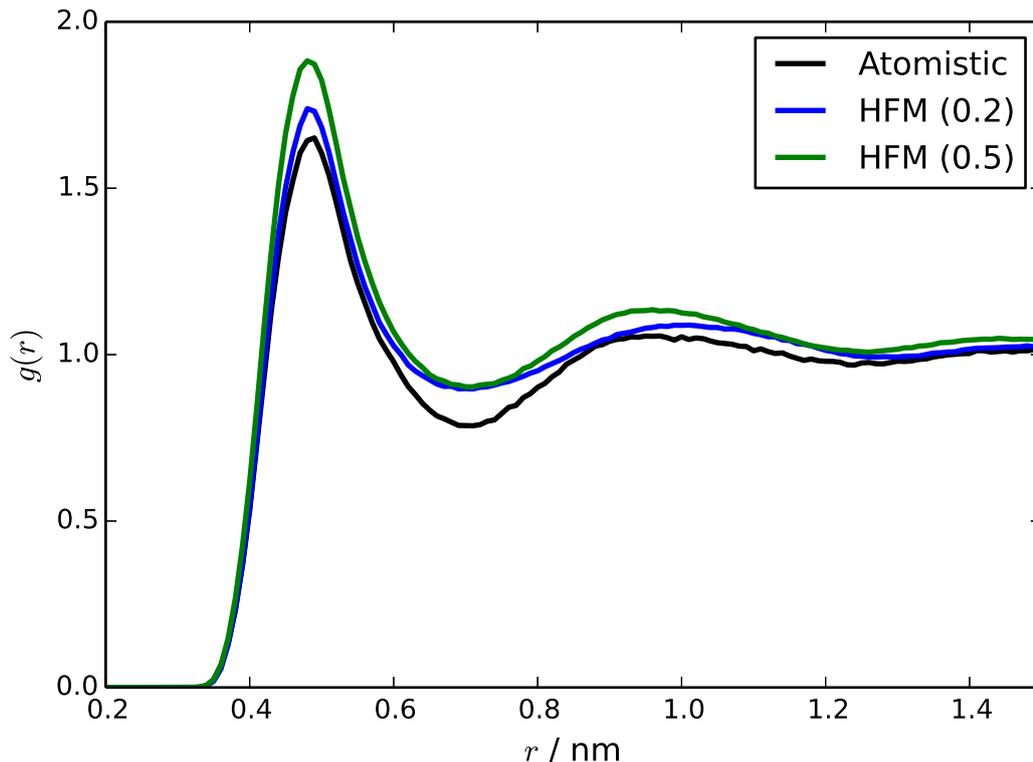
Figure 3.10: Radial distribution function for the A-A interaction for simulations of $x_{\mathrm{oct}} = 0.2$ at 298 K and 1 bar: atomistic reference, HFM parametrised for $x_{\mathrm{oct}} = 0.2$, and HFM parametrised for $x_{\mathrm{oct}} = 0.5$

points. As discussed above, this is in contrast to the excellent thermal expansion behaviour seen for simple Mie potentials fitted via the SAFT-γ Mie EoS (Figure 3.7). It should also be noted that the atomistic model does not show perfect agreement with experiment. The purpose of this work is to examine the connection between *bottom-up* models and the underlying reference, but this highlights that there are also issues with the transferability of atomistic models which must be considered if *bottom-up* models are to be used predictively.

3.3.5 Solvation free energies

The solvation free energies of octane and benzene are consistently overestimated by pressure-corrected HFM for all of the systems studied. However, as can be seen in Figure 3.12, there is no systematic relationship between the atomistic and HFM solvation free energies, with the difference between the two varying significantly over the concentration range.

HFM without pressure correction performed extremely poorly, as shown in Figure 3.12. All of the solvation free energies from these models are very far off the atomistic
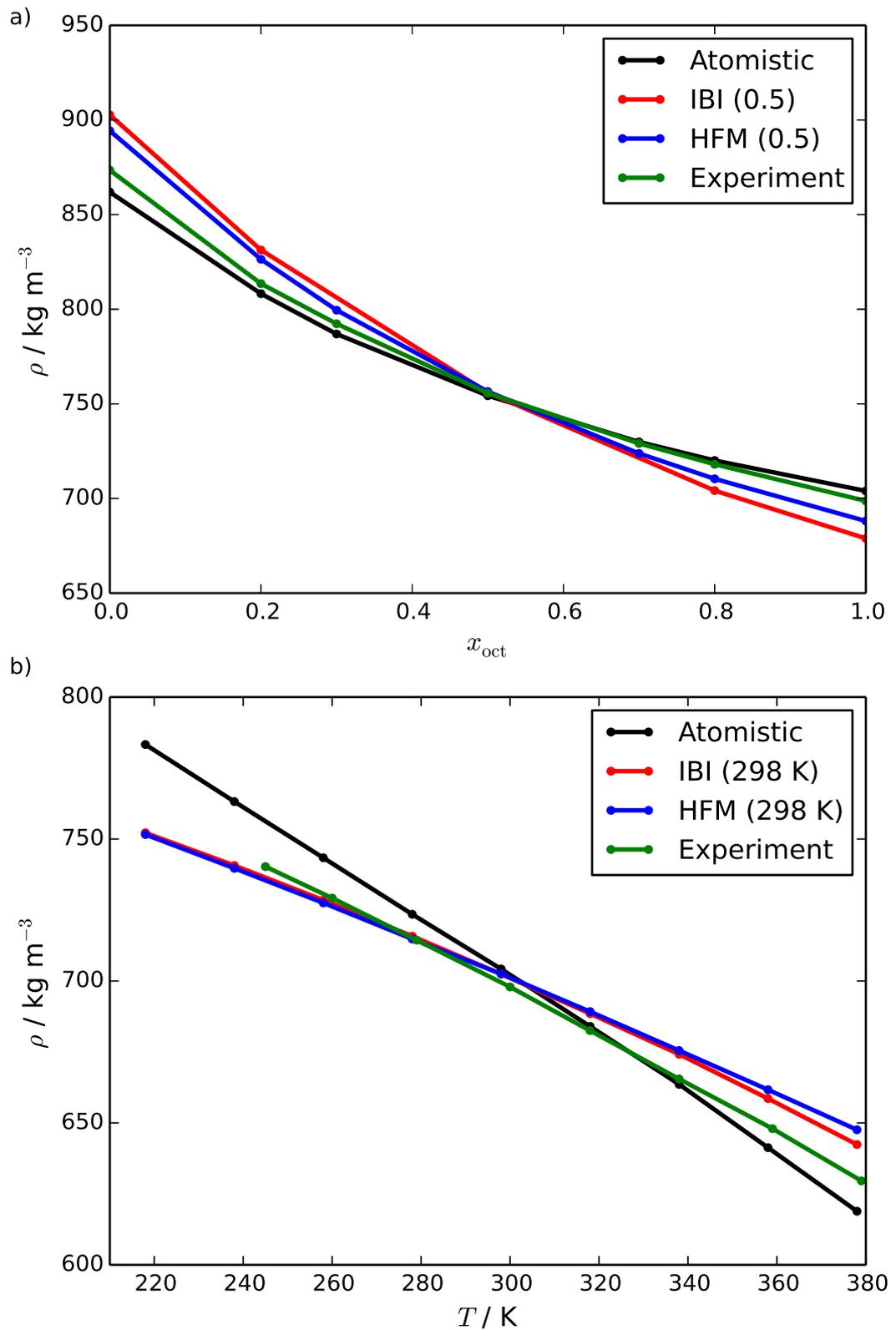
Figure 3.11: Transferability of densities calculated by the IBI and HFM models: a) parametrised at $x_{oct} = 0.5$, 298 K and 1 bar, and b) parametrised for pure octane at 298 K and 1 bar, and run at the specified temperature. Experimental data are also shown.[155,159]

values; most notably, benzene is predicted to be insoluble in pure benzene.

The difference between the pressure corrected and non-pressure corrected HFM models highlights the problem of coarse-grained models properly representing thermodynamic properties. It is known that the state-dependence of coarse-grained potentials affects the ability of coarse-grained models to reproduce the pressure, since the virial formula does not take into account the volume dependence of the potentials.[39] This has ramifications for the calculation of free energies for coarse-grained models. The incorrect representation of the pressure of the coarse-grained model leads to an incorrect representation of free energy changes, since $(\partial A/\partial V)_T = -P$; this explains the poor performance of the non-pressure corrected models, where P is 2–3 orders of magnitude too large. The application of pressure correction also improves the accuracy of free energy changes at a single state-point. However, because the linear pressure correction does not properly address the volume dependence of the coarse-grained potentials, this is unlikely to be transferable to other state-points.

For all systems, the IBI model predicts the solvation free energy of both benzene and octane more accurately than the HFM model. This could be attributed to the way in which the pressure correction is carried out for each method. In the IBI procedure, the pressure correction is carried out in conjunction with structure matching, and so the resulting model is guaranteed to reproduce both the structure and the pressure. This is only possible due to the iterative nature of IBI; it is trivial to add extra steps to the structure matching procedure to introduce the extra constraints. For force matching, on the other hand, the pressure correction is applied after the force matching procedure, and so there is less of a guarantee that both the forces and the pressure will be correct. Matching the pressure at the same time as the forces would require either the introduction of another set of equations or terms to the force matching equations (as shown, for example, by Das and Andersen)[128] or adapting the iterative force matching method suggested by Lu *et al.* to self-consistently match the pressure.[160]

The SAFT-γ Mie model is also able to predict solvation free energies, and their trend with respect to concentration, with fairly good accuracy. The SAFT equation of state is designed to accurately calculate free energies, so it is encouraging that a coarse-grained force field based on it performs reasonably well on such quantities.
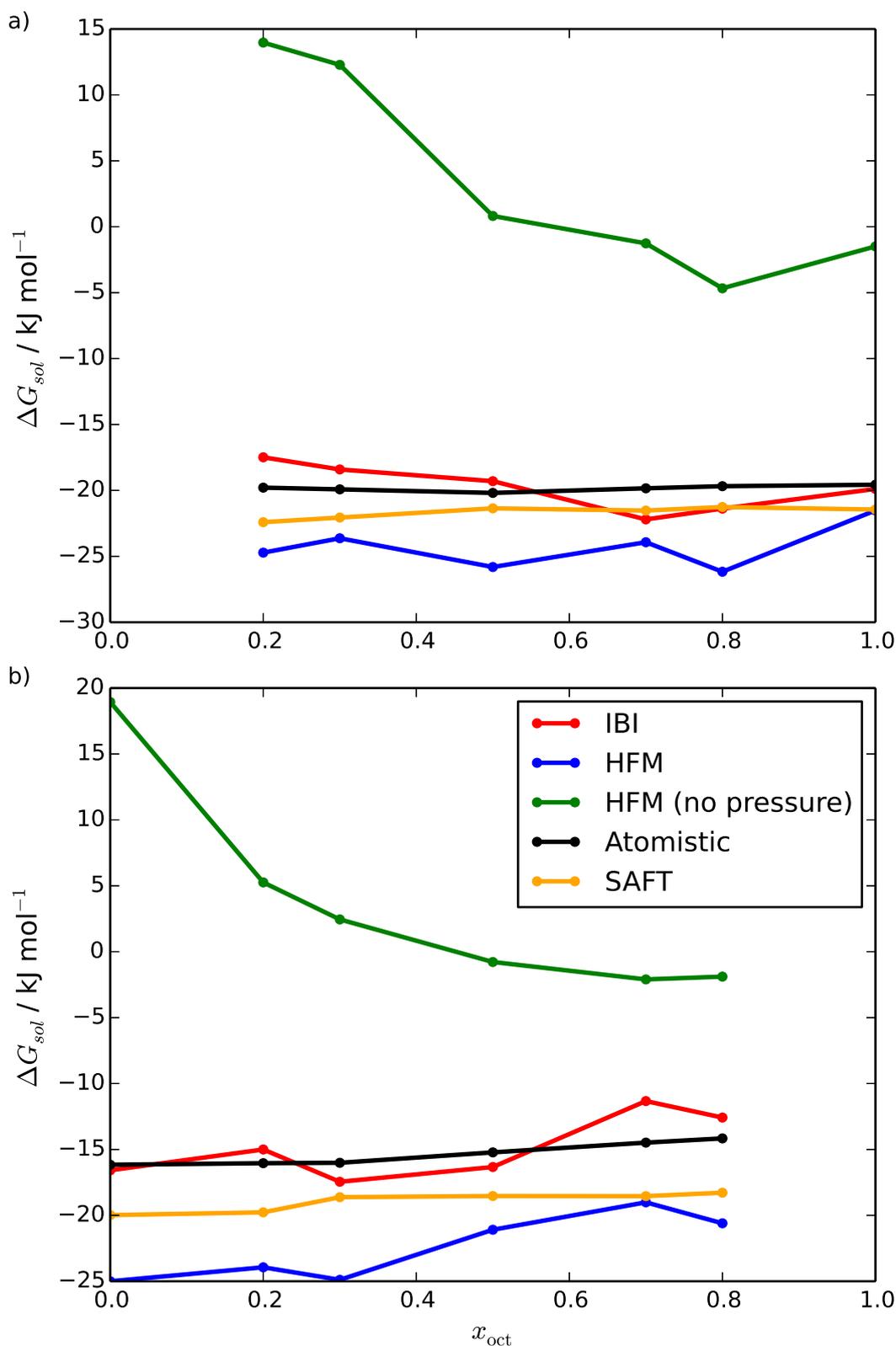
Figure 3.12: Free energies of solvation of a) octane and b) benzene as a function of solvent concentration, for atomistic, IBI, MS-IBI, HFM with pressure correction, HFM without pressure correction, and SAFT-γ Mie simulations. (The IBI and HFM *bottom-up* potentials are parametrised for each independent state point.) At this scale, the calculated errors in the free energy are smaller than the symbols used, and are generally on the order of $\pm 0.1 - 0.2$ kJ mol^{-1}.

3.4 Discussion

The Henderson uniqueness theorem[101] states that a pair potential which is able to reproduce the RDF of a simple liquid will be unique. This unique potential is the target of IBI and other structure-based coarse-graining methods. However, it is known from previous studies of IBI potentials that, even if a coarse-grained model exactly matches the RDFs from an atomistic system, this provides no guarantees that the thermodynamic properties will be reasonable.[108] Moreover, as seen in some previous work[106] potentials generated with the IBI method tend to have limited transferability, both in terms of concentration and temperature (though the degree of transferability is clearly system-dependent). The form of the effective pair potential differs significantly with concentration. This seems to be exaggerated by the presence of physically undesirable oscillations in the potential as a function of distance; so that the concentration-dependent pair potential changes even at large separations.

The HFM used in this work does remarkably well in terms of prediction of local structure, noting that it is not fitted to reproduce this. The simple linear pressure correction term that was introduced does not noticeably change local structure but does allow the pressure to be corrected by small changes to the effective pair potential, and thereby allows HFM to work under conditions of constant NPT. Unfortunately, for the octane-benzene mixture the effective pair potentials produced with HFM change even more than IBI potentials with concentration. For both methods, this translates to poor transferability. Chemically transferable models, whether between different concentrations or between similar chemical systems, are crucial, as they reduce the time otherwise spent parametrising models for each state point. Models developed by the conditional reversible work method have been shown to be transferable between different alkane chain lengths.[137] However, as the authors note, this method is not necessarily applicable to more complex systems. The extended ensemble approach of Mullinax and Noid,[132] in which the coarse-grained pair potentials are parametrised for an ensemble of systems simultaneously, allows for concentration transferable models, although at the cost of some representability compared to models parametrised for a single system. This approach was extended by Dunn and Noid[140] to include volume potentials, yielding a set of related models which, together, can give both temperature and concentration transferability.

The linear pressure correction which has been applied to the IBI and HFM models is very effective at correcting the coarse-grained pressure for a given state-point, without requiring the inclusion of any other parameters in the coarse-grained force field. However, it does not do so in a transferable way; in fact, it has been shown by Wang *et al.*[107] that correcting the pressure in this way means that the compressibility of the model will no longer be correct. The reason for this is that the linear correction does not take into account the underlying issue with the representation of pressure in coarse-grained models; the effective pair potentials generated by *bottom-up* coarse graining methods are volume-dependent, and so the virial formula for calculating pressure is not valid.[39] The volume potentials first suggested by Das and Anderson[128], on the other hand, are able to simultaneously represent pressure and compressibility,[139] because they explicitly take into account the volume-dependence of coarse-grained potentials. This approach has recently been used to construct a temperature transferable coarse-grained model,[141] although this is, of course, at the expense of adding extra parameters to the model, and therefore slowing down the coarse-grained simulations.

The SAFT-γ Mie EoS is an intriguing approach to *top-down* coarse-graining. As an EoS, SAFT-γ Mie is remarkably accurate when fitted to produce a wide range of experimental data. The possibility of being able to fit Mie potentials to reproduce, for example, vapour-liquid coexistence data over a range of temperatures guarantees that optimised effective pair potentials will be produced that are transferable in terms of thermodynamic free energies. However, there are problems with SAFT-γ Mie potentials in terms of local structure prediction. The origin of these problems arises mainly from the use of tangential spheres. This leads to a local structure that is fundamentally different to that seen in *bottom-up* coarse-grained models, and which does not directly map onto an atomistic representation of the system. Chapter 5 contains further discussion on the problems this can cause in a coarse-grained model. Within the SAFT framework, the use of shape factors[161] has recently been suggested as a method for creating structurally representative coarse-grained models.[162] However, this would be difficult to implement for systems with multiple bead types. Alternative *top-down* approaches may also be effective; for example, a recent study by An *et al.* presents a coarse-grained model for alkane chains parametrised using a particle swarm optimisation strategy. This model is thermodynamically transferable, and exhibits reasonably good structural accuracy.[163]

It is worth noting that in the current work, standard combining rules were used for benzene-octane cross-interactions. This is superficially very attractive because the parametrisation task for systems with many beads becomes far less onerous when all the cross-interaction terms do not need to be fitted. Unfortunately, for many systems the combining rules do not work as well.[164] However, simple improvements to the fitting of thermodynamic data within SAFT-γ Mie can be made by allowing the k_{ij} parameter to vary from zero.[165]

A further interesting comparison between the *bottom-up* and *top-down* approaches is the shape of the potentials they produce. As shown in Figure 3.8, the *bottom-up* methods produce a wide variety of potential shapes, which do not correspond to a common functional form. In principle, the use of numerical or spline potentials allows for greater flexibility when trying to match the reference properties. In practice, however, it can result in overfitting the model to a particular state point. The reason for the irregular bumps present in many of these potentials is that they are effective pair potentials, in which any multi-body contributions resulting from the coarse graining of the system are included only in an averaged way. This will only be applicable to the state point at which the model was parametrised since the local environment of a given bead will vary depending on the temperature and concentration. SAFT-γ Mie models, on the other hand, are based entirely on Mie potentials, with a distinct functional form described in Equation 3.6. Because of this well-defined shape, there is less danger of overfitting; therefore, they are likely to have much better transferability than IBI or HFM models.

Finally, given the difficulties in achieving the two key chemically desirable attributes (i.e. local structure and thermodynamics) within a single coarse-grained model, it is appropriate to ask the question as to whether it is possible to have this level of representability, together with transferability to other state points. We would argue that, in principle, it is; however in practice this may require some degree of compromise with computational cost, either in terms of a model which is more expensive to simulate, or a more expensive parametrisation procedure. In one sense, all classical models can be thought of as existing on a continuum scale of complexity: atomistic models are largely successful because typical force fields have achieved a reasonably high degree of representability and transferability. Yet we know that if, for example, a TIP4P water molecule is transferred from bulk water into the gas phase (an environment with a different density) or transferred to the surface of a

protein (an environment with a dielectric constant) then the TIP4P water is not as good a model as it is for bulk water. The way around this transferability problem is partly being tackled by polarizable models, such as AMOEBA;[166] i.e. models that can respond to environmental changes. For coarse-grained models it is desirable to do the same, without the obvious but prohibitively expensive addition of full three-body forces to a method such as HFM. Recent work provides some encouragement that this can be achieved due to the improvements of the transferability in DPD models via addition of a local density-dependent term (MDPD).[167] Crucially, the computational cost of this local density-dependent term scales with system size like a simple pair potential. A similar approach has been applied to transferable *bottom-up* coarse-grained models of a range of systems, including liquid-liquid and liquid-vapour equilibria.[143–146,168]

3.5   Conclusions

In this chapter, the accuracy and transferability of coarse-grained models parametrised using *top-down* (SAFT-γ Mie) and *bottom-up* (IBI, HFM) methods has been compared. Both approaches were found to have distinct advantages and disadvantages.

In terms of structural accuracy, IBI was found to be superior to both HFM and SAFT-γ Mie. HFM models were able to reproduce the structure of the systems studied fairly well, while SAFT-γ Mie models were not able to give an accurate picture of the local structure of any of the systems. Unfortunately, even for this simple system, the effective pair potentials derived using IBI and HFM vary considerably between state points, limiting transferability.

In terms of thermodynamics, although neither *bottom-up* method was able to reproduce exactly the solvation free energies of the atomistic system, the accuracy was good enough that the correct phase behaviour was observed. Pressure correction was found to be crucial for the reproduction of solvation free energies, confirming that matching structure or forces alone does not guarantee thermodynamic consistency when moving to a coarse-grained representation of a system. The model derived from SAFT-γ Mie was found to be thermodynamically transferable across the entire concentration range.

The main concern with *bottom-up* coarse graining appears to be transferability. All of the IBI and HFM models studied were parametrised for a specific state point,

and were less accurate when applied to a different state point. This can partially be attributed to the coarse graining process. Removing degrees of freedom will always reduce the entropy of the system, so in order to match the free energy, the enthalpy/entropy balance must be shifted. Since the entropic contribution to the free energy is temperature dependent (including the contribution to the entropy from degrees of freedom that have been removed), the overall free energy of the system will not scale correctly with temperature, and the force field will not be completely transferable. SAFT -γ Mie models tend to be considerably better in terms of thermodynamic transferability, due to fitting over a range of state points. Unfortunately, this is at the expense of very poor structural accuracy. However, as discussed, it may be possible to improve this transferability. Possible approaches to achieving this will be discussed in detail in the next chapter.

Chapter 4

Improving the transferability of *bottom-up* coarse-grained models

4.1 Introduction

One of the key shortcomings of the *bottom-up* models described in Chapter 3 is their poor transferability across both concentrations and temperatures, particularly in the context of reproducing densities. Poor transferability significantly reduces the usefulness of a coarse-grained model, as it necessitates reparametrising the interactions for different state points. Investigating methods for improving the transferability of *bottom-up* coarse-grained models will be the aim of this chapter.

The MS-IBI method has previously been shown to improve the chemical transferability of coarse-grained models.[117] This includes relatively complex biological systems with multiple bead types, although for such systems the structural representability is not consistently good.[118] However, the method has so far only been employed using the radial distribution functions of a range of states as targets. Also, the thermodynamic transferability of models parametrised with MS-IBI has not been well studied. With the findings of the previous chapter on the relationship between accurate pressures and free energies in mind, it is likely that accurate pressures across a range of state-points will be necessary for thermodynamic transferability. Moore *et al.* describe the use of both *NVT* and *NPT* states to account for the pressure density relationship.[118] However, it also seems plausible that a linear pressure correction[21] may be applied within the MS-IBI framework to match the pressures of multiple states; this would have the advantage of requiring fewer

reference states, reducing the complexity of the parametrisation process. The first aim of this chapter will be to investigate to what extent MS-IBI is able to yield transferable coarse-grained models in terms of both structure and thermodynamics, across the concentration and temperature ranges, by applying the same tests which were applied to the IBI and HFM models in Chapter 3.

One feature shared by the majority of *top-down* coarse-grained models which may partially explain their good transferability is that they are usually composed of pair potentials with defined functional forms rather than numerical potentials.[32,33] This has the advantage of preventing the over-fitting of the model to a particular state-point. However, *top-down* models often have rather poor structural representability when compared to *bottom-up* models, which can limit their usefulness. It will therefore be investigated whether requiring a defined functional form for the pair potentials (in this case a Morse potential) using a *bottom-up* structure-based parametrisation can result in a more transferable coarse-grained model, which still has good structural accuracy. This also has the advantage of simplifying the optimisation of the coarse-grained model by reducing the number of parameters to be fitted.[169]

Finally, the possibility of improving transferability using local-density dependent potentials in coarse-grained models will be discussed, and the preliminary steps which have been taken towards this aim will be described.

Some of the work within this chapter is included in the publication: "Assessing the transferability of common *top-down* and *bottom-up* coarse-grained molecular models for molecular mixtures", specifically the results described in Sections 4.3 and 4.4.1. It is reproduced from Ref. 138 with permission from the PCCP Owner Societies.

4.2 Methodology

4.2.1 Simulations

Atomistic reference simulations and coarse-grained simulations were carried out using the parameters and methods described in Chapter 3. The specific simulation conditions used will be described in the relevant sections.

Table 4.1: Scaling factors used, and iterations required, for the MS-IBI optimisation. The numbers in parentheses are the number of MS-IBI iterations which were carried out, after which pressure correction was applied. For the MS-3c and MS-4c models the same η value was used for each reference system, for the MS-2t and MS-lv models, a different η value was used for each reference.

Model	Iterations	η	f
MS-3c	319 (300)	0.2	0.0004
MS-4c	310 (295)	0.2	0.0004
MS-2t	248 (200)	0.7 (238 K), 0.5 (378 K)	0.0001
MS-lv	96	0.5 (liquid), 0.1 (vapour)	-

4.2.2 Parametrisation of multi-state IBI models

Coarse-grained models were parametrised using a version of the VOTCA 1.3[113] package modified to carry out MS-IBI. This used the existing iterative framework from the VOTCA package, but with the structure of the code modified by myself to allow for multiple simulations at each iteration, and scripts added to calculate the potential update from these simulations according to:

$$\Delta U = \frac{1}{N} \sum_{i=1}^{N} \eta_i \Delta U_i. \tag{4.1}$$

In this implementation, ΔU_i may be either the IBI update calculated for reference system i, or a pressure correction update.

The initial guesses for non-bonded potentials were obtained by taking the average of the Boltzmann inverted potentials of the reference systems included in the fits. At each MS-IBI step, a simulation was run for each of the reference systems, and the RDFs calculated. A single potential was then updated using Equation 4.1, using the IBI updates for each reference system scaled by a factor of η. Each potential was updated in turn. Once the RDFs had converged as much as possible, as determined by visual inspection, additional iterations were carried out in which only pressure correction was applied; here, η was set to 1 for each reference system and the potential updates were scaled by a factor of f.

Several MS-IBI models were considered. The MS-3c and MS-4c models used three and four octane/benzene mixtures of different concentrations as references. The MS-2t model used octane at two different temperatures. The MS-lv used octane references in the liquid and vapour states. The scaling factors used in the parametrisation of these models are shown in Table 4.1. The coarse-grained poten-
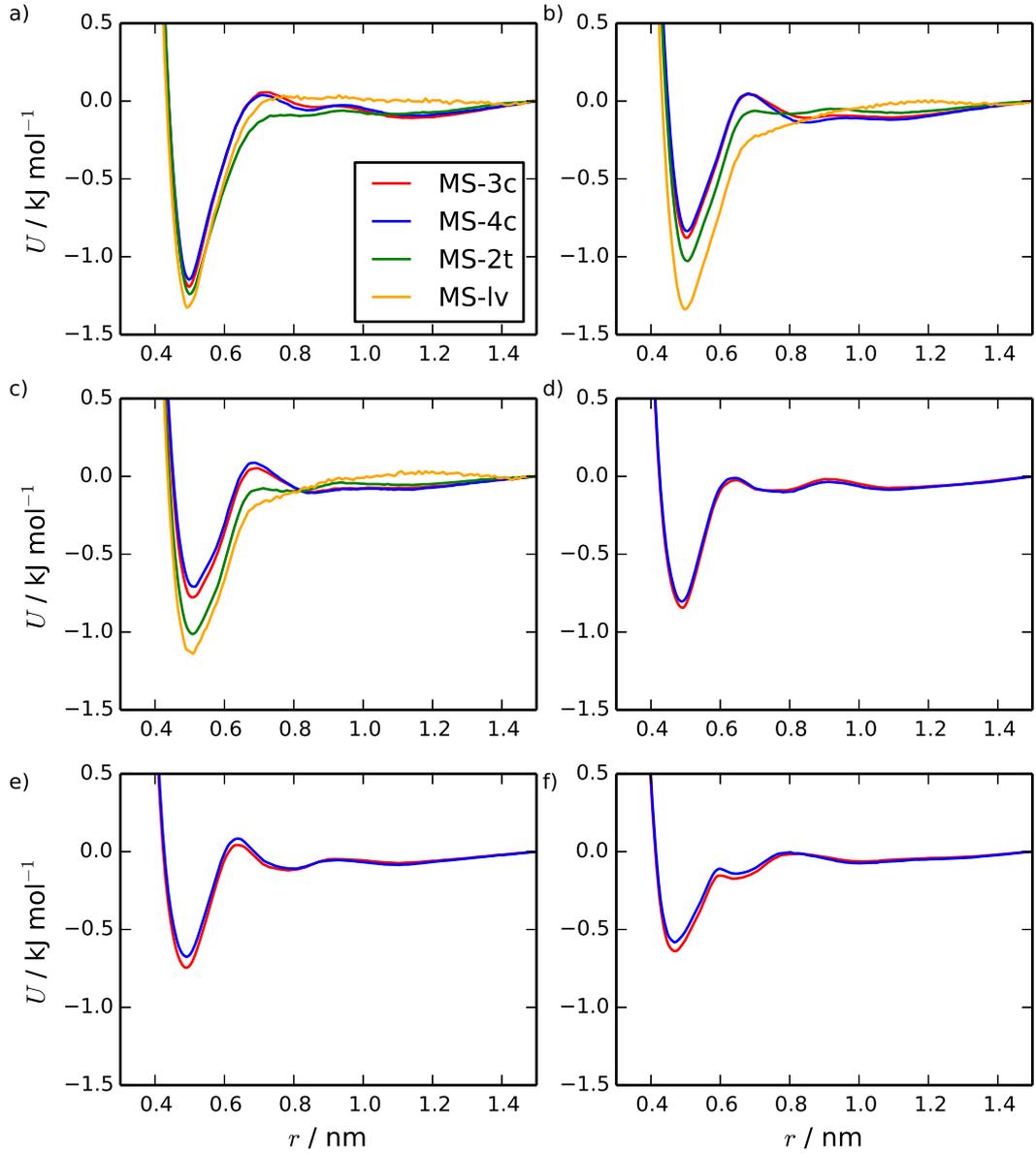
Figure 4.1: Non-bonded potentials parametrised using MS-IBI for the a) A-A, b) A-B, c) B-B, d) A-R, e) B-R, f) R-R interactions.

tials from each of the models discussed in this chapter are given in Figure 4.1.

4.2.3   Morse potentials using the simplex method

The simplex algorithm, as implemented in the VOTCA 1.3 package,[113,114] was used to optimise the parameters of Morse potentials, using the difference between the coarse-grained and reference atomistic RDFs as an objective function:

$$y(\boldsymbol{x}) = \sum_{r=0}^{r_{\mathrm{cut}}} |g_{\mathrm{ref}}(r) - g_{\mathrm{CG}}(r, \boldsymbol{x})|, \tag{4.2}$$

where $\boldsymbol{x}$ is the set of Morse potential parameters for the vertex being evaluated. In these models, the A and B beads in octane were treated as the same bead type, O; therefore, the non-bonded interactions to be parametrised for the octane-benzene mixture were O-O, O-R and R-R. The bonded interactions used were the same as those used in the other *bottom-up* models. Treating A and B as the same bead type reduces the number of interactions to be parametrised from 6 to 3, which speeds up the optimisation procedure. Although A and B do show different intermolecular RDFs in the reference system, these can largely be attributed to different intermolecular environments.

Models were parametrised using the simplex algorithm for pure octane and pure benzene. The vertices of the initial simplexes were calculated by carrying out Boltzmann inversion, and fitting the parameters of a Morse potential to the resulting numerical potential. These parameters were used for one vertex, and the remaining vertices were calculated by applying small random perturbations to these parameters.

Two models were then parametrised for an octane-benzene mixture with $x_{\text{oct}} = 0.5$. The first used the like interactions from the octane and benzene models and used the simplex algorithm to optimise the unlike interaction; this will be known as the Morse-MS model. The second re-optimised all interactions simultaneously for the mixture; this will be called the Morse-50 model. Both models used the Boltzmann inversion of the atomistic O-R partial RDF as a starting point for the unlike interaction. The Morse potentials from the single-component models were used as a starting point for the optimisation of the like interactions in the Morse-50 model.

The Morse potential parameters obtained from the simplex algorithm optimisations are given in Table 4.2.

4.3 Concentration transferability

Two MS-IBI models were parametrised to test concentration transferability. Each of these models used pure benzene and pure octane as reference systems for the benzene-benzene and octane-octane interactions, respectively. Additionally, one of the models used an $x_{\text{oct}} = 0.5$ mixture of octane and benzene (where x_{oct} is the mole fraction of octane); this model is referred to as MS-3c. The second model, called

Table 4.2: Morse potential parameters obtained from the simplex algorithm optimisations. The parameters are given to 5 significant figures.

Model	Interaction	r_0 / nm	ε / kJ mol^{-1}	α / nm^{-1}
	O-O	0.51399	0.79819	10.558
Morse-MS	R-R	0.46921	0.79806	10.532
	O-R	0.49286	0.79815	10.530
	O-O	0.50841	0.80022	10.511
Morse-50	R-R	0.47341	0.79800	10.428
	O-R	0.48989	0.79668	10.522

MS-4c, used two mixtures, with $x_{\text{oct}} = 0.3$ and $x_{\text{oct}} = 0.7$, as references in addition to the pure components. In both cases, the mixtures were used as references for all interactions, including like-like interactions. The results from these models were compared to results from a single-state IBI model parametrised at $x_{\text{oct}} = 0.5$. The interaction potentials for the MS-3c and MS-4c models are shown in Figure 4.1.

The performance of each model was tested across the concentration range, at $x_{\text{oct}} = 0.0, 0.2, 0.3, 0.5, 0.7, 0.8$ and 1.0. At each concentration, a constant *NPT* simulation was carried out at 298 K and 1 bar to calculate the density and RDF. The solvation free energies at each concentration were also calculated as described in Section 3.2.4.

The RDFs calculated using the MS-3c and MS-4c models at $x_{\text{oct}} = 0.2$ and 0.8, compared to atomistic results, are shown in Figures 4.2 and 4.3. In contrast to the results for standard IBI, MS-IBI was successful in constructing a concentration transferable model. The MS-3c model reproduced all of the RDFs quite well at $x_{\text{oct}} = 0.0, 0.5$ and 1.0. However, it should be noted that, at low octane concentrations ($x_{\text{oct}} = 0.2$ and 0.3), the RDFs involving benzene were reproduced slightly better than those involving octane and this was reversed at higher octane concentrations ($x_{\text{oct}} = 0.7$ and 0.8). However, the difference between the models is rather small, indicating that the additional computational expense of including an additional reference system does not result in a significant improvement in structural representabiliy or transferability in this case. While the MS-4c encompasses a wider concentration range for all of the interactions, including a single mixture concentration as a reference in addition to the pure components appears to be enough to capture all of the relevant contributions to the effective pair potential.

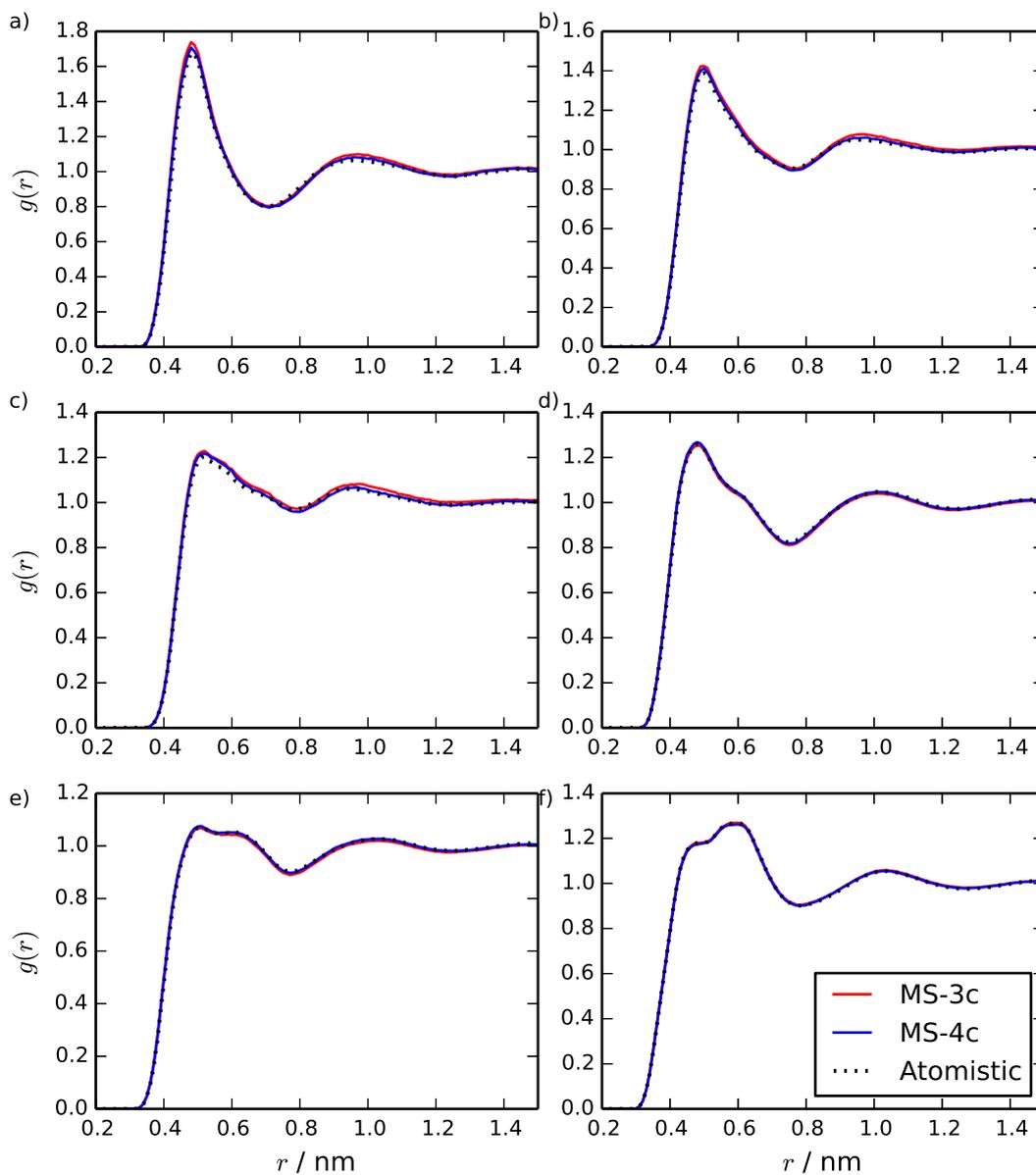After the application of pressure correction during the MS-IBI process, the MS-

Figure 4.2: Non-bonded distributions for $x_{oct} = 0.2$ from simulations of the MS-3c and MS-4c models, compared to atomistic simulation results, calculated at 298 K and 1 bar. The plots show the: a) A-A, b) A-B, c) B-B, d) A-R, e) B-R and f) R-R interactions.
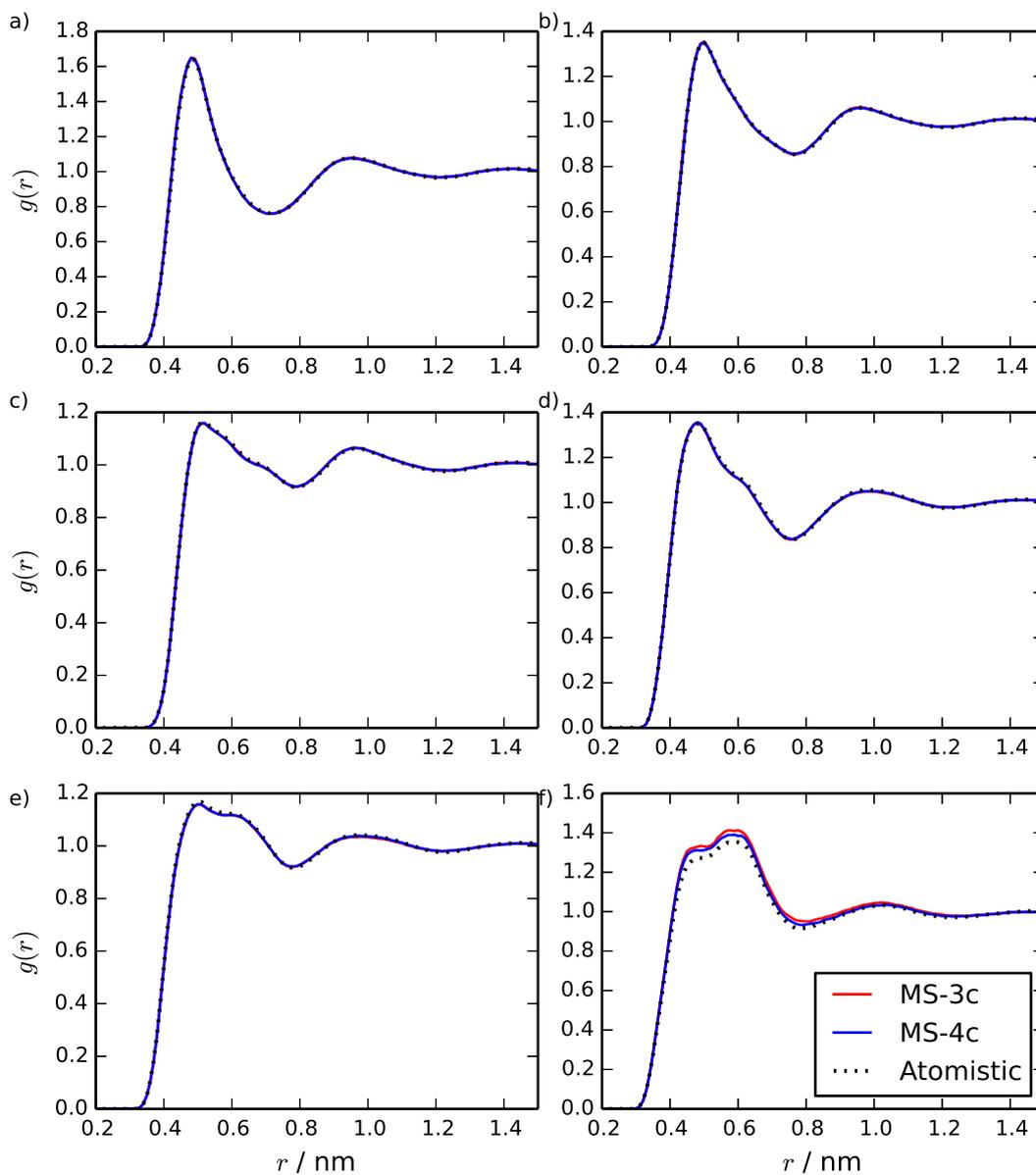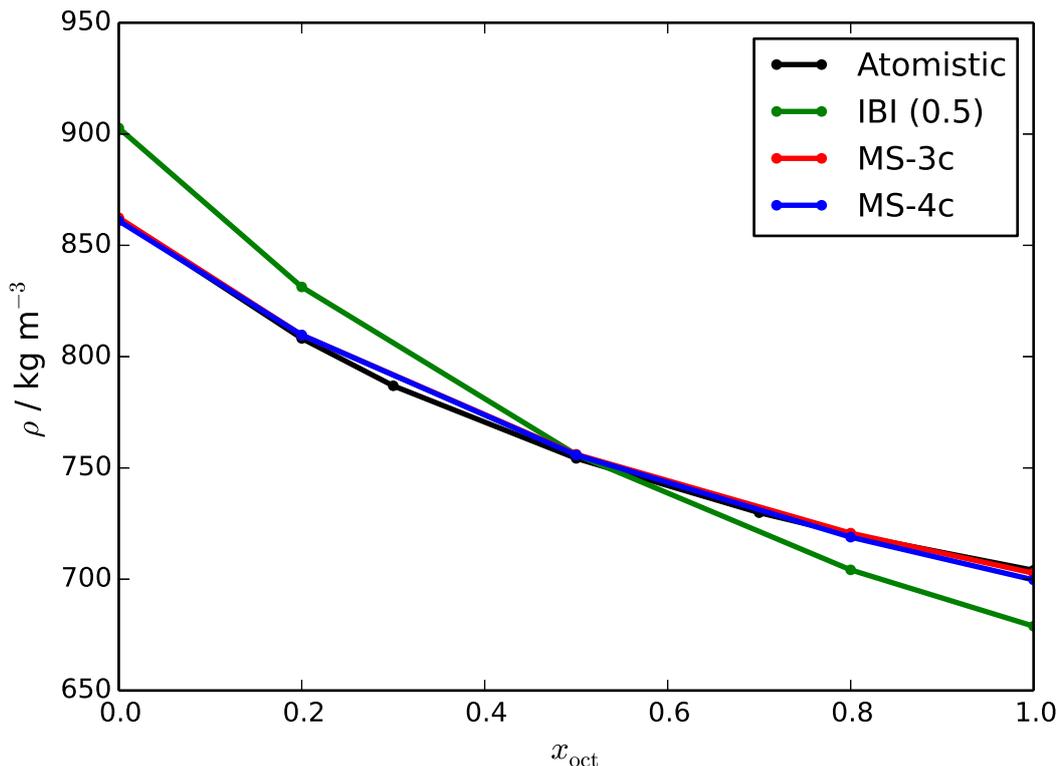
Figure 4.3: Non-bonded distributions for $x_{oct} = 0.8$ from simulations of the MS-3c and MS-4c models, compared to atomistic simulation results, calculated at 298 K and 1 bar. The plots show the: a) A-A, b) A-B, c) B-B, d) A-R, e) B-R and f) R-R interactions.

Figure 4.4: Density as a function of x_{oct}, calculated using atomistic, MS-3c, MS-4c and IBI parametrised at $x_{oct} = 0.5$.

3c and MS-4c models both gave pressures of close to 1 bar at the atomistic density for all of the reference concentrations used in the parametrisation. Therefore, both models were able to reproduce the density across the concentration range equally well, as shown in Figure 4.4. This is a marked improvement over the performance of the IBI model parametrised at $x_{oct} = 0.5$.

The solvation free energies calculated using the MS-3c and MS-4c models, here compared to IBI models parametrised at the relevant concentration, are shown in Figure 4.5. Different single state IBI models are used at each concentration, because, as shown in Figure 4.4, the densities (and therefore the pressures) will be incorrect at concentrations other than $x_{oct} = 0.5$. As shown for HFM models with incorrect pressures in Chapter 3, this will lead to incorrect free energy calculations.

The MS-IBI models show a marked improvement over the IBI models in the calculation of free energy changes, in terms of both representability and transferability. It was shown in Figure 4.4 that the MS-4c model is able to reproduce the atomistic pressure right across the concentration range; this translates to an accurate reproduction of the trend in solvation free energy across the same concentration range, as seen in Figure 4.5 The compromise of matching to multiple state-points does result
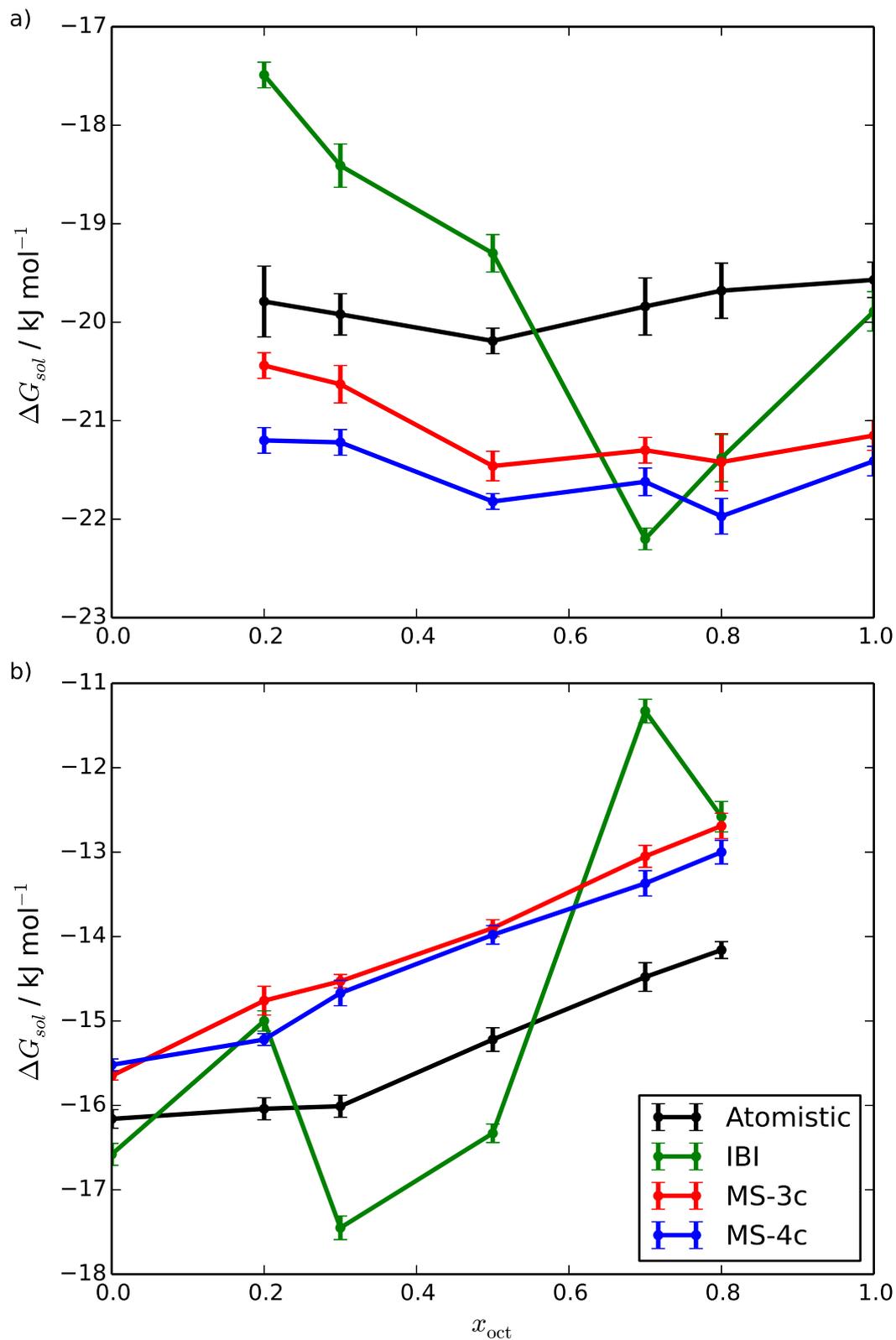
Figure 4.5: Gibbs free energies of solvation of a) octane and b) benzene in a solvent of an octane-benzene mixture with a concentration of x_{oct}, calculated using atomistic, MS-3c, MS-4c and IBI parametrised at the concentration at which the free energy was calculated.

in a systematic error for both MS-IBI models. This is in contrast to the IBI models, where a single model will not be thermodynamically transferable across the concentration range, and so accurate representation of the pressure and solvation free energy at a given concentration requires the use of a model parametrised specifically for that concentration. The necessity of using a separate coarse-grained model for each state-point impacts the ability of IBI to replicate the trend in solvation free energy across the concentration range; this is shown by the differing line-shapes produced by the IBI models, compared to those of the atomistic and MS-4c models.

4.4   Transferability of octane models

4.4.1   Using two temperature references for MS-IBI

Initially, the temperature transferability of MS-IBI models was tested by constructing a model using two reference systems of pure octane at 238 K and 378 K. This model is referred to as MS-2t.

During the parametrisation of the MS-2t model, a point was reached where it was no longer possible to improve the structural accuracy of the model for both references simultaneously. This occurred at the point where the potential update for the high temperature and low temperature systems nearly cancelled out, so the net potential update was close to zero. Changing the weighting of the two systems in determining the potential update changed the point at which the updates began to cancel out, but did not allow both systems' RDFs to be matched at the same time. During the pressure correction, the same effect was observed when one system's pressure reached a negative value while the other remained positive; this is shown in Figure 4.7. Again, changing the weighting of the potential updates did not fix the problem.

Figure 4.6 shows the site-site RDFs for the MS-2t model measured at 238 K and 298 K. Despite the issues matching the RDFs exactly at both state-points, the structural representability of the MS-2t model is still reasonably good. More problematic is that the model was not able to reproduce the atomistic densities across the range of temperatures any better than the IBI model, as shown in Figure 4.8. This was expected given that the pressures of the reference systems did not converge to 1 bar.
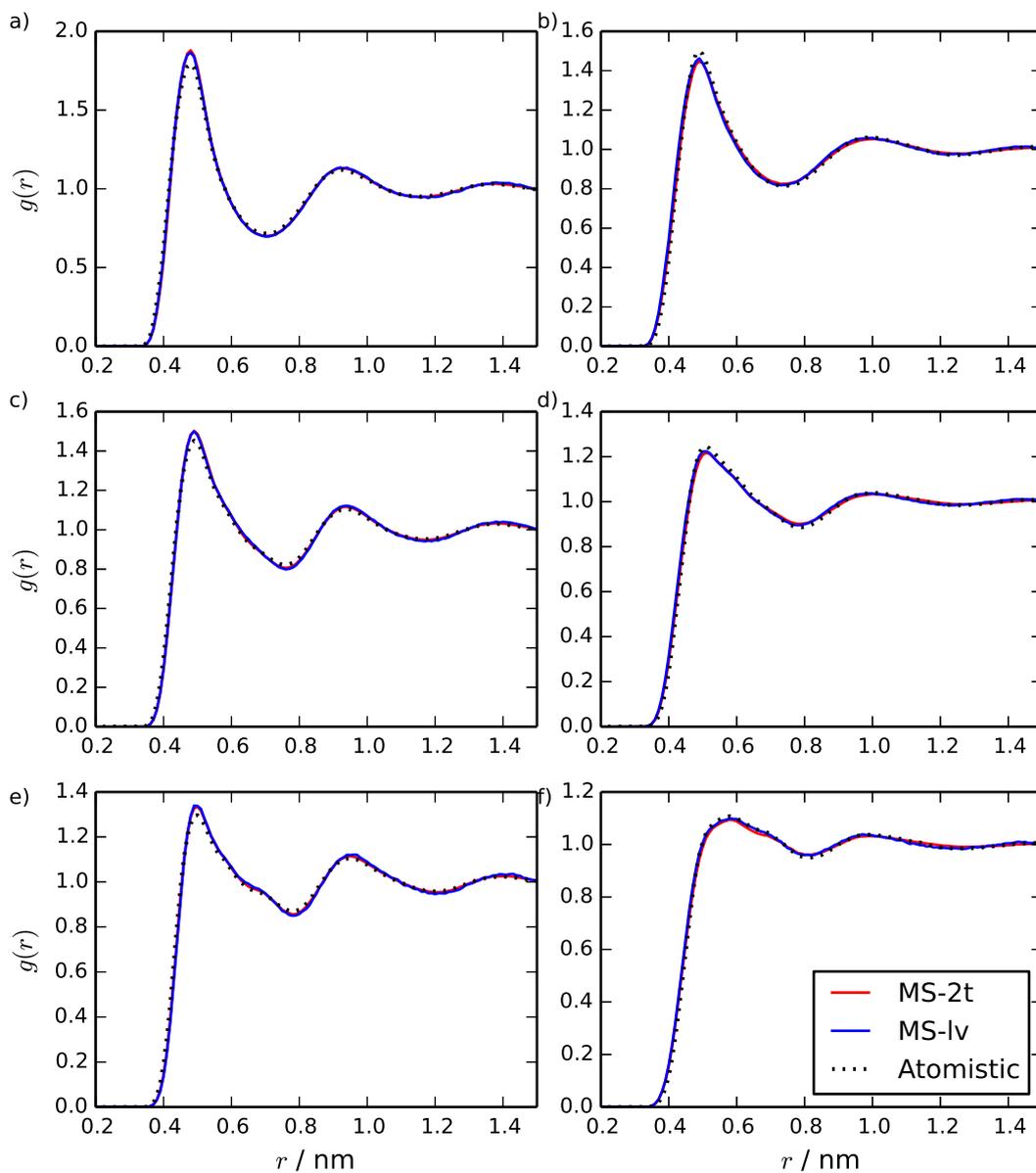
Figure 4.6: Site-site RDFs from simulations of pure octane using the MS-2t and MS-lv models, compared to the atomistic references. The plots show a) A-A at 238 K, b) A-A at 378 K, c) A-B at 238 K, d) A-B at 378 K, e) B-B at 238 K and f) B-B at 378 K.
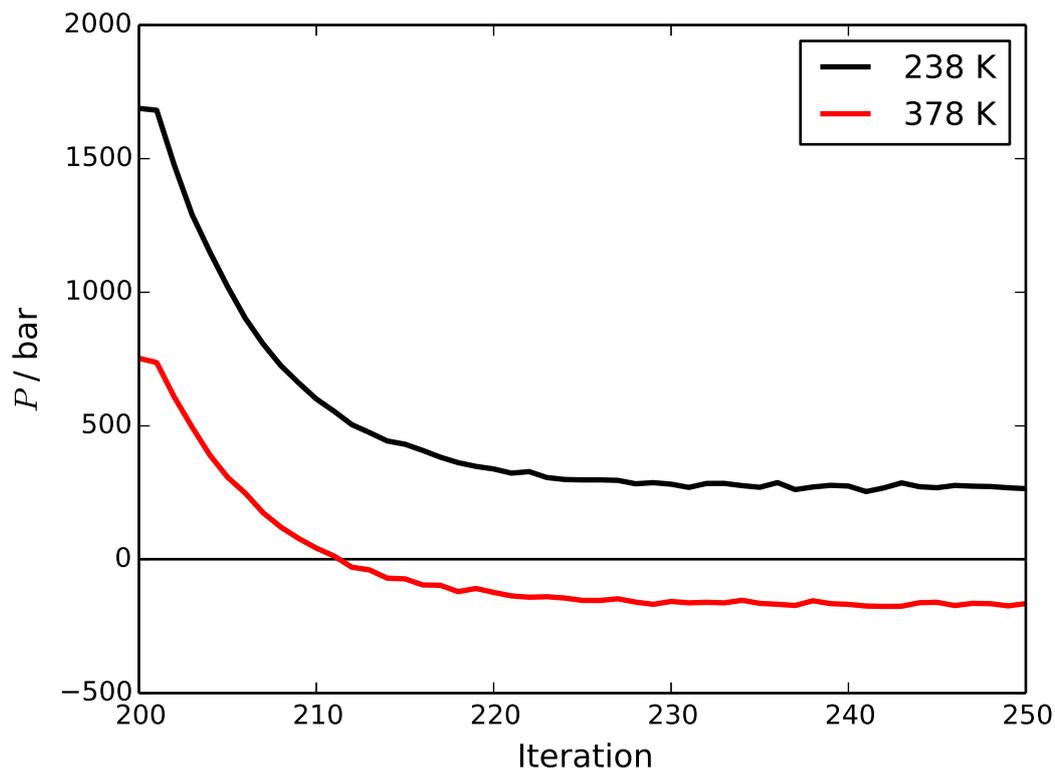
Figure 4.7: Pressure of octane measured at 238 K and 378 K, at the atomistic densities, during the pressure correction iterations for the MS-2t model parametrisation. The horizontal line indicates the target pressure of 1 bar.
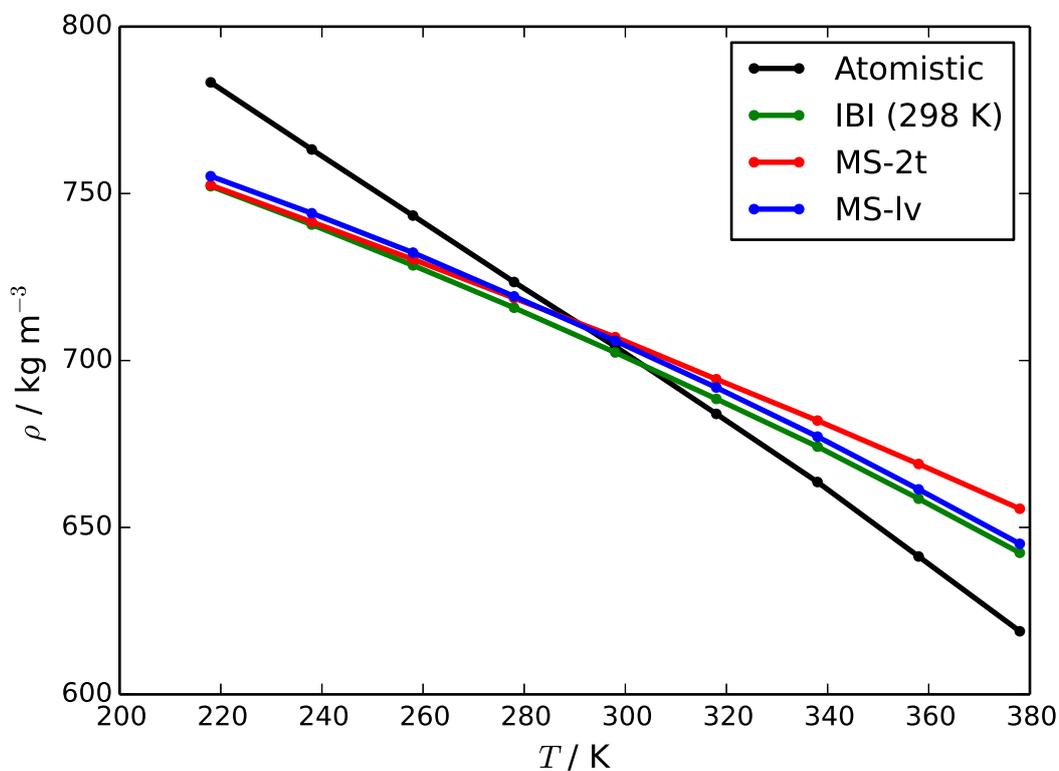

Figure 4.8: Density of octane at 1 bar as a function of temperature, calculated using atomistic, MS-2t, MS-lv and IBI parametrised at 298 K.

4.4.2 Using constant *NPT* simulations in the MS-IBI process

Moore *et al.* have shown that constant *NPT* simulations may be used during the MS-IBI process to give the correct pressure/density relationship at a single thermodynamic state-point.[118] Therefore, the use of constant *NPT* simulations at the two reference temperatures, in place of constant *NVT* simulations, was tested as a method of achieving an MS-IBI model which was transferable across the temperature range.

The very high pressures associated with *bottom-up* coarse-grained models will of course cause issues when running constant *NPT* test simulations. Care was therefore necessary when selecting starting potentials for the MS-IBI iterations. Simulations of the average Boltzmann-inverted pair potentials at constant *NPT*, which gave pressures of 2887 and 1597 bar for the low and high temperature references respectively, were not stable; the system expanded rapidly, and the simulations crashed within 100 ps. The potentials from the final MS-2t model were therefore chosen as a more stable starting point, with the hope that attempting to match the RDFs exactly in the constant *NPT* ensemble would improve both the structural accuracy and the pressure/density relationship of the model across the temperature range.

This methodology was not successful in producing a temperature transferable model. Figure 4.9 shows how the densities at the two reference temperatures changed over the MS-IBI iterations. The densities at both temperatures decreased steadily throughout the process. For the low temperature reference, this meant that the pressure/density relationship actually worsened. The high temperature reference crossed the correct density, but there is no indication that it was close to converging at that point.

4.4.3 Liquid-vapour references for MS-IBI

A study by Lyubartsev *et al.* describes the parametrisation of a coarse-grained water model using a liquid-vapour water system as a reference system.[110] An interesting aspect of the resulting model is that, although it was parametrised to match the RDF of the liquid-vapour system, it is able to match both the structure and pressure of the system reasonably well at the experimental liquid density of water. However, the method presented is not without issues. The authors found a signifi-
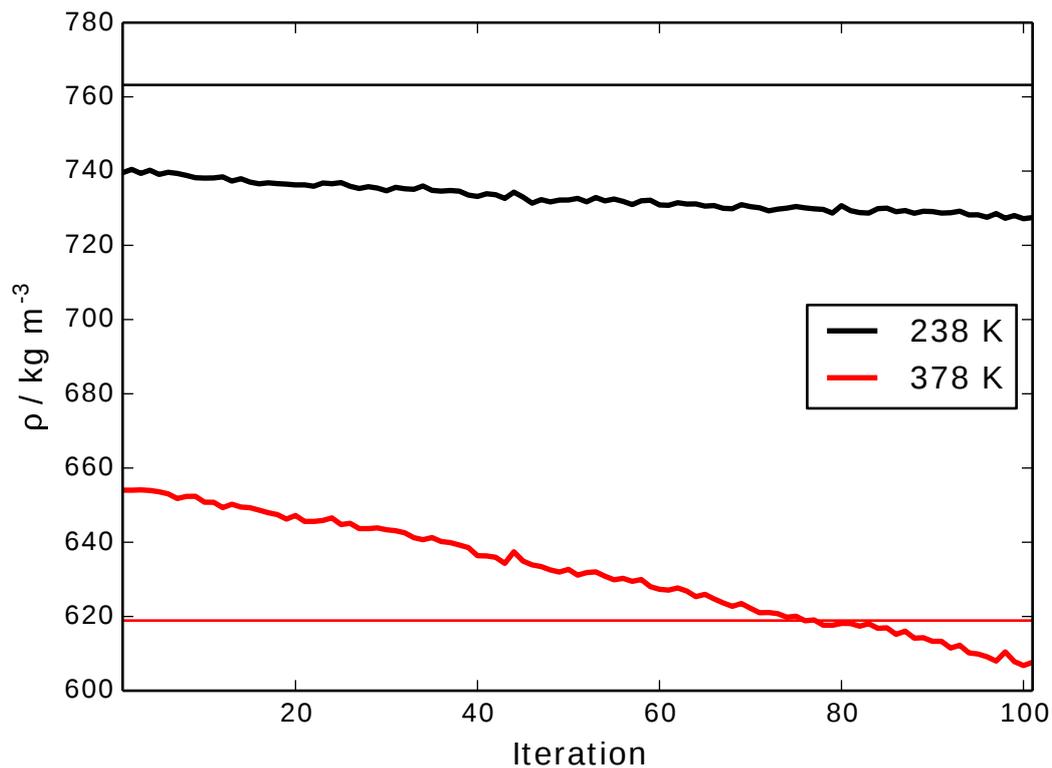
Figure 4.9: Densities of octane measured at 238 K and 378 K during the MS-IBI iterations with constant NPT test simulations. The horizontal lines indicate the densities measured using atomistic simulations at the reference temperatures.
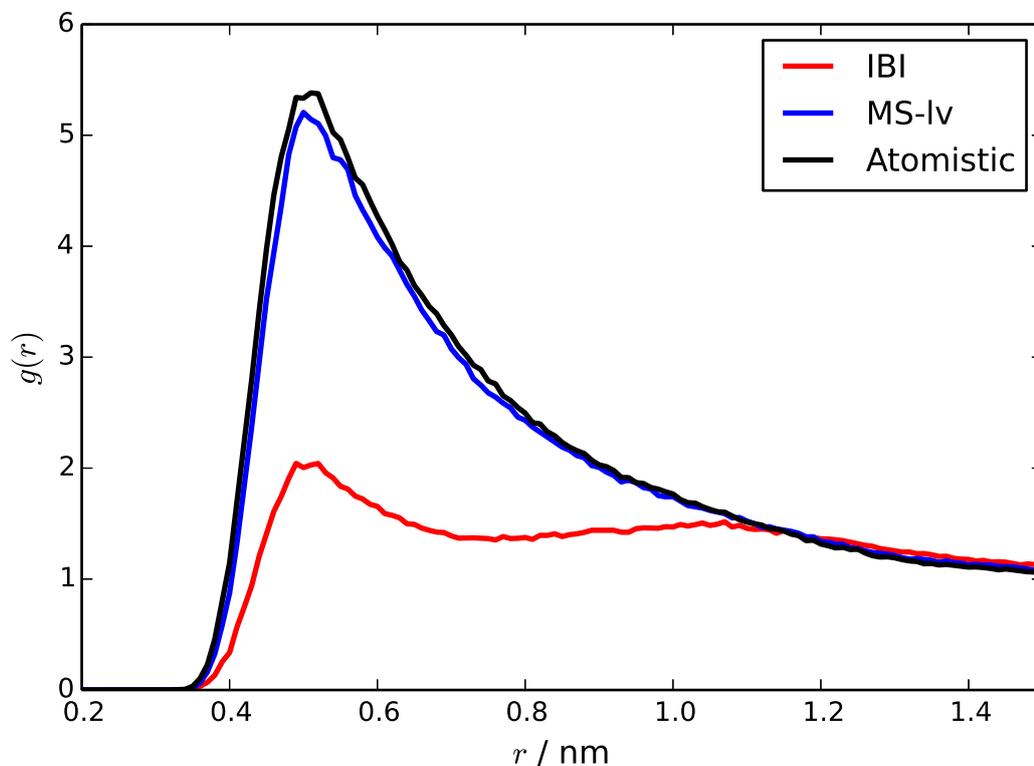

Figure 4.10: A-A partial RDFs calculated in the vapour phase at 298 K, from simulations of the pure octane IBI, MS-lv and atomistic models.

cant correlation between the relative sizes of the liquid and vapour phases and the pressure of the resulting coarse-grained system; in principle, by altering the density of the system, any desired pressure could be achieved. Although the authors do not directly address the reasons for this correlation, it seems likely that, at different densities, the extent to which the RDF of the system is dominated by the interface varies, and this has a large effect on the pressure of the coarse-grained model.

A similar approach would be to carry out an MS-IBI parametrisation, using separate liquid and vapour systems as references. This would have the advantage of removing the influence of the phase boundary on the pressure of the coarse-grained model. A model which is able to match the structure of a system on different sides of the vapour/liquid phase transition has the potential to improve upon the transferability of a standard IBI model. This approach was tested on the pure octane system.

The octane liquid reference system was the same system used for the $x_{\text{oct}} = 1.0$ IBI model in Chapter 3. The vapour phase reference also contained 1600 molecules simulated at constant NVT, at 298 K and a volume of 216000 nm^{-3} (corresponding to a density of 1.41 kg m^{-3}). MS-IBI iterations were carried out, with no pressure correction steps. Atomistic and coarse-grained vapour phase simulations were run using a stochastic integrator, and the simulation time was increased by a factor of 4 to gather sufficient statistics for calculating RDFs.

During various attempts at the parametrisation using different scaling factors, a correlation was observed between the MS-IBI iterations converging to good structural accuracy and good accuracy with regards to the pressure. This was particularly encouraging, given that pressure was not a target of the MS-IBI procedure. After 96 MS-IBI steps, the RDFs of both reference systems were matched to a high level of accuracy, with a pressure of -17 bar. When the same vapour system was simulated using the octane IBI model, the structural accuracy was worse than the octane MS-lv model; this is shown in Figure 4.10. From these plots, it appears that the atomistic vapour system includes some small octane clusters, which are cpatured by the MS-lv model, but not by IBI. Therefore, the MS-lv model has improved on the transferability of IBI to different densities at the same temperature. The MS-lv potentials also lack the pronounced hump which is present in the IBI potentials. Including the vapour reference in the parametrisation, which has a less structured RDF, prevents overfitting of the potentials to the more complex shapes of the liquid

RDFs, which occurs when using IBI and HFM.

However, the transferability of the MS-lv model across temperatures is not significantly better than the other coarse-grained octane models studied. The thermal expansion behaviour is similar to the IBI and MS-2t models, as shown in Figure 4.8. The radial distributions across the range are reasonably accurate, and are in fact extremely similar to those calculated from simulations of the MS-2t model. Therefore, transferability between volumes does not necessarily translate to transferability across temperatures.

4.5   Morse potentials

The RDFs calculated using the Morse-MS and Morse-50 models, at the concentrations for which they were parametrised and other concentrations, are shown in Figures 4.11 and 4.12. The level of detailed structural accuracy which can expected of these models is of course less than models employing numerical potentials. Nevertheless, the models all do reasonably well at reproducing the pairwise structure across the concentration range. This indicates that a good level of accuracy and transferability can be achieved for liquid structure without the overfitting that was observed in IBI and HFM.

There is relatively little difference between the Morse-MS and Morse-50 models in terms of their structural accuracy across the concentration range. Re-optimising the like–like interactions for the mixture does not appear to be worth the additional computational cost, at least for this system. This hybrid approach could lead to the efficient parametrisation of coarse-grained models for a range of liquid mixtures. The use of analytical potentials may also allow combination rules to be tested. A recent study suggested a set of combination rules for Morse potentials:

$$\varepsilon_{12} = \frac{2\varepsilon_{11}\varepsilon_{22}}{\varepsilon_{11} + \varepsilon_{22}} \tag{4.3}$$

$$r_{0,12} = r_{0,11}r_{0,22}\frac{r_{0,11} + r_{0,22}}{r_{0,11}^2 + r_{0,22}^2} \tag{4.4}$$

$$\sigma_{12} = \sigma_{11}\sigma_{22}\frac{\sigma_{11} + \sigma_{22}}{\sigma_{11}^2 + \sigma_{22}^2} \tag{4.5}$$

Table 4.3: Weightings given to matching pressure and RDFs during the attempted pressure correction of the octane Morse model, and the pressure achieved.

Weight (pressure)	Weight (RDF)	P / bar
0.0	1.0	1262
0.01	10.0	1119
0.05	10.0	412
0.01	1.0	55
1.0	0.0	3.1

Here, σ is the distance at which $U_{\mathrm{Morse}} = 0$. α can be calculated using:

$$\alpha_{12} = \frac{\ln 2}{r_{0,12} - \sigma_{12}} \tag{4.6}$$

These mixing rules were effective for noble gas mixtures,[170] and it would be interesting to test whether this could further improve the efficiency of parametrising Morse potentials. This would, however, require extensive validation on a range of liquid mixtures, since simple combination rules are not guaranteed to be effective for all coarse-grained models.[171]

Of course, as observed in Chapter 3, if we are interested in studying thermodynamics, it is vital that the coarse-grained model is able to represent the pressure of the atomistic system well. The linear pressure correction method is not applicable to potentials with fixed functional form like the Morse potential. However, in principle, the simplex method provides a framework for matching any simulation observable, including pressure. Therefore, an attempt was made to parametrise a Morse potential which matches the pressure of pure octane by using the difference between the coarse-grained pressure and the target pressure (1 bar) as a target.

If pressure alone was used as a target, then the resulting coarse-grained model had the correct pressure. However, looking more closely at the potential and structure obtained showed serious issues: the Morse potential obtained was purely attractive, resulting in overlapping octane molecules. If the pressure and the RDF were used as dual targets for the optimisation, then the weights of the two targets could be altered such that either the pressure or the RDF was matched well, but never both simultaneously. As an illustration of this, Table 4.3 shows the pressures obtained with different scaling factors, and Figure 4.13 shows the RDFs of the different models.
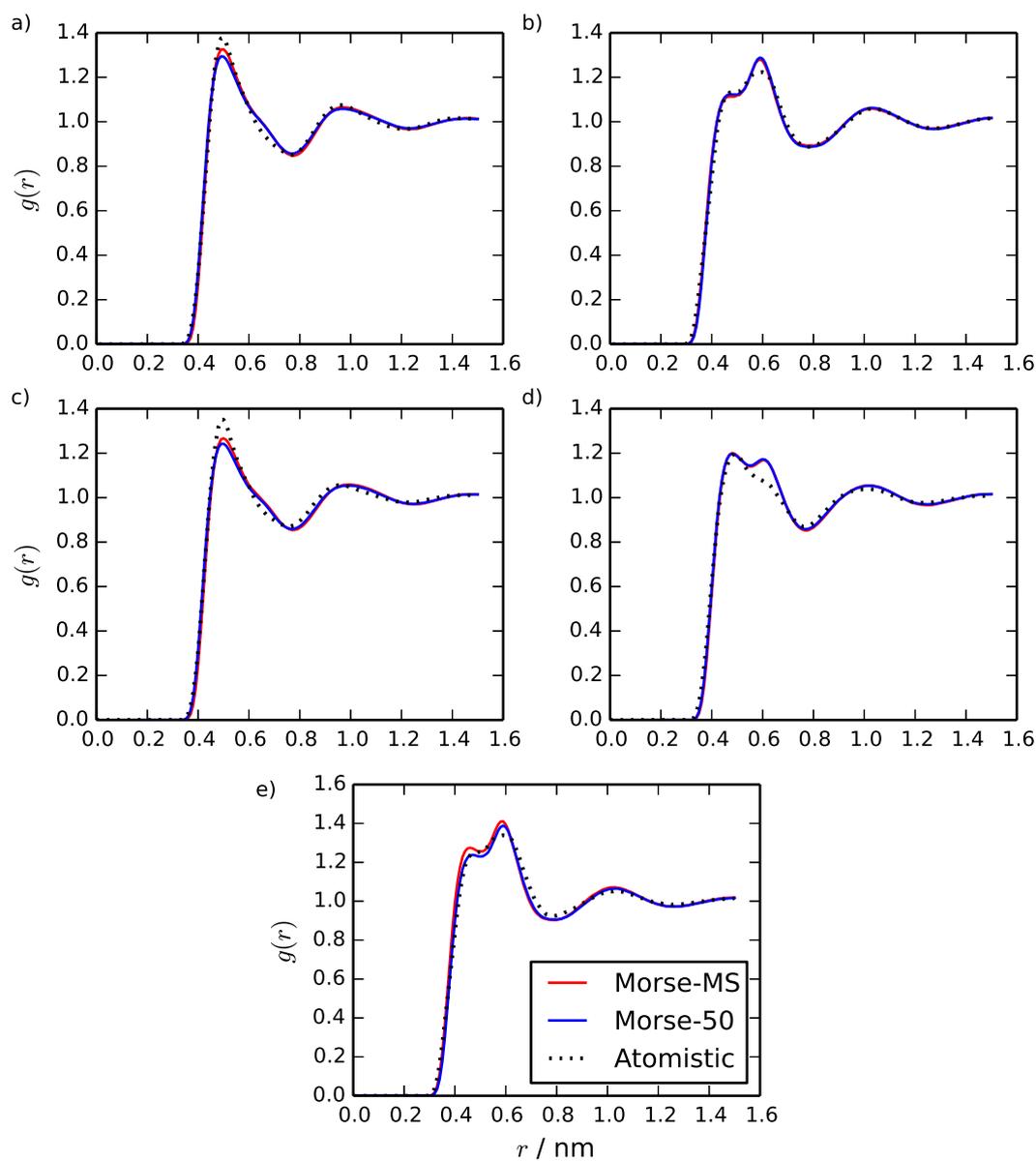
Figure 4.11: Site-site RDFs calculated at 298 K and 1 bar using the Morse-MS and Morse-50 models, compared to the atomistic references, from simulations of a) pure octane, b) pure benzene and a mixture with $x_{oct} = 0.5$, the c) O-O d) R-R and e) O-R interactions.
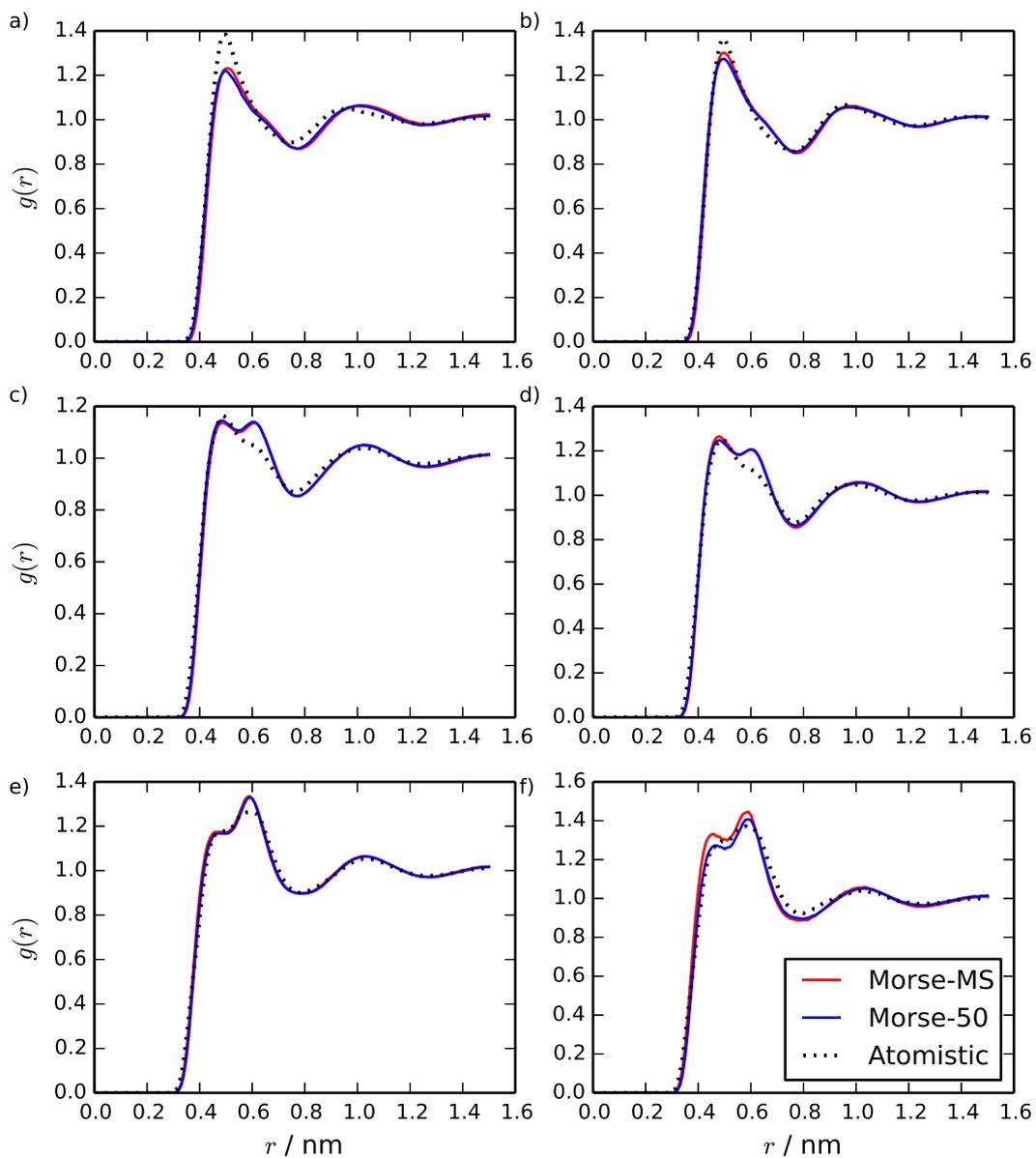
Figure 4.12: Site-site RDFs calculated at 298 K and 1 bar using the Morse-MS and Morse-50 models, compared to the atomistic references, from simulations of octane/benzene mixtures with $x_{oct} = 0.2$: a) O-O, c) O-R, e) R-R, and $x_{oct} = 0.8$: b) O-O, d) O-R, f) R-R.
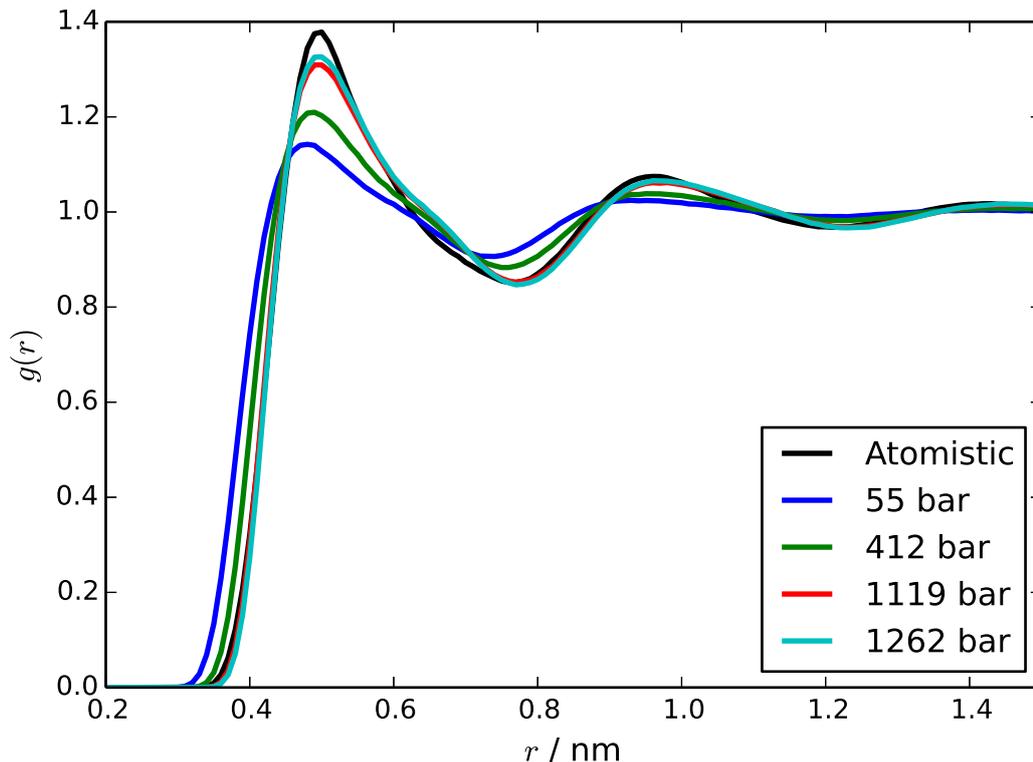
Figure 4.13: RDFs from simulations of octane at 298 K and the atomistic density using a range of pressure-corrected Morse potentials. The pressures of these models are indicated.

Therefore, while the fixed functional form of a Morse potential is advantageous in preventing overfitting to a single state-point and simplifying the optimisation of the coarse-grained model, the lack of flexibility does appear to have downsides when it comes to simultaneously representing structural and thermodynamic properties, even at a single state-point. The key advantage of the linear pressure correction to numerical potentials is that it has the largest effect on the long-range part of the potential, while the attractive well, to which structural properties are most sensitive, is mostly unaffected. When the functional form is constrained to a Morse potential, however, it becomes very difficult to alter one part of potential while keeping other parts unaffected, and so pressure correction becomes more difficult.

4.6 Local-density potentials - preliminary work

The linear pressure correction is useful for obtaining the correct pressure in a coarse-grained model at a single state point. However, as shown in Chapters 3 and 4, it does not guarantee a transferable model, particularly across a range of tem-

peratures. This is because it does not address the underlying reason for the poor representation of pressure in coarse-grained models, the state point dependence of the pair potentials.[39] Therefore, while the idea of a coarse-grained model based on simple pair potentials, which is transferable and has good structural and thermodynamic representability, is very attractive, actually obtaining such a model proves to be extremely difficult.

One solution to this is to use a *top-down* approach, where transferability and good thermodynamics appear to be easier to come by, and to attempt to improve the representability of structures for these models, which are often rather poor. From a *bottom-up* perspective, however, improving transferability may require including additional terms in the potentials. For example, three-body terms could be included in a coarse-grained model. For simple systems, this is known to improve representability.[127,172] It may also allow a coarse-grained model to react better to its environment, for example in the modelling of a phase-separated liquid mixture.[173] However, it is not guaranteed that three-body terms alone would address all the causes of poor transferability in coarse-grained models; specifically, the multi-body PMF for the system contains many contributions that are state-point dependent, which it may not be possible to match using only configuration-dependent potentials.[43,174]

The use of local-density dependent potentials (described in Chapter 2) is a promising idea, which has been explored in the literature recently.[144–146] Given their connection to volume-dependent potentials,[144] which allow the volume-dependence of a coarse-grained model to be correctly addressed[139], local density potentials may be useful as a form of pressure correction.

Recent publications on the use of local-density potentials have generally involved the simultaneous parametrisation of numerical pair and local-density potentials, for example using the relative entropy method. However, given the promise of the Morse potentials described in the previous section as a simplified method for efficiently obtaining structurally representative and chemically transferable coarse-grained models, it would be interesting to see whether these existing potentials could be combined with local-density potentials to give a coarse-grained forcefield which was also thermodynamically representative, and transferable across temperatures.

Methodology for pressure correction with volume potentials,[140] and for converting volume potentials to local density potentials already exists.[144] However, it would

be advantageous to have a way to directly carry out pressure correction with local density potentials without the intermediate stage. Aside from reducing the number of steps in the parametrisation of a coarse-grained model, this would be particularly useful for structurally complex systems, and those with many bead types. Volume potentials, by their nature, only consider bulk properties; local density potentials, on the other hand, inherently consider changes to the local environment of a bead, and so by directly parametrising them we avoid missing crucial contributions to the energy of the system.

4.6.1 Implementation in DL_POLY

There is no widely available molecular dynamics software package which includes local-density potentials as standard; therefore, it was necessary to modify a package to include this interaction type. The DL_POLY 4.08 package was chosen for this purpose because: a) it is open-source; and b) it was written in a modular way, with the intention that users could easily write and include their own modules.[175]

Local-density potentials were implemented in DL_POLY according to the equations given in Section 2.7.2. The functional form of $\psi(r_{ij})$ implemented by Sanyal and Shell[143] was used. This is given by:

$$\psi(r_{ij}) = \begin{cases} 1 & r_{ij} \leq r_0 \\ c_0 + c_2 r_{ij}^2 + c_4 r_{ij}^4 + c_6 r_{ij}^6 & r_0 < r_{ij} < r_\text{c} \\ 0 & r_{ij} \geq r_\text{c} \end{cases} , \qquad (4.7)$$

where r_0 and r_c are in inner and outer cutoff, and the coefficients are given by:

$$\begin{aligned} c_0 &= \frac{1 - 3r_0^2/r_\text{c}^2}{(1 - r_0^2/r_\text{c}^2)^3}, & c_2 &= \frac{1}{r_\text{c}^2} \frac{6r_0^2/r_\text{c}^2}{(1 - r_0^2/r_\text{c}^2)^3}, \\ c_4 &= -\frac{1}{r_\text{c}^4} \frac{3(1 + r_0^2/r_\text{c}^2)}{(1 - r_0^2/r_\text{c}^2)^3}, & c_6 &= \frac{1}{r_\text{c}^6} \frac{2}{(1 - r_0^2/r_\text{c}^2)^3}. \end{aligned} \qquad (4.8)$$

The potentials were implemented as a tabulated potential, to allow for total flexibility in the functional form. Multiple local density potentials may be defined between different pairs of sites, and it is possible for U_LD^{ij} and U_LD^{ji} to have different functional forms.
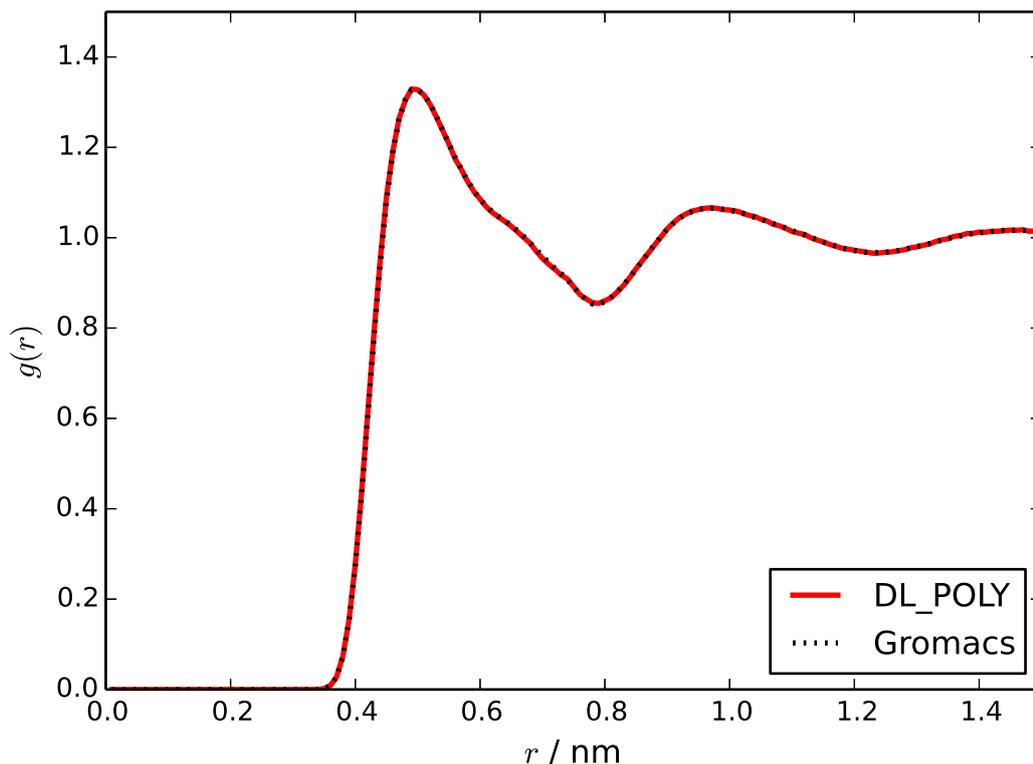
Figure 4.14: RDFs from simulations of octane at 298 K and 1 bar, using the same Morse potential with Gromacs and DL_POLY.

4.6.2 Testing the code

Before any parametrisation of local-density potentials, it was necessary to confirm that stable molecular dynamics simulations could be run on DL_POLY using the Morse pair potentials, and that these simulations gave the same results as those run using Gromacs. Small differences between simulation packages are unavoidable due to small differences in the way various algorithms are implemented; however, it is important that the results are at least closely comparable.

Initial tests were carried out on the pure octane system, using the Morse potential parameters optimised using the simplex algorithm. Immediate issues were found with tabulated dihedral potentials, which would regularly cause simulations to crash. Similar issues were found when an analytical dihedral potential was used. However when no dihedral potential was used, and 1–4 interactions were described with the Morse potential, the simulations no longer crashed. It was verified that doing the same thing in Gromacs made little difference to the RDF or pressure of the system.

Once a working model had been established in DL_POLY, the results obtained using it were compared to those from the same model simulated using Gromacs.
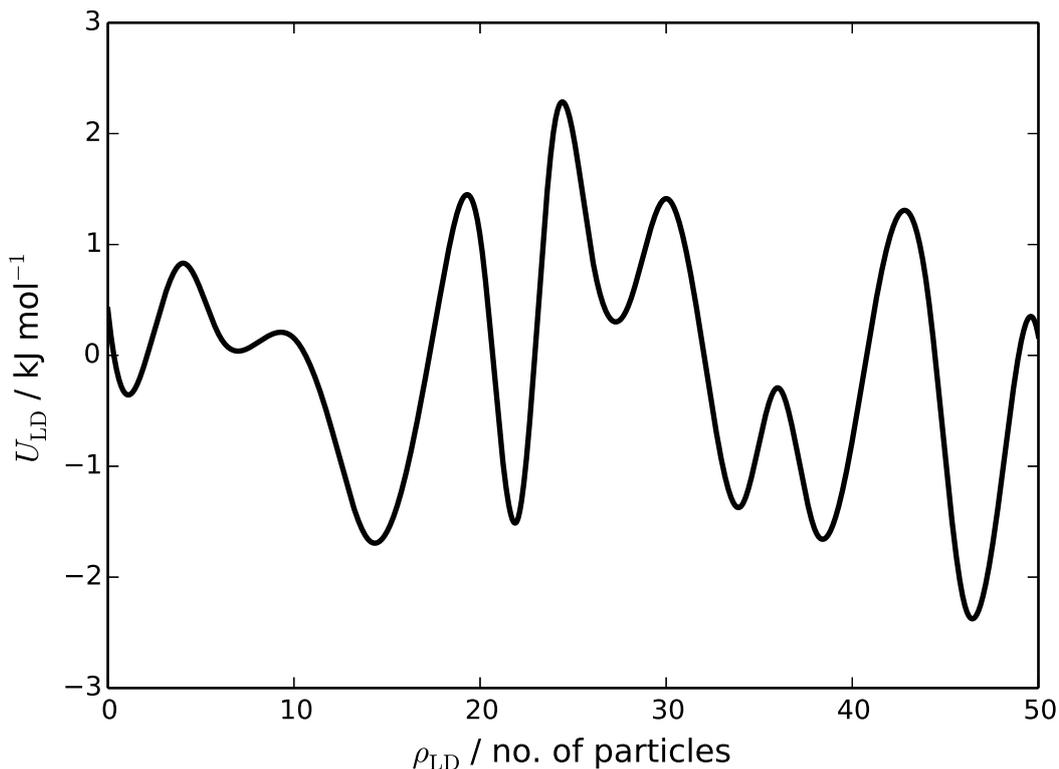
Figure 4.15: Example of a local density potential obtained during the attempted parametrisation using CMA-ES

The RDFs from simulations using the two packages are given in Figure 4.14; it can be seen that these are very similar. The pressures obtained from the two packages are also very similar (1.27 kbar for Gromacs and 1.28 kbar for DL_POLY).

Finally, a basic check of the new code was carried out by doing a simulation of the working model with a flat local-density potential (i.e. one with the same potential for all values of the local density). This simulation gave the same results as one without a local-density potential, indicating that the new code does not affect the functioning of the existing code.

4.6.3   Implementation in VOTCA

Preliminary work was also carried out to allow the parametrisation of local-density potentials using the iterative framework of VOTCA 1.3. Since this iterative framework is extremely flexible, this involved adding scripts to automate the construction of tabulated local-density potentials from a set of grid-points. The values of these grid-points are to be parametrised using the existing optimizers in VOTCA (either the simplex or CMA-ES method).

In an initial test, the CMA-ES method[115] was used to try to optimise a numerical local-density potential for octane at 298 K, simulated using the octane Morse model described in Section 4.5. The CMA method was used because it is often more effective than the simplex algorithm for optimising a large of number of parameters.[113] The local-density potentials seen throughout the optimisation all resembled the one shown in Figure 4.15. The target pressure used was 1 bar; however, at no point was convergence to this value observed. The likely reason for this is that the starting point used (a flat potential) is not close to the optimal potential, which makes it difficult for the optimiser to find a minimum in parameter space.

Further work here is clearly required. This should focus on finding a good starting point for the algorithm, including whether there are any convenient functional forms for the local density-dependent potential which may simplify the optimisation. Since the high pressure causes the system to expand, a local density potential which counteracts this may be a good starting point. For example, a simple linear potential which favours higher local densities, and therefore acts as an attractive potential, may be sufficient.

4.7   Conclusions

The MS-IBI approach clearly improves on the results from IBI. In practice, MS-IBI offers a compromise where slightly worse fits to pairwise RDFs allows a better representation of some thermodynamic properties and improved transferability. However, this is at some additional computational cost in terms of fitting. Moreover, the oscillations of the effective potential are not eliminated by this method and for pressure-consistent potentials to be produced, further pressure corrections may be required. This can make it more difficult to achieve a single temperature transferable potential, since a high pressure in one reference system can be balanced by a too low pressure in another. It is also important to note that the burden of producing MS-IBI potentials increases dramatically both with the number of different types of coarse-grained beads and the number of reference systems. As each pair interaction needs to be fitted, the practical application of IBI/MS-IBI for complex systems is extremely computationally expensive.

The failure of MS-IBI to produce a single temperature transferable model highlights the difficulty of achieving this in a coarse-grained model. Some success has

been achieved by scaling or reweighting a model developed at one temperature to other temperatures, using the MS-CG[176] and IBI[46,116] methods. In terms of capturing densities, the conditional reversible work method has also been fairly successful at achieving temperature transferablity, although the model has worse representability at the temperature for which it was parametrised when compared to pressure-corrected IBI.[106] Ultimately, the issue of temperature transferability in a pair potential is one that is yet to be solved, and doing so will require addressing the differences in the free-energy decomposition with temperature which come to the fore when coarse-graining.[177]

The simplex algorithm was shown to be a successful method for parametrising structurally accurate coarse-grained models. The resulting models also have good chemical transferability. However, attempts to improve their thermodynamic representability without compromising on their good structural accuracy proved difficult. Overall, the use of a fixed functional form in *bottom-up* coarse-grained models does simplify their parametrisation considerably, but does not necessarily represent a solution to the transferability problem. The models from such a method do in fact share the state-dependence of other *bottom-up* coarse-grained models, as shown by their inability to correctly represent pressure.

Ultimately, the issue of transferability in coarse-grained models is one which is yet to be solved, and it seems that coarse-grained models with good transferability, structural representability and thermodynamic representability may not be feasible when the interactions are limited to only pair potentials. The groundwork has therefore been laid for parametrising local density-dependent potentials; these may be able to address the issue of state-point dependence in coarse-grained models, which is crucial to solving the transferability problem. Future work could focus on efficient methods for parametrising these potentials to reproduce the pressure of the underlying atomistic system.

Chapter 5

Parametrising coarse-grained models of a chromonic liquid crystal

5.1 Introduction to liquid crystals

5.1.1 Liquid crystal phases

Liquid crystals are a class of soft matter which exhibit fluid-like behaviour characteristic of a liquid, along with some kind of anisotropic ordering of the molecules, like a crystal.[178] The different liquid crystal phases are characterised by the degree and type of positional and orientational order present. The phases exhibited, and the conditions under which they are formed, are highly system-dependent, resulting from a subtle balance between the interactions between components. Broadly speaking, liquid crystals can be divided into thermotropics and lyotropics, which form a variety of liquid crystal phases depending on temperature and concentration, respectively.[178,179]

Thermotropic liquid crystal phases are formed by anisotropic molecules, at a range of temperatures between the usual liquid and crystalline phases. A number of thermotropic phases exist, but the most common is the nematic phase; here, the orientations of the molecules are aligned, but the positional order is like that of a liquid. The driving force for the alignment of the molecules can be explained by excluded volume arguments, initially suggested by Onsager.[180] A pair of directionally aligned, rod-shaped, molecules will have a smaller excluded volume than two

molecules at an angle to each other; alignment therefore maximises the positional entropy of the system. In thermotropic liquid crystals, liquid crystal phases will occur when there is sufficient thermal motion for the system to flow like a liquid, but not so much that the alignment of the molecules is lost.

Thermotropic mesogens can be split into two general types: calamitic (rod-like) and discotic (disc-like).[178] Each of these types has a characteristic set of phases which it generally forms, although there is significant variation between different mesogens. Thermotropics are most commonly found in liquid crystal displays; however, their unique physical properties have given rise to many more applications.[181,182]

Most lyotropic liquid crystals are formed from amphiphilic molecules in aqueous solution. These molecules typically consist of a hydrophilic head group and a hydrophobic tail. In aqueous solution, the differing solubility of the two groups leads to the formation of micelles, in which the head is preferentially exposed to the solvent. This behaviour is driven by the hydrophobic effect, a phenomenon whereby hydrophobic molecules aggregate to avoid contact with water. This is favoured above a certain concentration (the critical micelle concentration) for both entropic and enthalpic reasons. The formation of structured, hydrogen-bonded solvation shells of water molecules around solute molecules has an entropic cost; the aggregation of solute molecules therefore increases entropy by reducing the surface area in contact with water. Additionally, there is an enthalpic driving force which arises from favourable interactions between hydrophobic chains inside micelles, and between the hydrophobic tails and water.[183,184] Different aggregate structures are seen depending on the concentration, from spherical micelles at low concentrations, to cylindrical and lamellar aggregates at higher concentrations. These differing shapes give rise to distinct liquid crystal phases. Lyotropics are extremely common, particularly in the surfactants found in soap, and the lipids which form cell membranes.[185,186]

5.1.2 Chromonic liquid crystals

Chromonic liquid crystals are a class of lyotropics characterised by the self-assembly of molecules in aqueous solution into columnar aggregates, rather than micelles, which go on to form liquid crystal phases. The molecules which form chromonics tend to consist of a hydrophobic, usually aromatic, core surrounded by hydrophilic groups, which can be either ionic or non-ionic. While the driving
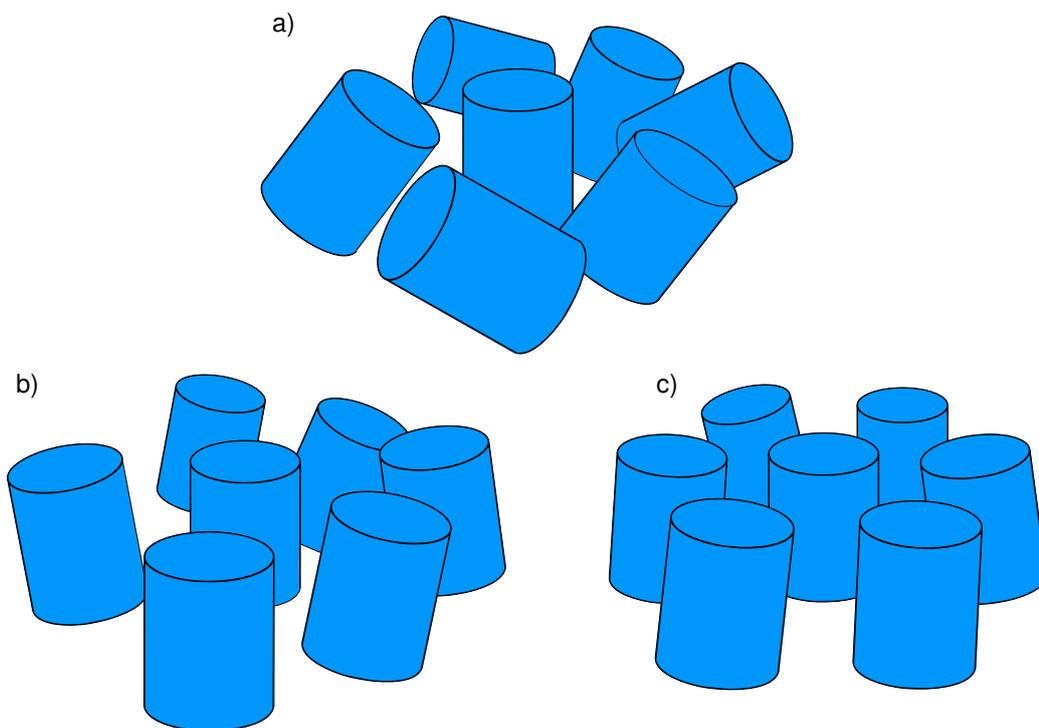
Figure 5.1: Liquid crystal phases commonly observed in chromonic systems: a) isotropic (I), b) nematic (N_C) and c) hexagonal columnar (C_H).

force behind aggregate formation (the hydrophobic effect) is similar to conventional lyotropics, the different aggregate shape leads to a distinctive set of chromonic liquid crystal phases.[187,188] Due to their polyaromatic cores, they are often found in dyes. Their optical properties do not arise directly from their liquid crystalline behaviour, although the formation of aggregates in dyes is known to affect the position of the absorption spectrum bands, and therefore the colour.[189,190]

Unlike conventional lyotropics, where the formation of micelles occurs only above the critical micelle concentration, the formation of short chromonic stacks occurs even at very low concentrations.[186] These stacks can be categorised according to whether there is an offset between molecules in the stack; stacks with large offsets are called J-aggregates and those without an offset are H-aggregates. In general, the phase diagram of a chromonic liquid crystal consists of: the isotropic (I) phase, in which the columns have no orientational order; the nematic (N_C) phase, which is analogous to that found in thermotropic liquid crystals; and the hexagonal (C_H) phase, in which the aligned columns pack in a hexagonal manner. At high concentrations, chromonics often lose their liquid crystalline nature, forming a crystalline phase. These phases are illustrated in Figure 5.1.[191]

One of the more commonly studied examples of a non-ionic chromonic is 2,3,6,7,10,11-
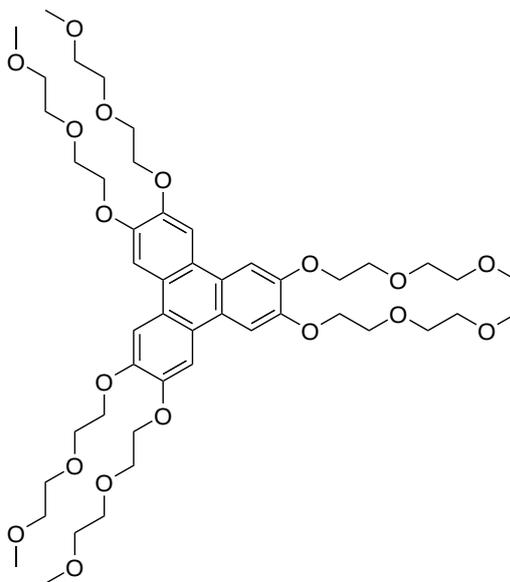
Figure 5.2: Molecular structure of TP6EO2M.

hexa-(1,4,7-trioxa-octyl)-triphenylene, more commonly known as TP6EO2M, the structure of which is shown in Figure 5.2. Like other chromonics, the molecule consists of a hydrophobic aromatic core and hydrophilic arms, in this case short ethylene oxide chains. As with all liquid crystal mesogens, the molecule's ability to form chromonic liquid crystal phases is extremely sensitive to the relative sizes of these two portions of the molecule.[192] For example, shortening the arms significantly reduces the water solubility, while lengthening them prevents the molecules from forming any sort of aggregate.[193–195]

TP6EO2M has been studied extensively both experimentally[193–197] and by atomistic simulation.[86,198] The formation of TP6EO2M stacks has been suggested to be "quasi-isodesmic", which means that the free energy of assocation of a molecule onto a stack is the same regardless of the stack size, with the exception of the formation of the initial dimer.[86] The same simulation study showed, by analysis of the temperature dependence of the free energy of association, that it arises from favourable enthalpic and entropic contributions in roughly equal proportions. These are likely to be similar in nature to the hydrophobic effect seen in conventional lyotropics, with the π–π stacking favoured enthalpically and an entropic contribution arising from the structuring of water in the solvation shell.[86] This agreed with an earlier experimental study in which the free energy of association was calculated at different temperatures using NMR.[199] The phase diagram of TP6EO2M is shown in Figure 5.3. Like most chromonics, it shows transitions from isotropic to nematic, and then
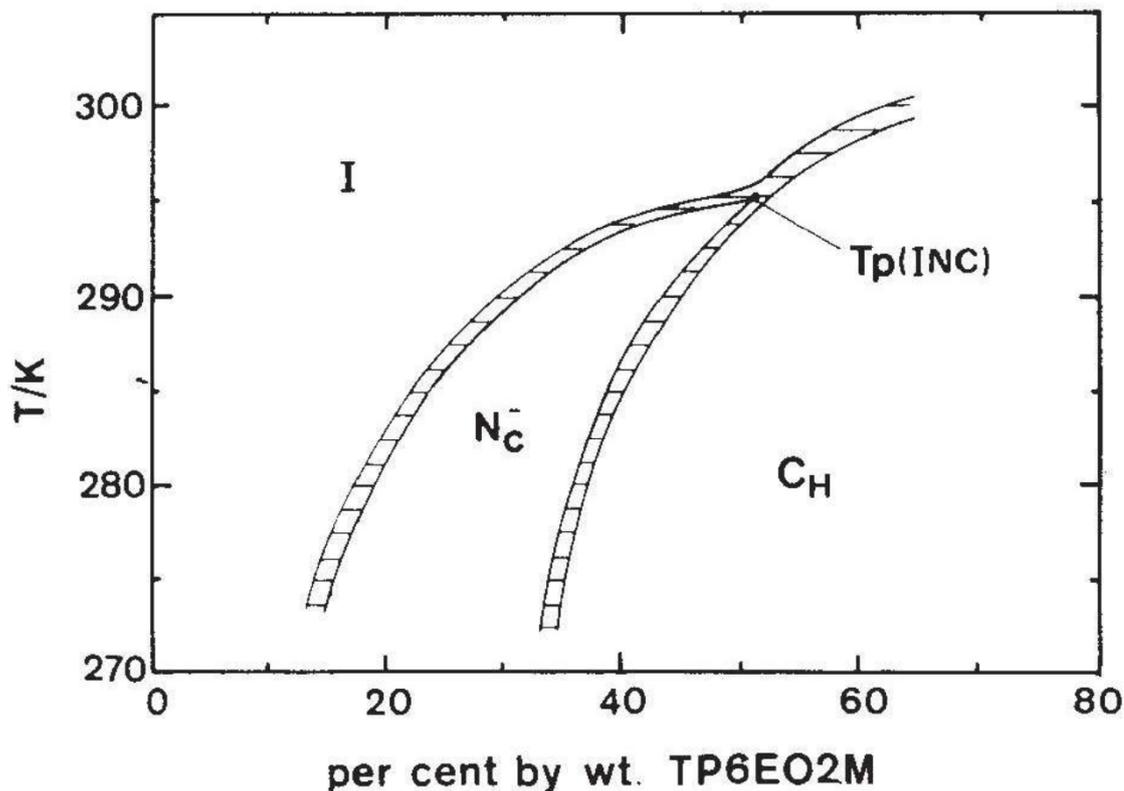
Figure 5.3: Part of the experimental phase diagram for the TP6EO2M/water system, showing the isotropic (I), nematic (N_C) and hexagonal columnar (C_H) phases. Reprinted from Ref. 193 with permission from the Taylor & Francis Group.

to hexagonal, as the concentration increases.[195]

5.1.3 Simulating chromonics

The simulation of chromonic liquid crystals presents a number of challenges. Self-assembly into stacks occurs over periods of tens or, in most cases, hundreds of nanoseconds. Moreover, formation of liquid crystal mesophases, by self-organisation of stacks, takes place over time scales that are considerably longer. In two recent studies, progress has been made in understanding and validating the structures expected in short chromonic stacks via long atomistic simulations.[86,200] There have also been previous attempts to produce coarse-grained models of chromonics. Chromonic self-assembly and phases have been seen in simple disc models[201,202] and, at longer length scales, dissipative particle dynamics studies have recently demonstrated the first full simulations of chromonic liquid crystal phase diagrams using up to 5000 mesogens in water.[28,41]

However, there is a real challenge to produce coarse-grained models that capture the key structural features responsible for the local molecular packing within

stacks and yet are computationally tractable, allowing longer time scale events to be seen. Such a model would provide a bridge between atomistic models and simple phenomenological models and would potentially shed light on the mechanisms involved in chromonic self-assembly. The chemical specificity of systematic coarse-grained models would also allow comparisons to be made between specific systems, going beyond the general trends which have been observed in generic coarse-grained models.

The parametrisation of systematic coarse-grained models of liquid crystals comes with a range of issues beyond those seen for simple homogeneous systems. Despite the issues surrounding representability and transferability which have been discussed in previous chapters, there is still a relatively large parameter space which will give qualitatively correct liquid behaviour in these simple systems. This is less true for the more complex structures found in liquid crystals. Chromonics present a particular challenge, since the formation of chromonic stacks and the subsequent alignment of these stacks to form liquid crystal phases must both be stabilised. In order for both of these processes to occur, a very delicate balance between the different interactions in the system must be satisfied. Firstly, the hydrophilicity and hydrophobicity of the tails and core must be correct to stabilise the stacking process; secondly, the interactions between stacks, and between stacks and water, must be balanced correctly so that liquid crystal phases are able to form. This has already been seen in an atomistic study of this system by Akinshina *et al.*,[86] where the OPLS force field showed chromonic stacking, while the GAFF force field did not. Previous attempts to parametrise *bottom-up* coarse-grained models of chromonics have been unsuccessful in modelling chromonic stacking,[203] so further work is clearly needed in this area.

5.1.4 Aims

The aim of this chapter will be to investigate methods for the systematic coarse graining of the chromonic liquid crystal TP6EO2M in water. This system was chosen because it has a phase diagram which includes multiple liquid crystal phases (nematic and hexagonal), and because there are experimental and atomistic simulation studies against which results can be tested. In addition to this, there are only three distinct chemical environments (aromatic core, ethylene oxide tails and water)

and there are no strongly charged particles. This takes away some complications which were faced in previous attempts to coarse grain an ionic chromonic,[203] so that the fundamental challenges involved in modelling liquid crystals at the systematic coarse-grained level may be more easily investigated.

In the case of *bottom-up* coarse graining, even carrying out appropriate reference simulations is not straightforward. The ordered structures mean that it is difficult for an atomistic reference simulation to adequately sample all distances and orientations. This is especially true for chromonic liquid crystals; modelling an entire liquid crystal phase, for a sufficient amount of time, would require an extremely large all-atom simulation. Therefore, the feasibility of running a smaller reference simulation, and whether this results in a representative coarse-grained model, will be examined using the HFM method. Methods for achieving a more computationally efficient coarse-grained mapping (implicit water and clustering algorithms) will also be tested and discussed.

Given the difficulty in parametrising *bottom-up* coarse-grained models of liquid crystals, *top-down* coarse graining methods will also be examined. The MARTINI 2 force field, while it is mainly intended for biomolecules, has been used for a wide range of systems, and so its application to the TP6EO2M system will be tested. These results will be compared to results from the new, and currently unpublished, MARTINI 3 forcefield, where four different parametrisations will be examined. Recent work using the SAFT-γ Mie approach will then be discussed as an alternative *top-down* method. The degree of success of each model will be interpreted in terms of its balance of hydrophilic and hydrophobic interactions. The parts of this section which discuss the MARTINI 2 and SAFT-γ Mie models have been previously published by this author in "Development of new coarse-grained models for chromonic liquid crystals: insights from *top-down* approaches".[40] These sections are reproduced and adapted with permission from the Taylor & Francis Group.

Finally, a series of large-scale coarse-grained simulations of TP6EO2M will be carried out using the MARTINI 3 forcefield. The ability of the model to self assemble liquid crystal phases from a dispersed starting configuration will be tested. The transferability of the model across concentrations (in terms of its ability to represent the different chromonic liquid crystal phases) will then be examined, by carrying out a range of simulations with pre-assembled columnar starting configurations.
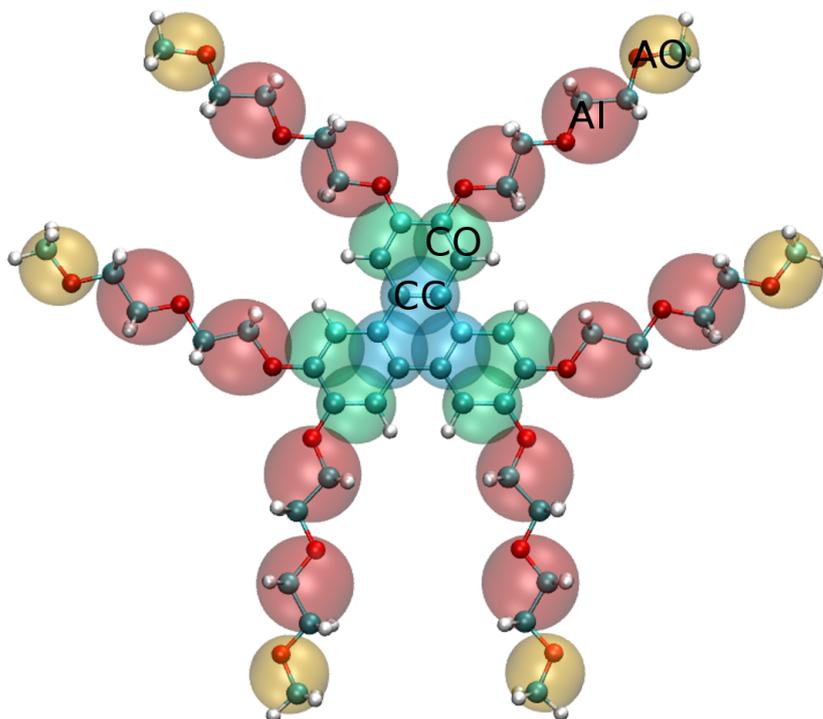
Figure 5.4: Coarse-grained mapping used in this chapter for TP6EO2M.

5.2 Computational methodology

5.2.1 Coarse-grained mapping

The mapping scheme used for TP6EO2M is shown in Figure 5.4. The TP6EO2M model consists of four bead types: the central core (CC), outer core (CO), inner arm (AI) and outer arm (AO). With the exception of the MARTINI 3 model, the CC and CO beads, and the AI and AO beads, use the same interaction parameters, and differ only by their masses. The mapping used for water varies between coarse-grained models. The MARTINI water models map four water molecules to one coarse-grained bead. For the HFM model (with the exception of the models discussed in Sections 5.3.5), a single water molecule was mapped to one coarse-grained bead.

The same set of bonded interactions was used for all of the coarse-grained models of TP6EO2M. Bonds and angles were described by simple harmonic potentials, with bond lengths and angles taken from the energy minimized atomistic structure, mapped onto a coarse grained representation. Six improper dihedrals were required to keep the core of the molecule planar, but no other dihedral interactions were used in this model. Bonded interaction parameters are given in Tables 5.1, 5.2 and 5.3.

Table 5.1: TP6EO2M bond parameters. The CO–CO* bond is along the long edge of the aromatic core (see Figure 5.6.

Bond	r_0 / nm	k_{bond} / kJ mol^{-1} nm^{-2}
CC–CC	0.212	15000
CC–CO	0.212	15000
CO–CO	0.212	15000
CO–CO*	0.424	15000
CO–AI	0.276	10000
AI–AI	0.328	10000
AI–AO	0.282	10000

Table 5.2: TP6EO2M angle parameters.

Angle	θ_0 / deg	k_{angle} / kJ mol^{-1} deg^{-2}
CC–CO–AI	180	85.0
CO–AI–AI	130	85.0
AI–AI–AO	130	85.0

5.2.2   Atomistic reference simulations

Atomistic simulations of TP6EO2M in water were used as references for the HFM model of TP6EO2M. The system was modelled using the OPLS forcefield,[70] with the same parameters used by Akinshina *et al.* to model the same system.[86] The equations of motion were solved using the leap-frog integrator with a time step of 1 fs. All simulations were carried out at a temperature of 280 K and a pressure of 1 bar, using the Nosé-Hoover thermostat[79,80] and the Parrinello-Rahman barostat.[81] The van der Waals, neighbour list and coulomb cutoffs were set to 1.2 nm.

Two separate reference systems were used for different HFM models: an equilibrated chromonic stack of 10 TP6EO2M molecules in 20928 waters (2.5 wt%); and a pre-equilibrium mixture of short stacks of varying sizes, with 50 TP6EO2M and 14433 water molecules (15.3 wt%). The starting configuration for the stacked reference was provided by Martin Walker, and obtained by equilibration of a pre-assembled stack for 300 ps at constant *NVT*, followed by 80 ns at constant *NPT*. These systems were each simulated for 20 ns, and snapshots of positions and forces were taken every 5 ps.
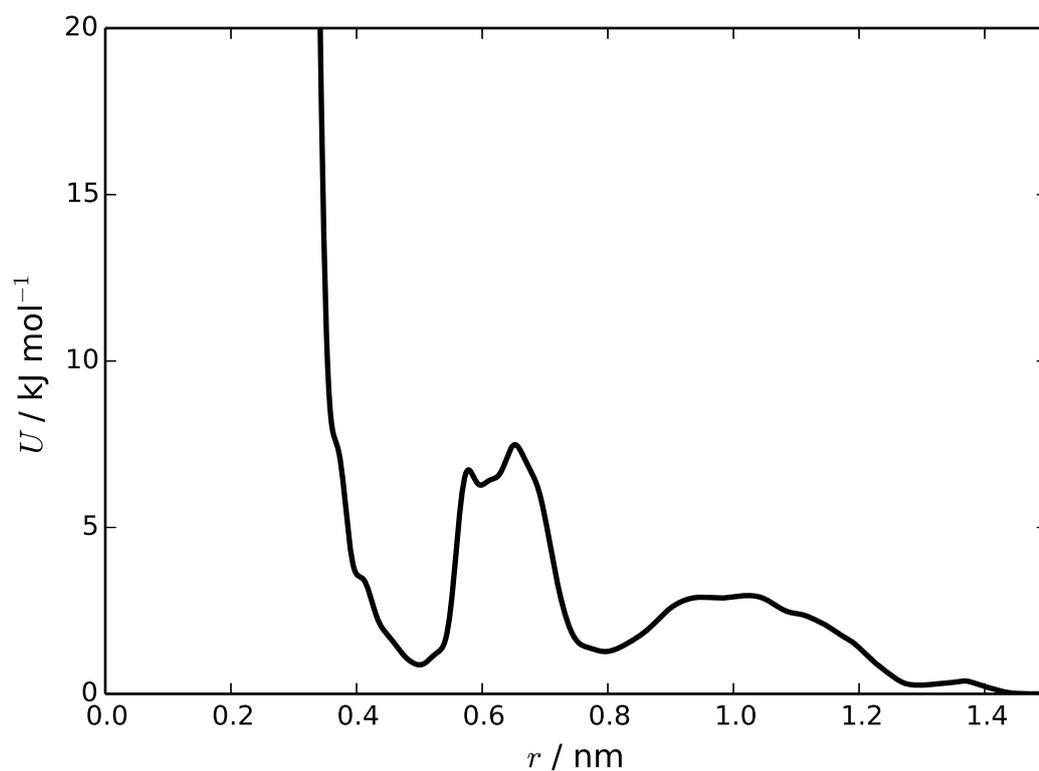
Figure 5.5: C-C interaction potential calculated using HFM, after excluding only coarse-grained bonds and angles from the reference trajectory. The sharp peaks in this are due to 1–4 C-C interactions, and disappear when these pairs are excluded.
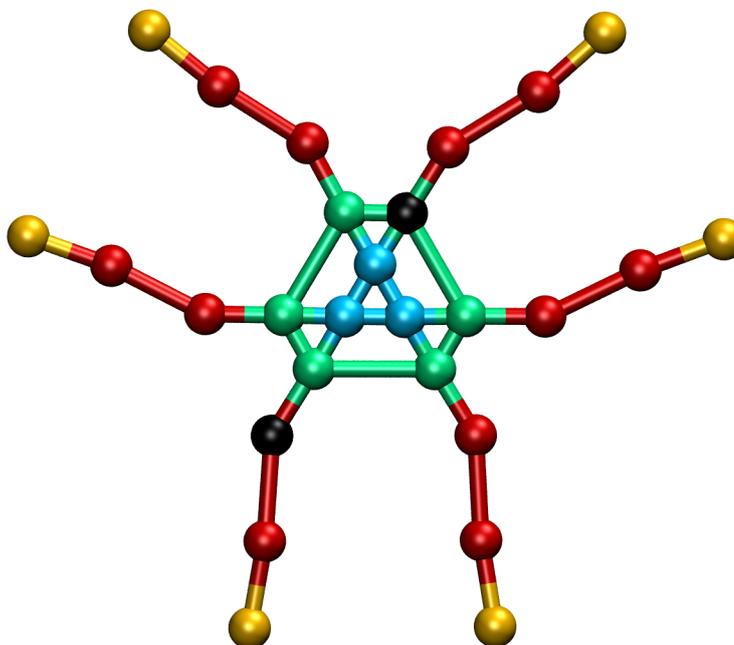

Figure 5.6: Connectivity of the coarse-grained TP6EO2M model. All 1-3 and 1-4 interactions, as well as the 1-5 interactions of the type indicated by the black beads, were excluded from the reference trajectory.

Table 5.3: TP6EO2M improper dihedral parameters.

Dihedral	ϕ_8 / deg	k_ϕ / kJ mol^{-1}
CO–CC–CC–CC	0	30.0

5.2.3 Force matching calculations

Force matching calculations for the TP6EO2M model were carried out using the Bottom-up Open-source Coarse-graining Software (BOCS) package, which implements the normal equation method for force matching shown in Equation 2.42.[31] BOCS was chosen over VOTCA for this system because of its significantly lower memory usage, and the fact that it can be run in parallel; these features make it more feasible to run the larger reference systems which are required for the TP6OE2M system.

The molecular structure of TP6EO2M introduces some complications to the HFM procedure. The rigidity of the aromatic core of TP6EO2M means that, at some distances, the intramolecular interactions dominate the forces in the reference system. The effect of this on the interaction potentials is highlighted in Figure 5.5. A solution to this was to determine which pairs of coarse-grained beads were causing the sharp peaks by examining the effect on the coarse-grained forcefield of excluding certain pairs. The interactions which were found to contribute to the issue were then excluded from the atomistic reference. These include all 1-3 and 1-4 interactions, as well as the 1-5 interaction shown in Figure 5.6. When carrying out coarse-grained simulations, the interactions between these additional excluded pairs were modelled using the non-bonded potentials obtained from the force matching calculations. This is a valid approach in the HFM framework, in which there are no restrictions on which interactions are excluded, and how the excluded interactions are modelled.

A number of HFM models of TP6EO2M are discussed in Section 5.3. The FM-N and FM-Q model use neutral and charged coarse-grained beads, respectively, and differ in how electrostatics are treated in the reference system (details are given in Section 5.3.1). Pressure correction was applied to the FM-Q model, yielding the FM-QP model. The FM-D model uses a dispersed mixture of stack sizes as a reference system, as discussed in Section 5.3.4. Finally, the FM-4S and FM-4N use a 4-to-1 coarse-grained mapping for water, and differ in how the forces on water beads are treated (see Section 5.3.5).

Table 5.4: Non-bonded parameters used in the MARTINI 2 model. 'C' refers to both the CC and CO beads and 'A' refers to the AI and AO beads. P4 and BP4 refer to standard and antifreeze water particles, respectively.

Interaction	σ / nm	ϵ / kJ mol^{-1}
C–C	0.43	2.625
A–A	0.43	3.375
C–A	0.43	2.325
P4–P4	0.47	5.000
BP4–BP4	0.47	5.000
P4–BP4	0.57	5.600
P4–C	0.47	2.700
P4–A	0.47	4.000
BP4–C	0.47	2.700
BP4–A	0.47	4.000

5.2.4 MARTINI parametrisation

MARTINI 2 model

The bead types for the MARTINI 2 model were taken from the MARTINI benzene[32] and polyethylene oxide[91] models, which use the SC4 and SNa bead types. Here, "S" denotes a small bead compared to the MARTINI standard of 4 heavy atoms, and "a" denotes a hydrogen-bond acceptor. MARTINI water (P4) with antifreeze particles (BP4) was used, as this has been shown to prevent problems with the crystallisation of water in MARTINI simulations. BP4 has the same interactions with all other MARTINI bead types as P4, with the exception that the P4-BP4 interaction has larger σ and ϵ values, disrupting the formation of any crystal structure. Lennard-Jones parameters were obtained by consulting the interaction matrix provided in the original MARTINI paper. For interactions between two S beads, σ was reduced to 0.43 nm and ϵ scaled by a factor of 0.75, according to the rules provided.[32] Table 5.4 contains the Lennard-Jones parameters for the interactions in this system.[*]

MARTINI 3 model

The MARTINI 3 forcefield has not yet been formally published, and so full details of its parametrisation are unfortunately not yet available; however, it is currently in

---

[*]Note that this description of how the interactions were determined has been corrected from the original publication.[40]

open beta, and parameters (and some documentation) are freely available online.[204] This new version of MARTINI includes more parameters for different levels of coarse-graining. The standard mapping is, as in MARTINI 2, four heavy atoms to a coarse-grained bead. However, for each standard bead type, there are separate parameters for beads with three and two heavy atoms, denoted by S (small) and T (tiny) prefixes in the bead name. The MARTINI 3 model for TP6EO2M, therefore, uses different Lennard-Jones parameters for the AO and AI beads. MARTINI 3 no longer uses antifreeze water particles, instead having one water bead (WN) which, again, represents four water molecules. The interactions between water beads are weaker than in MARTINI 2 (see Tables 5.4 and 5.5, which seems to have lowered the melting point of water.

The MARTINI 3 open beta documentation includes suggestions for the mapping of different chemical groups to bead types, which were used to determine bead types for the TP6EO2M model. The TC4 bead type was chosen for the aromatic core. For the arms, four bead types were considered: N1, N0, N1a and N0a (N denotes the non-polar bead type, where N0 has a lower polarity than N1, and 'a' denotes a hydrogen bond acceptor). For the purposes of this molecule, the N1-WN interaction has a higher ϵ than the N0-WN interaction, so the N1 bead is more soluble in water. Beads with the 'a' descriptor have both weaker self-interactions and weaker interactions with water than their non-hydrogen bonding counterparts. The Lennard-Jones parameters used in the MARTINI 3 models are given in Table 5.5. Throughout this chapter, the MARTINI 3 models will be referred to according to bead type used for the arms.

5.2.5 Coarse-grained simulations

All molecular dynamics simulations were carried out using GROMACS 4.6.7. The equations of motion were solved using the leap-frog integrator with a time step of 2 fs. All simulations were carried out at a temperature of 280 K using the Nosé-Hoover thermostat. Constant NPT simulations were carried out at a pressure of 1 bar using the Parrinello-Rahman barostat. The van der Waals, neighbour list and Coulomb cutoffs were set to 1.2 nm (MARTINI) and 1.5 nm (HFM and SAFT). More specific details of the simulations carried out can be found in the relevant sections.

Table 5.5: Non-bonded parameters used in the MARTINI 3 models. 'TC4' was used for both the CC and CO beads, 'WN' for water and either 'N0', 'N1', 'N0a' or 'N1a' for the AO (T prefix) and AI (S prefix) beads.

Interaction	σ / nm	ϵ / kJ mol^{-1}			
		N0	N0a	N1	N1a
C–C	0.330	1.45	1.45	1.45	1.45
AO–AO	0.330	1.70	1.45	1.70	1.45
AI–AI	0.400	2.54	2.11	2.54	2.11
AO–AI	0.370	2.07	1.57	2.07	1.57
C–AO	0.330	1.20	1.20	1.20	1.20
C–AI	0.370	1.57	1.57	1.57	1.57
W–W	0.470	4.65	4.65	4.65	4.65
W–C	0.411	1.19	1.19	1.19	1.19
W–AO	0.41	2.40	2.15	2.67	2.40
W–AI	0.440	3.02	2.81	3.23	3.02

Simulations using the MARTINI model generally use time steps of up to 20 fs, which further improves the computational speed-up of the model. However, in this case, the highly interconnected bonded structure of the aromatic core resulted in very unstable molecular dynamics when time steps of larger than 2 fs were used, and so this smaller value was chosen for the TP6EO2M simulations. While this removes one advantage of coarse graining, the model still represents a significant speed-up compared to the atomistic model, as shown in Section 5.5.

5.2.6 Potential of mean force (PMF) calculations

The PMF for the separation of a TP6EO2M dimer in aqueous solution was calculated for each coarse-grained model by constraining the separation distance between the centres of mass (COMs) of the cores of two TP6EO2M molecules at a range of distances and calculating the average constraint force at each distance. Only the distance between the COMs was constrained, the molecules were allowed to rotate freely during the calculations. The potential of mean force was then calculated at each distance according to Equation 2.29.

For the MARTINI 2 model, a dimer was placed in a simulation box and solvated with water, and an energy minimisation was carried out. The system was equilibrated in the constant NPT ensemble for 1 ns with the Berendsen thermostat and barostat, followed by 5 ns with the Nosé-Hoover thermostat and the Parrinello-

Rahman barostat. A further 5 ns constant NVT equilibration was carried out at the equilibrium density from the constant NPT simulations. An initial pull simulation was carried out, with the distance between the COMs of the aromatic cores of the two molecules constrained and increased at a rate of $0.01\,\mathrm{nm\,ps^{-1}}$. A short push simulation was also carried out to generate structures with shorter intermolecular distances. Frames were then selected from these simulations representing a range of distances, clustered more closely around the expected energy minimum, and these were used as starting structures for the constrained simulations. The MARTINI 3 PMF calculations used the same set of starting structures which were generated using the MARTINI 2 model. At each distance, a 25 ns simulation was carried out with COM–COM distance constrained, the last 20 ns of which was used to generate data for the PMF calculation.

To generate the HFM starting structures, a similar procedure was used. The unconstrained equilibration of the dimer was carried out using the atomistic model. Coarse-grained mapping was then carried out, and the pull simulations were carried out using the HFM model.

5.2.7 Structural analysis of liquid crystal structures

The structures of liquid crystal phases were analysed in two ways. The nematic order parameter, S_{nematic}, was calculated to determine the degree of orientational order of the system. This is given by:

$$S_{\mathrm{nematic}} = \left\langle \frac{3\cos^2\theta - 1}{2} \right\rangle, \tag{5.1}$$

where θ is the angle between a specified vector, $\mathbf{d}$, for one liquid crystal molecule and the average value of $\mathbf{d}$ for that snapshot.[178] A value of 1.0 indicates a totally ordered system, and the value decreases with the amount of order in the system. For a TP6EO2M molecule, $\mathbf{d}$ was defined as perpendicular to the plane of central aromatic core, and was calculated from the three CC beads i, j and k, using:

$$\mathbf{d} = \mathbf{u}_{ij} \times \mathbf{u}_{jk}. \tag{5.2}$$

This means that S_{nematic} measures the alignment of that molecular axis, although in principal any vector calculated from the molecular structure could be chosen.

In practice, S_{nematic} is calculated by diagonalisation of the ordering tensor:

$$Q_{\alpha\beta} = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{3}{2} \hat{\mathbf{d}}_{i\alpha} \hat{\mathbf{d}}_{i\beta} - \frac{1}{2} \delta_{\alpha\beta} \right), \tag{5.3}$$

where α and β denote components in the x, y and z directions and N is the number of molecules. S_{nematic} is the largest eigenvalue of $\mathbf{Q}$.[205]

The degree of positional order was analysed using the two-dimensional pair distribution function, $g(u,v)$, of pairs of TP6EO2M molecules. For this quantity, the system director for a frame was calculated by taking the global average of $\mathbf{d}$. Two orthogonal vectors, $\mathbf{u}$ and $\mathbf{v}$, were then defined normal to the director. u and v were calculated by projecting the distance between a pair of molecules (calculated from the centres of mass of the cores) along these two vectors. A vertical cutoff along the global system director of 0.5 nm was applied so that only pairs within a thin segment of the box were considered.

5.3 Hybrid force matching (HFM) results

5.3.1 Treatment of electrostatics

The models discussed in this section were calculated from 2000 frames (corresponding to 10 ns) of the stack reference simulation.

The initial HFM forcefield for TP6EO2M/water, referred to as the FM-N model, is shown in Figure 5.7. This model contains a number of interesting features. The interactions involving water all look as expected. The shape of the water self-interaction resembles previously published force matched waters. The water-arm and water-core interactions also make physical sense; looking at these interactions, the arms appear more soluble than the cores due to their deeper and wider potential wells. The interactions between the arm and core beads, however, make less intuitive sense. The arm-arm and core-core interactions are both largely repulsive, with a shallow attractive well. The core-arm interaction, on the other hand, is very attractive, with a double well shape. This is surprising because, intuitively, the self interactions would be expected to be attractive, particularly the core-core interactions.

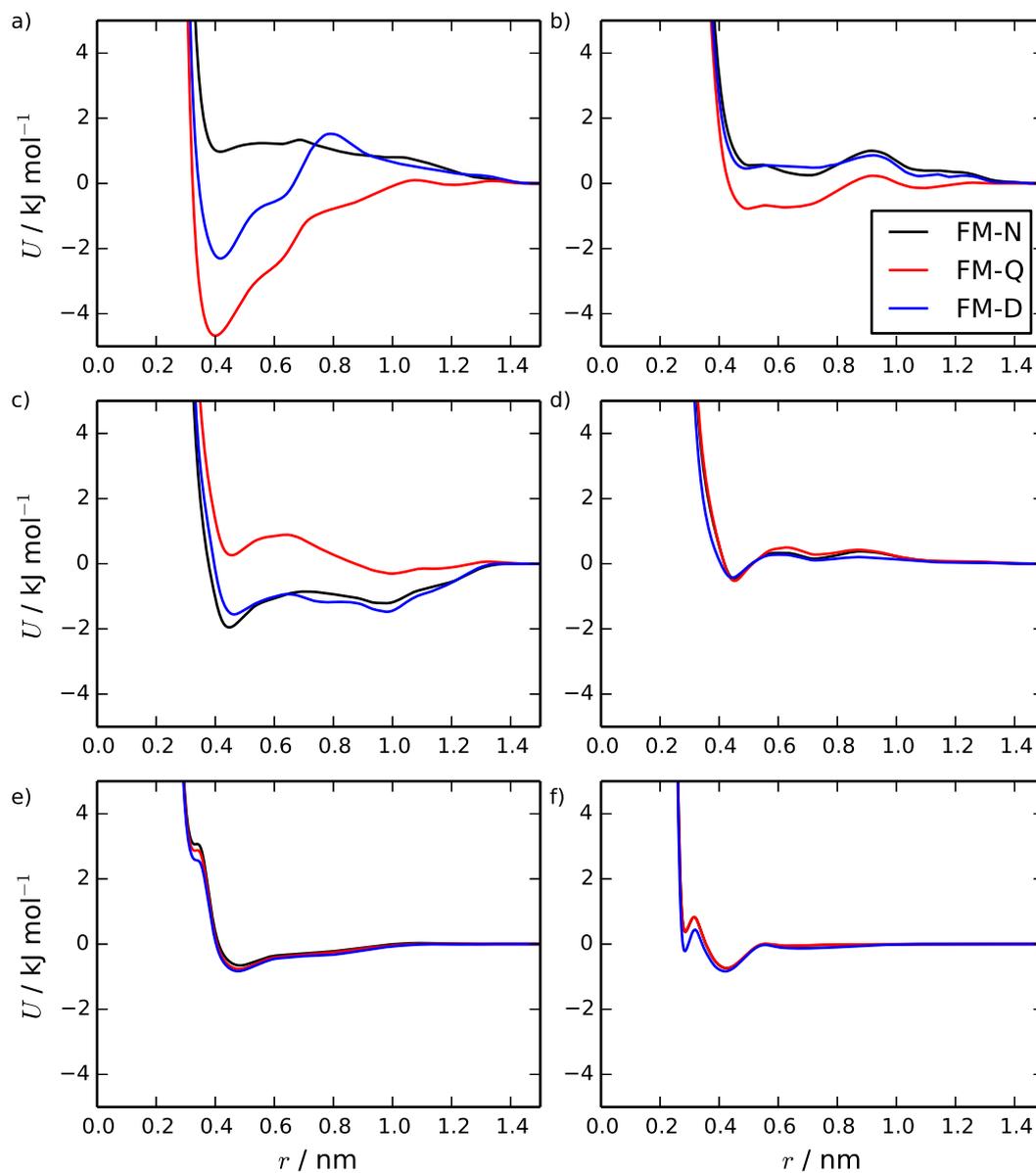The unexpected shape of the potentials in the initial HFM model comes from

Figure 5.7: HFM potentials calculated for the TP6EO2M/water system at 280 K, for the FM-N and FM-Q models (Section 5.3.1, as well as the FM-D model (see Section 5.3.4). Showing the a) C-C, b) A-A, c) C-A, d) C-W, e) A-W and f) W-W potentials.

the way in which electrostatics were treated in the parametrisation process. It is being assumed that, like the MARTINI models, all of the coarse-grained beads in the HFM model are neutral. However, when the charges from the atomistic model are carried over to the coarse-grained model, the AI and CC beads are neutral, but the AO and CO beads have charges of $-0.2e$ and $+0.2e$ respectively. The electrostatic forces between the charged beads are being included in the force matching, and then effectively averaged over all the beads; this is origin of the repulsive like interactions and the attractive unlike interactions. This is incorrect for two reasons: the electrostatic interactions should apply only to the AO and CO beads, not all of the beads; and long-ranged electrostatic interactions are being implicitly included in the short-range vdW potentials, and so not treated correctly in the coarse-grained simulations.

To address this issue, an additional step was added to the preparation of the reference. For each frame of the mapped trajectory, the electrostatic interactions coming from the mapping of the atomistic charges to the coarse-grained beads were calculated, giving the coarse-grained electrostatic forces for the system. The electrostatic forces were then subtracted from the reference forces (after the exclusion of intramolecular interactions), giving a new reference trajectory which does not include the coarse-grained electrostatic forces. The new reference was then used for force matching. Coarse-grained simulations of this model included explicit charges on the AO ($-0.2e$, where e is the electronic charge) and CO ($+0.2e$) beads, and the electrostatic forces were calculated using the PME method. It should be noted that this method can only deal with electrostatic interactions between atoms which map onto charged beads; this neglects, for example, any electrostatic interactions involving water, which is neutral overall. However, this is still expected to be an improvement over the FM-N model.

The FM-Q model potentials are shown in Figure 5.7. The interactions involving water are very similar to those in the FM-N model, since the water beads are neutral in both models. However, the other interactions have changed significantly. The core-core interaction is now strongly attractive, the arm-arm is more weakly attractive and the core-arm interaction is overall slightly repulsive, but with two small attractive wells. Comparing the two models, the dominant influence of the implicit electrostatic interactions on the potentials of the FM-N model are apparent.

The two coarse-grained models exhibit rather different behaviour in simulations.
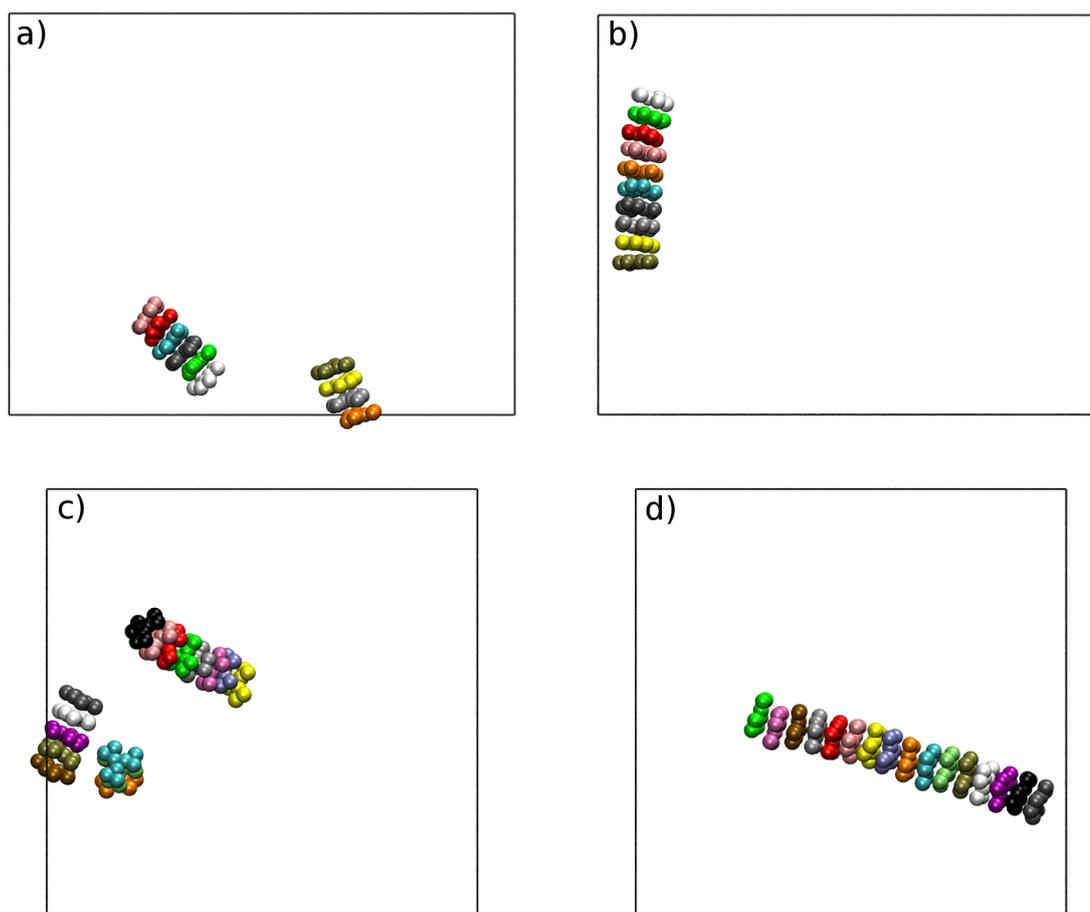
Figure 5.8: Snapshots taken from simulations of 10 (a and b) and 16 (c and d) TP6EO2M molecules, from dispersed starting configurations, using the FM-N (a and c) and FM-Q (b and d) models.

When a stack of 10 molecules (2.5 wt%) is simulated, it remains stable over a 100 ns simulation in both cases. However, the self-assembly of this stack occurred over very different timescales for the two models. Figure 5.8 shows snapshots from simulations using the two models, starting from dispersed configurations. In the simulation of the FM-N model, the assembly of two short chromonic stacks was observed within 35 ns, and the formation of the full 10 molecule stack took 100 ns of simulation time. The FM-Q model, on the other hand, was able to produce a 10 molecule chromonic stack within only 4 ns of simulation time. The difference between the two models was more pronounced when self-assembly at a higher concentration (16 molecules at 4.7 wt%) was attempted. Here, the FM-Q model formed a 16 molecule stack within 20 ns of simulation. The FM-N model was only able to form several smaller stacks, which often formed then unformed later in simulation, over 100 ns. Additionally, at several points, side-on aggregation of two stacks was observed, and was stable over 10s of ns.

Therefore, not only do the interaction potentials of the FM-Q model make more physical sense, but self-assembly is more reliable, and occurs over shorter timescales than the FM-N model. The charge-subtraction methodology appears to be useful for dealing with even weakly charged coarse-grained beads, and it is anticipated that it would be of even more use for more polar or ionic systems.

5.3.2 Structure

The partial RDFs calculated using the FM-Q model are plotted in Figure 5.9. There are significant quantitative differences between them and the atomistic RDFs. However, in most of the RDFs, the overall shape and peak positions are still qualitatively captured. The exception to this is the C-W RDF, where the atomistic and FM-N models give noticeably different peak patterns; this interaction is less well-sampled than the others in the stacked configuration, so the HFM interaction potentials may be less representative.

Two distinct measures of the distance between two molecules in a stack have been used in previous studies of TP6EO2M. The first, d_{com}, is simply the distance between the centres of mass of two molecules, while the stacking distance, d_S, is the distance between the molecules projected along the average of the vectors normal to the cores of the two molecules. Using both measures together provides insights into the nature of the intermolecular stacking; distributions of the two quantities will show different features depending on, for example, whether there are offsets between adjacent molecules, or bends in the stack. These quantities were calculated for the TP6EO2M stack simulated using the atomistic and FM-Q models, and these are plotted in Figure 5.10. The centres-of-mass of the aromatic cores were used for calculating distances, for consistency with previous atomistic studies of the system.[86,198] The positions of the first peaks in both distributions are well represented by the FM-N model; the model gives a d_S of 0.370 nm and d_{com} of 0.380 nm, which compare well to the atomistic values of 0.365 nm and 0.375 nm. The FM-N distributions do not have the pronounced tails present in the atomistic distributions, indicating that the molecules within the stacks are shifted by less relative to their neighbours. The small second peaks in the atomistic distributions represent bends within the stack structure, as illustrated in Figure 5.11; these peaks are not present in the FM-Q distribution. The shifts and bends present in the coarse-grained starting structure
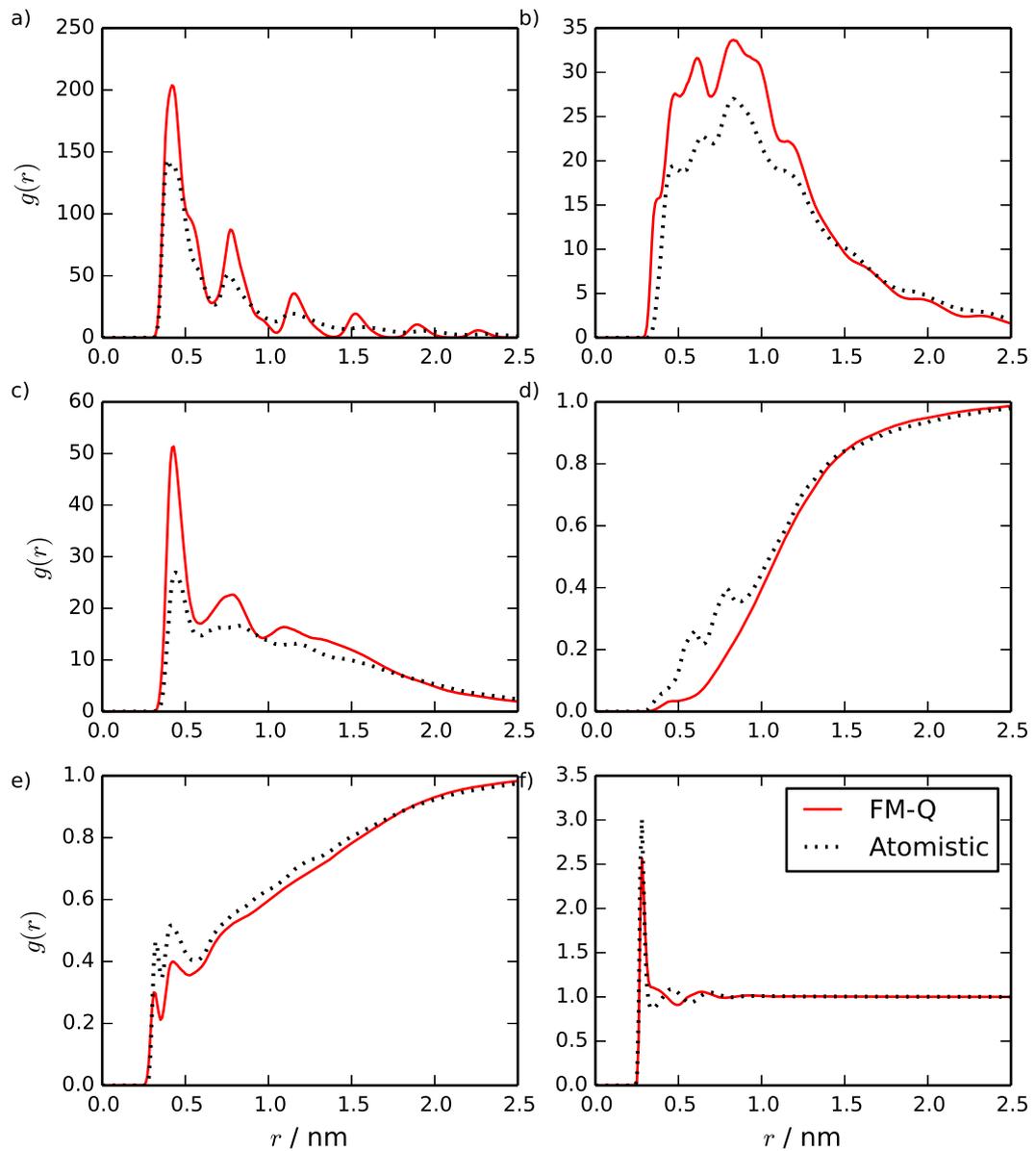
Figure 5.9: Site-site intermolecular RDFs calculated from a stack of 10 TP6EO2M molecules, using the FM-Q and atomistic models: a) C-C, b) C-A, c) A-A, d) C-W, e) A-W and f) W-W distributions.
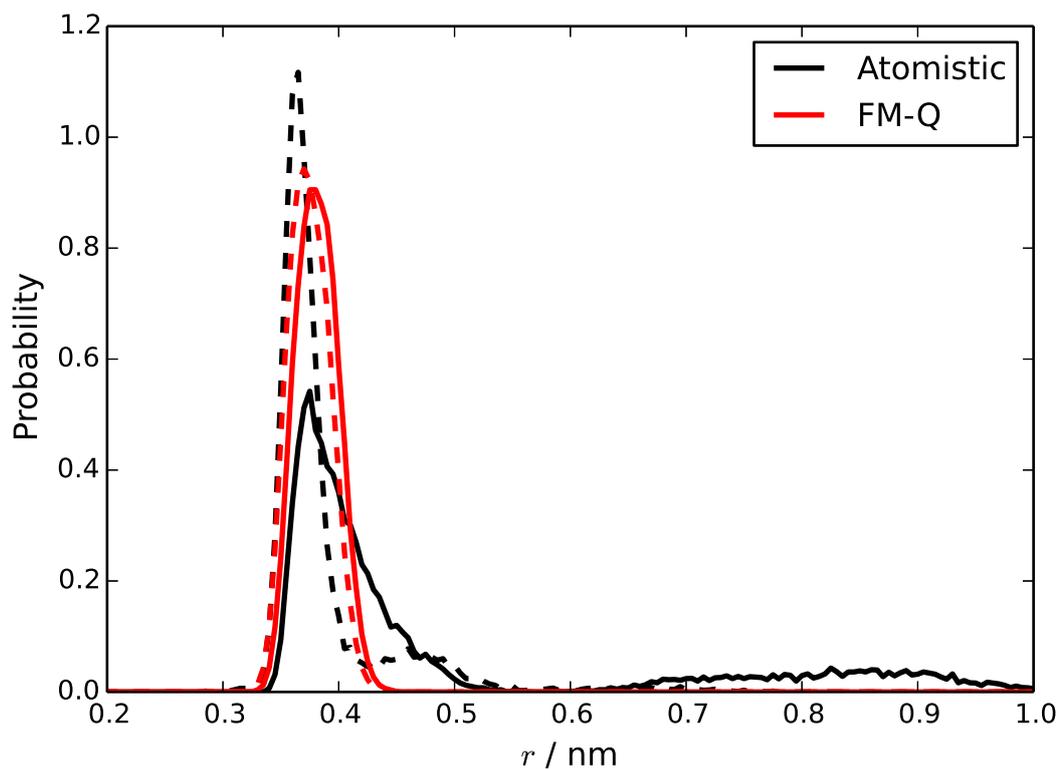
Figure 5.10: Distributions of d_S (dashed lines) and d_{com} (solid lines) for a stack of 10 TP6EO2M molecules, calculated using the atomistic and FM-Q models.
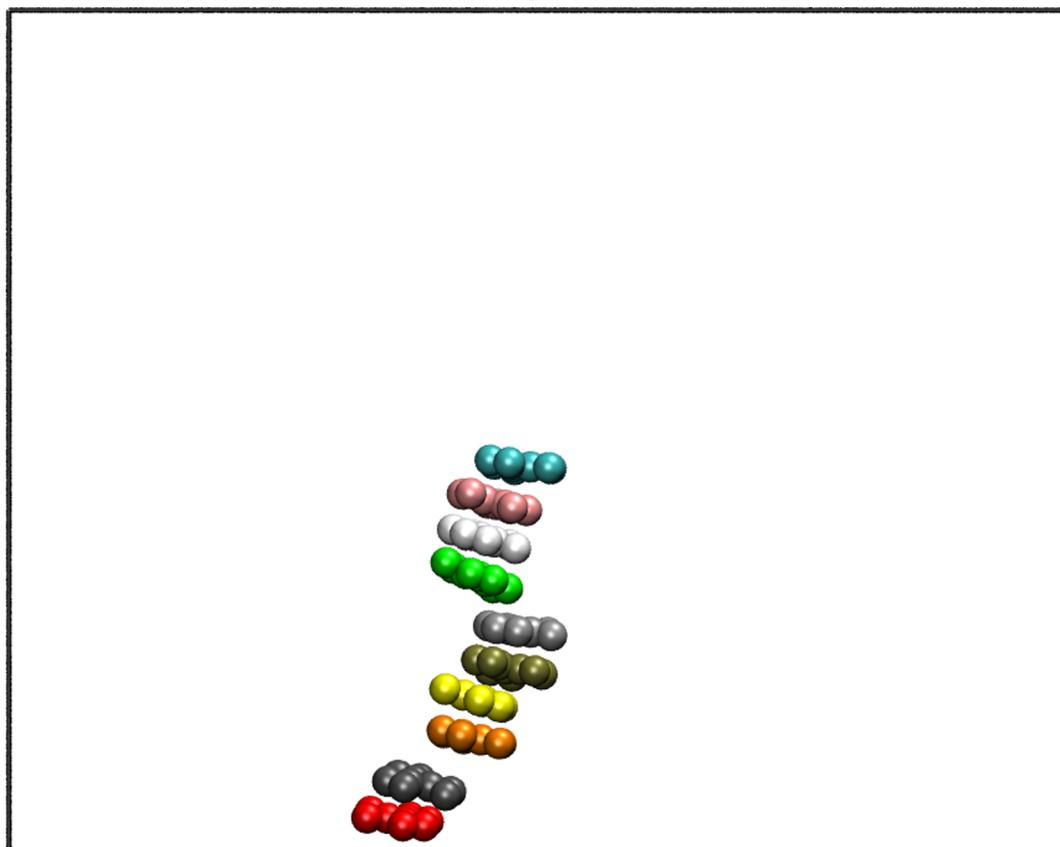

Figure 5.11: Starting structure for the atomistic reference simulation, showing only the cores mapped to a coarse-grained representation. This mapped structure was used as the starting structure for coarse-grained simulations.
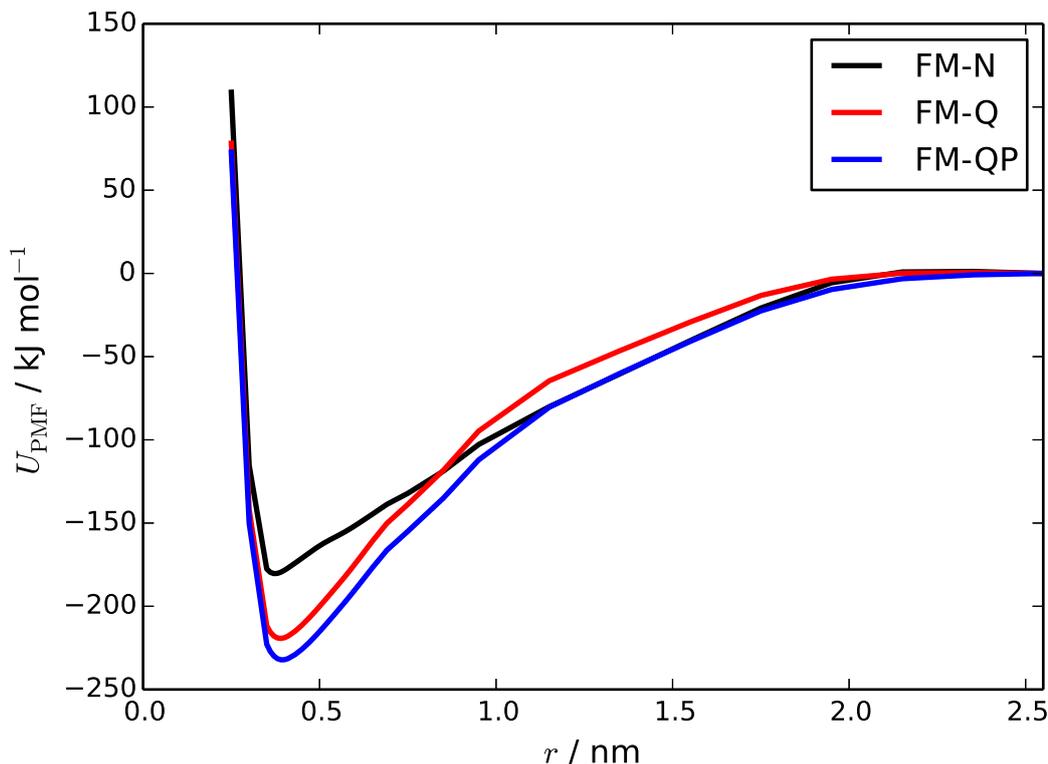
Figure 5.12: PMFs for the formation of a TP6EO2M dimer, calculated using the FM-N, FM-Q and FM-QP models.

disappeared within 10 ps of simulation, so this aspect of the structure is not well represented by the FM-Q model.

Overall, the structural representability of the FM-Q model is not as good as the the HFM models for the octane/benzene mixtures discussed in Chapter 3. However, this system is much more structurally complex than an isotropic liquid, so the level of structural accuracy achieved is still encouraging. In particular, the molecules are able to form stable stacks with accurate stacking distances, which is an improvement over previous attempts to apply *bottom-up* coarse graining to chromonic liquid crystals.[203]

5.3.3 Thermodynamics

The PMFs calculated using the FM-Q and FM-N models are shown in Figure 5.12. The free energy of association, ΔG, calculated from the FM-Q PMF is -219 kJ mol^{-1}, which is more than $5\times$ the atomistic value. For the FM-N model, ΔG is -180 kJ mol^{-1}; this is lower than the FM-Q value, but still significantly higher than the atomistic value. The difference between the two models is consistent with

the improved face-to-face aggregation in the FM-Q model, but clearly neither model represents the thermodynamics of the system particularly well.

Since pressure correction was found to help with the representability of free energy changes in Chapter 3, the linear pressure correction was applied to this system, using the methodology in Section 3.2.2; the resulting model is called FM-QP. It should be noted that, while pressure correction was carried out for the 10 molecule stack system, the FM-QP model also represents the pressure of the dimer system well. However, the PMF calculated using the pressure-corrected was no more accurate than the non-pressure-corrected PMF; the calculated ΔG was in fact more negative, at $-232 \mathrm{~kJ~mol}^{-1}$.

5.3.4 Alternative reference systems

One possible reason for the poor representation of the thermodynamics of dimer formation in the force-matched models of TP6EO2M system is that there are many intermolecular orientations which are not sampled by the reference system. This is important because, in the dimer simulations constrained to higher distances, a range of different molecular orientations are seen which may not be well described by the force field. The use of an alternative reference simulation of a mixture of shorter stacks in solution was tested, where it was hoped that more distances and orientations would be sampled in the interactions between the stacks. The starting structure for this reference was 50 monomers dispersed in water, which were allowed to assemble into short stacks over 20 ns. It should be noted that, at this concentration, the formation of several longer stack would be expected to occur eventually,[86] so this reference is not sampling the equilibrium of the system.

The HFM model calculated using this reference (referred to as FM-D, see Figure 5.7 for the interaction potentials) was not successful in simulating a long chromonic stack. Simulations from a dispersed starting structure resulted in a mixture of monomers and short stacks (similar to the reference simulation), and a pre-assembled stack broke apart to form an aggregate of two stacks joined by their arms. Comparing the FM-Q and FM-D interaction potentials, the FM-D has a less attractive core-core interaction, but the other interactions have similar strengths. This has the effect of destabilising face-to-face stacking in favour of other aggregate structures. It appears that, for the long stack to be stable within a *bottom-up* coarse-grained

model, there must be a long stack present in the reference; this highlights the problem of chemical transferability in *bottom-up* coarse-graining, particularly where the system is difficult to simulate atomistically.

For the coarse-grained model to be truly representative of the system, it must be able to capture even those configurations which are rarely seen. A possible solution to this could be a multi-state parametrisation, where both a stack and a dispersed system are used as references. This could be combined with the local-density potentials discussed in Chapter 4, which in this case could be used as a metric for whether a molecule is a monomer, at the end of a stack or in the middle of a stack, and alter the interaction with other molecules accordingly. A related approach has been used for a hydrophobic polymer, where the LD potential described whether a particular region of a polymer was part of an aggregate.[143]

5.3.5 Using clustering algorithms for a coarser water model

One major disadvantage of the HFM model compared with *top-down* models is the coarse-grained mapping scheme for the water, which is at a 1-to-1 level. Water makes up a large proportion of the TP6EO2M system, so this finer mapping reduces the computational efficiency of the model significantly. The standard centre-of-mass mapping method used so far in this thesis requires the assignment of atoms to specific beads, and this assignment does not change over a reference trajectory. Since molecules which begin in close proximity may move a long distance apart over the course of a simulation, it is not trivial to map multiple molecules to a coarse-grained bead. This is not an issue when using *top-down* methods because, although the coarse-grained beads still represent specific groups of atoms, there is usually no need for direct mapping between atomistic and coarse-grained configurations.

If a coarser mapping is required for water (or any other small molecule), the mapping must instead by carried out on a frame-by-frame basis. The K-Means algorithm is a clustering algorithm which was applied to frames of an atomistic trajectory to assign solvent molecules to coarse-grained beads consisting of multiple molecules.[206,207] In this case, it is guaranteed that all molecules belong to a coarse-grained bead, but not that all beads consist of an equal number of molecules.

K-Means clustering has so far only been applied only to structure-based coarse graining methods. It was anticipated that the use of force-matching may be more

problematic for the K-Means algorithm, because the differing compositions of the beads would result in wildly different forces on each bead when the reference was mapped to a coarse-grained representation. Nonetheless, the algorithm was tested, initially on a pure water system, then on the TP6EO2M/water mixture. Two methods for calculating the forces on the water beads were tried: summing the forces on the constituent molecules as in standard force-matching (the FM-4S model); and normalising the summed forces according to the number of water molecules in that cluster (the FM-4N model). For the FM-4N model, the normalised force on a bead I was calculated using:

$$\mathbf{F}_I = \sum_{i=1}^{n} \frac{4\mathbf{F}_i}{n}, \tag{5.4}$$

where $\mathbf{F}_i$ is the total force on water molecule i and n is the number of molecules mapped to bead I. The force field was then calculated by subtracting the coarse-grained electrostatics from the total forces, as in the FM-Q model. The resulting interaction potentials are plotted in Figure 5.13.

In the case of pure water, a good match was achieved between the structure of the mapped atomistic reference, and the force-matched system, using both force-calculation methods. In the case of TP6EO2M, however, neither method was able to achieve good structural representability, as illustrated by the structures in Figure 5.14. These simulations were run from a stacked starting structure, and in both cases the stack had collapsed within 20 ps of simulation. In the case where the water forces were calculated simply by summing the reference forces, a vacuum region formed in the water around the TP6EO2M molecules. Some of the molecules also contorted into extremely unfavourable configurations (this was confirmed by energy minimisation of isolated contorted structures, which converged back to a planar configuration). Clearly the imbalance in the forces across the different beads in this model has caused significant issues. When the forces were normalised according to bead composition, the 10 molecule stack collapsed into smaller aggregates with more complex stacking behaviour.

Although the models discussed here were not successful in simulating chromonic stacks, the 4-to-1 water mapping does represent a significant computational speed-up over the 1-to-1 mapping. For the 10 molecule stack, speeds of 104 ns/day on 32 cores were achieved, compared to 26 ns/day on 48 cores for the FM-Q force field.
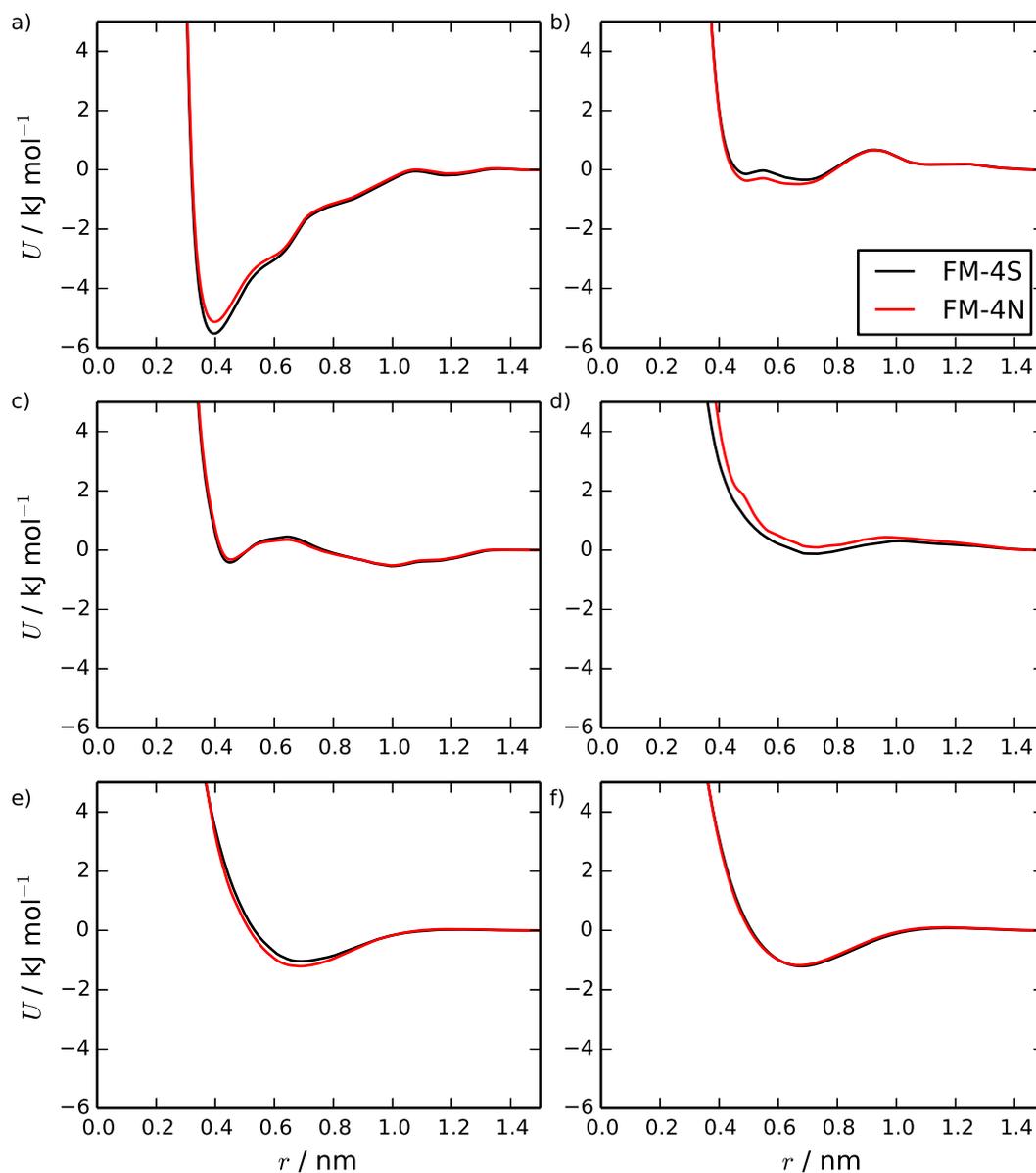
Figure 5.13: HFM potentials for the TP6EO2M/water system, for the FM-4S and FM-4N models. Showing the a) C-C, b) A-A, c) C-A, d) C-W, e) A-W and f) W-W potentials.
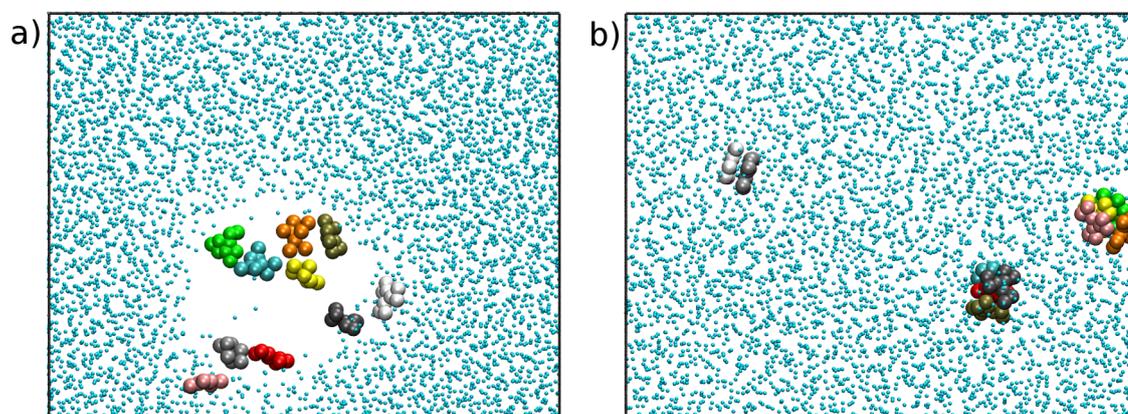

Figure 5.14: Structures after 100 ns simulations using the a) FM-4S and b) FM-4N models, starting from a 10 molecule TP6EO2M stack.

Therefore, the problem of how to map multiple water molecules to 1 coarse-grained bead, with accurate forces, is one which is still worth pursuing. There are two major issues which must be solved.

Firstly, there is a problem with how to calculate the forces on the coarse-grained water beads. Simply summing the forces from all of the atoms is problematic because it results in a large imbalance in the forces on the water beads depending on the number of molecules which were put into each cluster. However, normalising the forces for a four-molecule cluster does not necessarily accurately reflect what the forces on a four-molecule bead at that position would be, and the force calculated may not be consistent with the forces on other beads which are nearby. Neither is a significant problem for the pure water system, but for the more complex stack, the structural representability is affected. The difficulty in mapping the forces on the water molecules affects the balance between the forces from all of the interactions, which causes the resulting model to exhibit unphysical behaviour. This problem may be solved by applying an efficient clustering algorithm which ensures that all molecules are grouped into equally sized clusters. One method which has been published is the CUMULUS coarse-graining method, which applies a Monte Carlo algorithm to assign molecules to beads, although this method does not guarantee that every molecule in the reference is mapped to a coarse-grained bead.[208]

The second issue relates to the enthalpy/entropy balance required for chromonic stacking. The formation of a TP6EO2M stack relies on entropic and enthalpic contributions of roughly equal magnitude.[86] In going from a 1-to-1 to a 4-to-1 water mapping, the entropy of the water is reduced, which will affect any process with an entropic driving force. This is true for all coarse-grained models, and the reduced entropy term is generally balanced by a larger enthalpy term so that the overall free energy of the system is correct. For a *bottom-up* model to be successful, it is important that this is achieved; in this case it is not clear that the entropy lost by further coarse graining of water is being balanced by an increased enthalpy term.

5.4 *Top-down* results

5.4.1 MARTINI 2

Figure 5.15 shows a series of snapshots taken from an initially dispersed system of TP6EO2M in water, simulated using the MARTINI 2 model. The system spontaneously self-assembles over time to form short chromonic stacks, with aggregates growing through addition of a monomer or merging of two short stacks. If a stack is pre-assembled and solvated, it is stable for long simulation times (> 100 ns). The structure of the self-assembled stacks is similar to that observed in the HFM models discussed above.
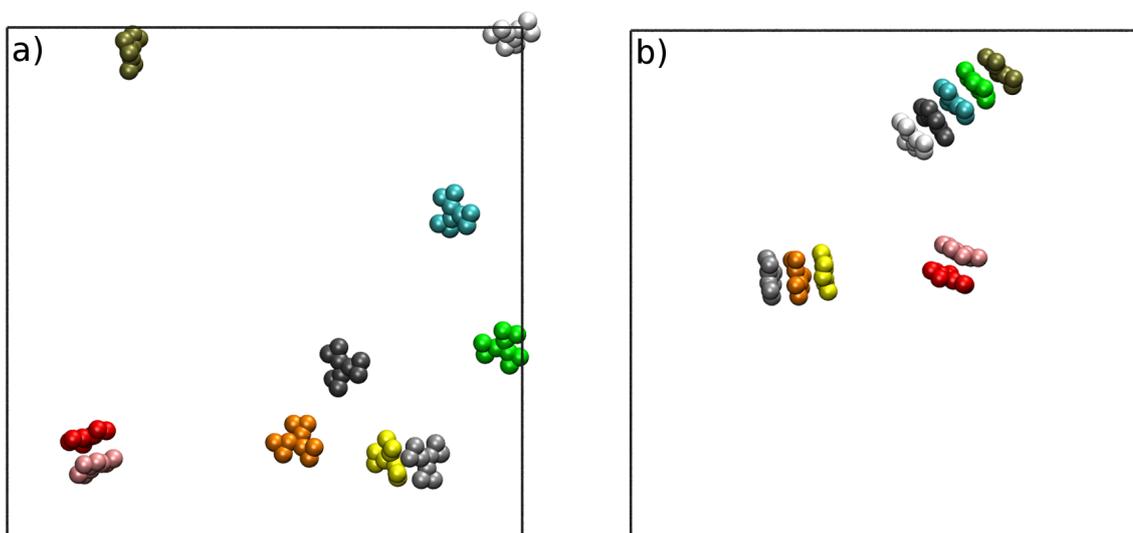

Figure 5.15: Snapshots taken from the simulation of the MARTINI model of TP6EO2M in water at 280 K, showing the central aromatic core of molecules, taken at a) the start of the simulation and b) 90 ns. The simulation trajectory shows the spontaneous self-assembly of short chromonic stacks.

Probability distributions for d_S and d_{com} are shown in Figure 5.16. The maximum of the peak d_S for adjacent molecules is 0.440 nm. The distribution for d_{com} is slightly shifted, so that the maximum is at 0.445 nm. This reflects the fact that there is a slight offset between molecules in the stack. This stacking demonstrates H-aggregation, rather than J-aggregation. The latter would have a larger systematic offset. As expected, the stacking distances between molecules are close to the contact distance of the MARTINI non-bonded potentials for the CG-sites. These separations are the minimum possible for this model due to the size of beads used, but are (of course) slightly larger than the experimental and atomistic value of ~ 0.35 nm, which correspond to the distance of separation for single aromatic carbon atoms.
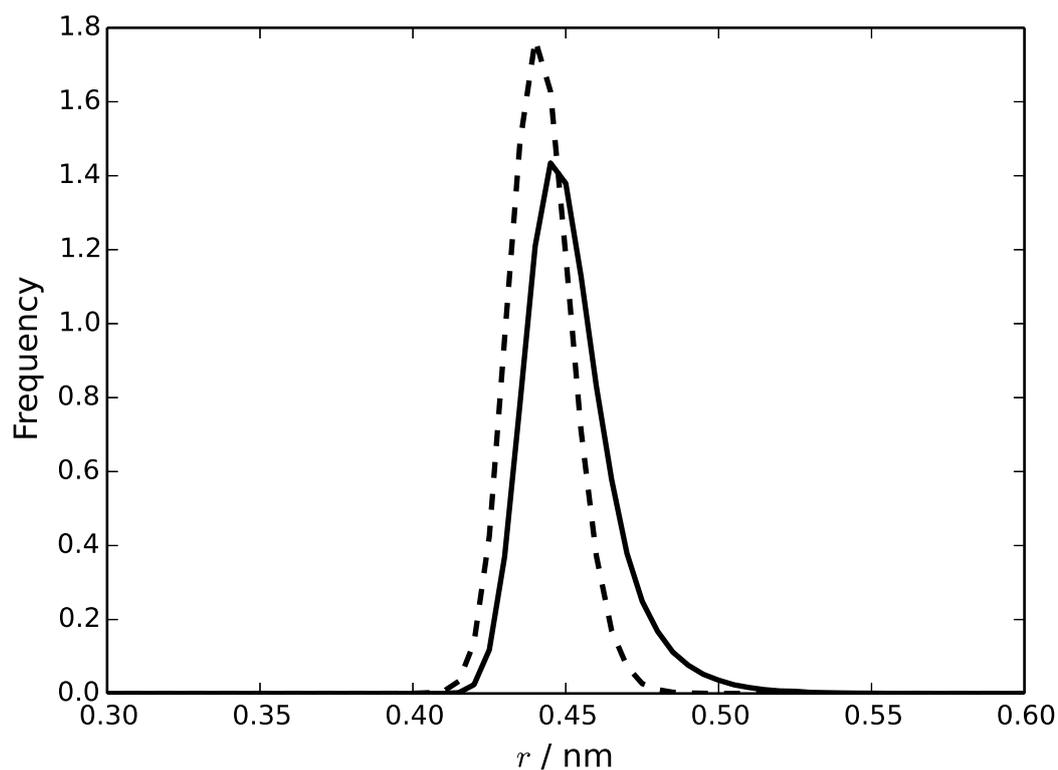
Figure 5.16: Probability distributions for the pair distances d_{com} (solid line), and d_S (dashed line) for a single chromonic stack simulated at 280 K using the MARTINI 2 model for TP6EO2M.
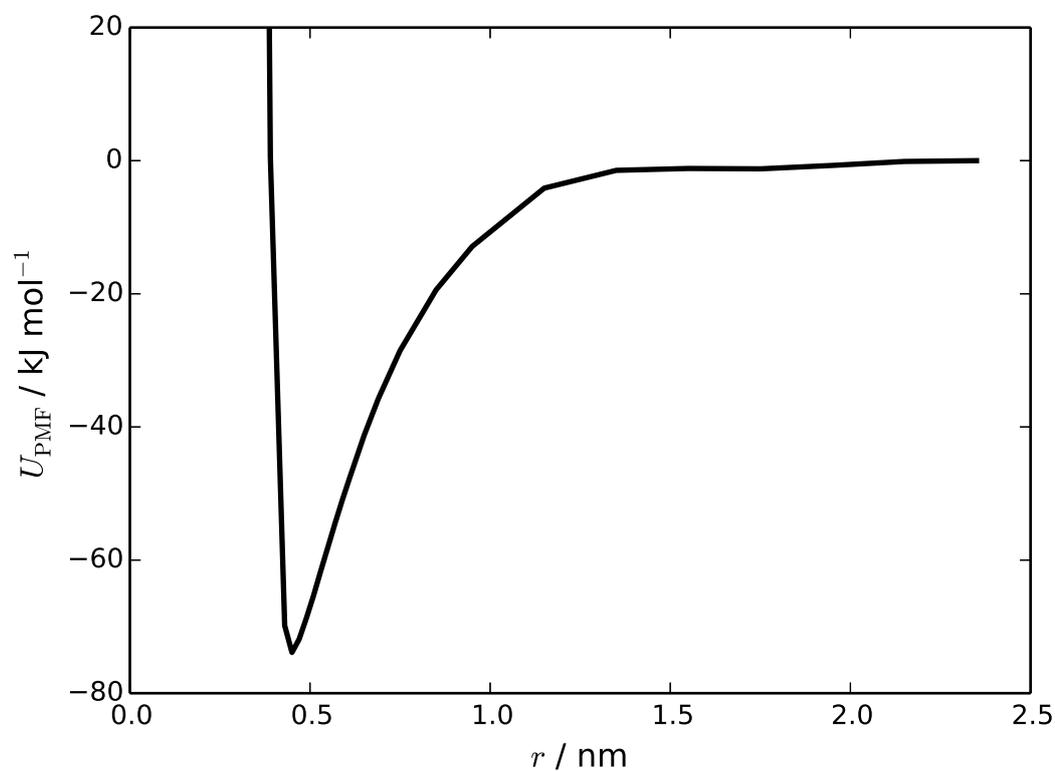

Figure 5.17: Potential of mean force curve calculated for the separation of a dimer of two molecules using the MARTINI 2 model.

The potential of mean force for the separation of a TP6EO2M dimer is given in Figure 5.17. The PMF has a similar shape to that reported for the atomistic system by Akinshina *et al.*,[86] with a short range repulsive and a long range attractive part. The free energy of association, ΔG, for the formation of a dimer can be obtained from the maximum well depth of the PMF. For this model, $\Delta G = -73.9 \pm 1.5$ kJ mol^{-1} at a concentration of 1.7×10^{-3} mol dm^{-3}, which is higher than the value of -40.7 ± 1.4 kJ mol^{-1} for a PMF obtained for the atomistic system. Therefore, the MARTINI model provides only semi-quantitative thermodynamic data. The original MARTINI paper recommends caution when using the force field for systems relying on the stacking of aromatic groups,[32] so the relatively poor accuracy achieved here is perhaps not surprising.

5.4.2   MARTINI 3

Simulation of the four different MARTINI 3 models for TP6EO2M, from a starting configuration of 10 TP6EO2M molecules dispersed in water, reveals very different behaviour depending on the parameters used for the AI and AO beads. Figure 5.18 shows snapshots taken after 100 ns of simulation for each of the models. The N0 and N0a models both give a 10 molecule chromonic stack in that time frame. The N1 model gives two shorter chromonic stacks, while the N1a model gives two short stacks and one monomer.

Simulations of a stack of 10 TP6EO2M molecules were run using each of the four molecules. Despite the differences in self-assembly, the pre-assembled stack was found to be stable over 100 ns for all of the models. Figure 5.19 shows distributions of d_{com} and d_{S} calculated for the nearest neighbours in the stack. The stacking distances are all significantly closer to the all-atom value (0.37 nm) than the MARTINI 2 value. As with the MARTINI 2 model, the d_{com} values are shifted to higher values compared to d_{S}, indicating slight offsets between adjacent molecules in the stack. The d_{com} distribution has a long tail similar to that found in the atomistic distribution (Figure 5.10), indicating that the MARTINI 3 stacks have a similar degree of flexibility as the atomistic model. Compared to the MARTINI 2 and HFM models, adjacent molecules in the MARTINI 3 model are less tightly bound due to the smaller C–C ϵ value, and so are more free to move relative to each other. The models do exhibit some variation in their stacking distance, despite using the same bead type for the
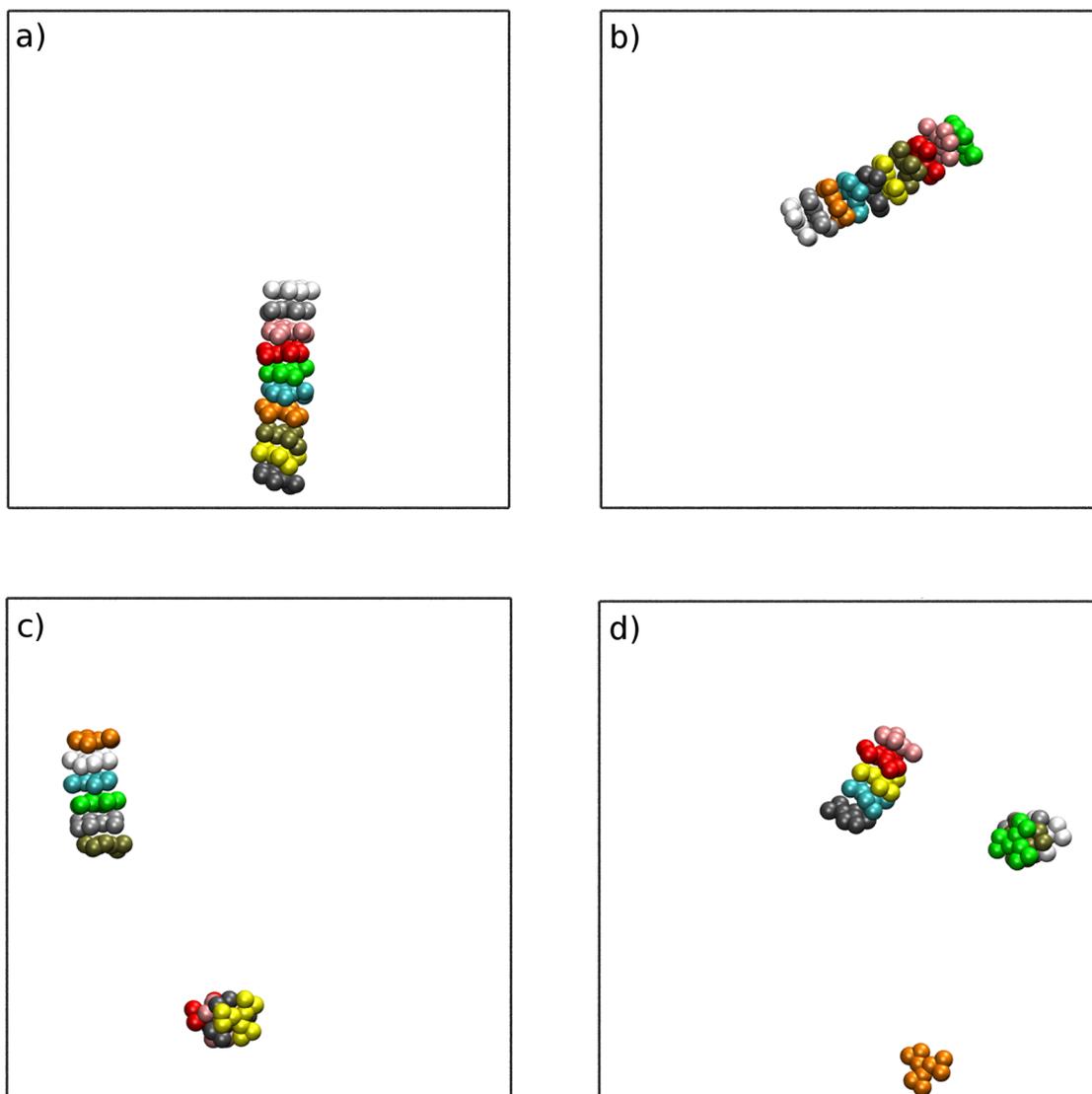
Figure 5.18: Final configurations from 100 ns simulations of 10 TP6EO2M molecules in water, using the a) N0, b) N0a, c) N1 and d) N1a models. For clarity, only the aromatic core beads are shown, and are coloured by molecule.

core. This may be rationalised in terms of solubility; the models with more soluble A beads have a larger d_S, allowing for greater solvation of the hydrophilic arms.

Neither MARTINI model is able to quantitatively match the site–site intermolecular RDFs from the atomistic model, as shown in Figure 5.21. However, MARTINI 3 appears to do better than MARTINI 2 in predicting the positions of the major peaks in all the RDFs shown. It appears that the introduction of more distinct particle sizes has helped the representation of local structure within the force field.

The PMFs calculated using the MARTINI 3 models are as shown in Figure 5.20. The ΔG values from these PMFs, for the N0, N1, N0a and N1a models respectively, are: -46.3 ± 1.5, -36.7 ± 1.5, -42.6 ± 1.5 and -32.4 ± 1.4 kJ mol^{-1}. These all represent a significant improvement over the MARTINI 2 model, in terms
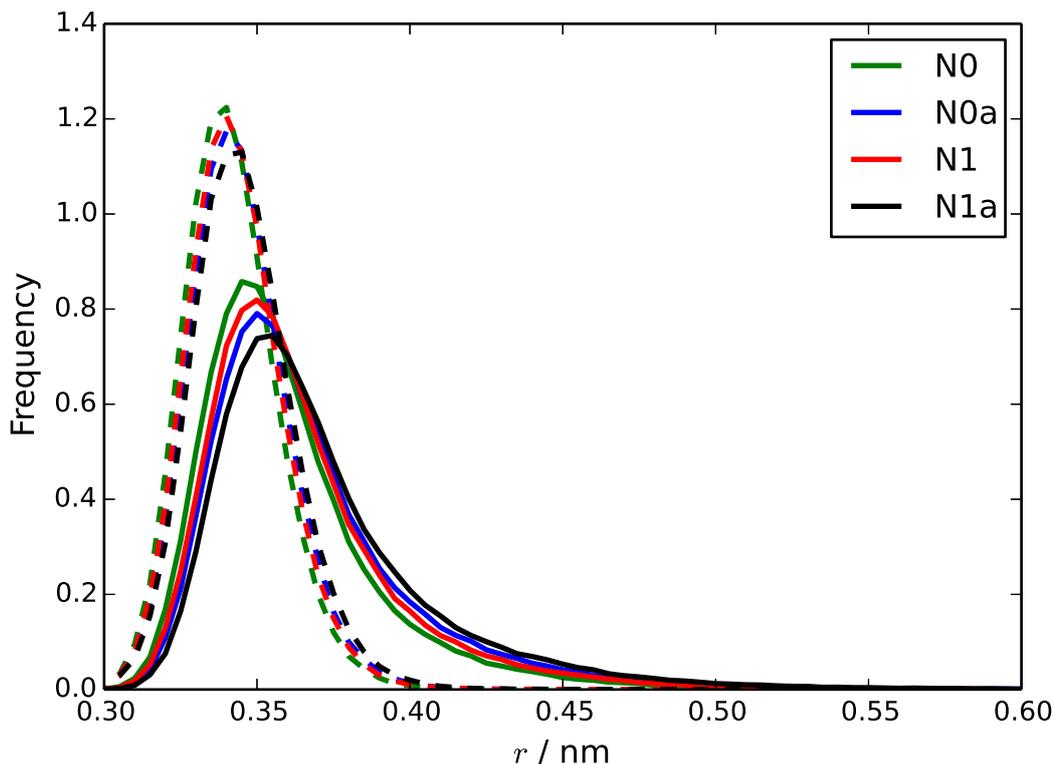
Figure 5.19: Distributions of d_{com} (solid lines) and d_{S} (dotted lines) for each of the four MARTINI 3 models.

of reproducing the free energy of assocation of a dimer. The differences between the PMFs are consistent with the behaviour of the models in simulations. The N0 and N0a models have deeper potential wells, and therefore more negative ΔGs. All of the PMFs have small bumps, which are larger for the N1 and N1a models; these can be viewed as small energy barriers to association. These two features can both be explained in the context of the interaction potentials of the models. As two TP6EO2M molecules approach each other, they will initially be fully solvated by water. At a certain distance, there will no longer be sufficient room between the molecules for all of the beads to be fully solvated, and this will carry a larger energy penalty for the more water soluble N1 and N1a models. Once the energy barrier has been overcome, the less soluble N0 and N0a will form a more stable dimer, since there is less of an energy penalty associated with their arms not being fully solvated.

The MARTINI 3 force field represents a total re-parametrisation of all of the non-bonded interactions in the MARTINI 2 model, and this has translated to noticeably different behaviour when the force fields are used in simulations of TP6EO2M. The MARTINI 3 open-beta documentation highlights improved stacking behaviour of aromatic molecules as one target for the parametrisation of the new forcefield. The

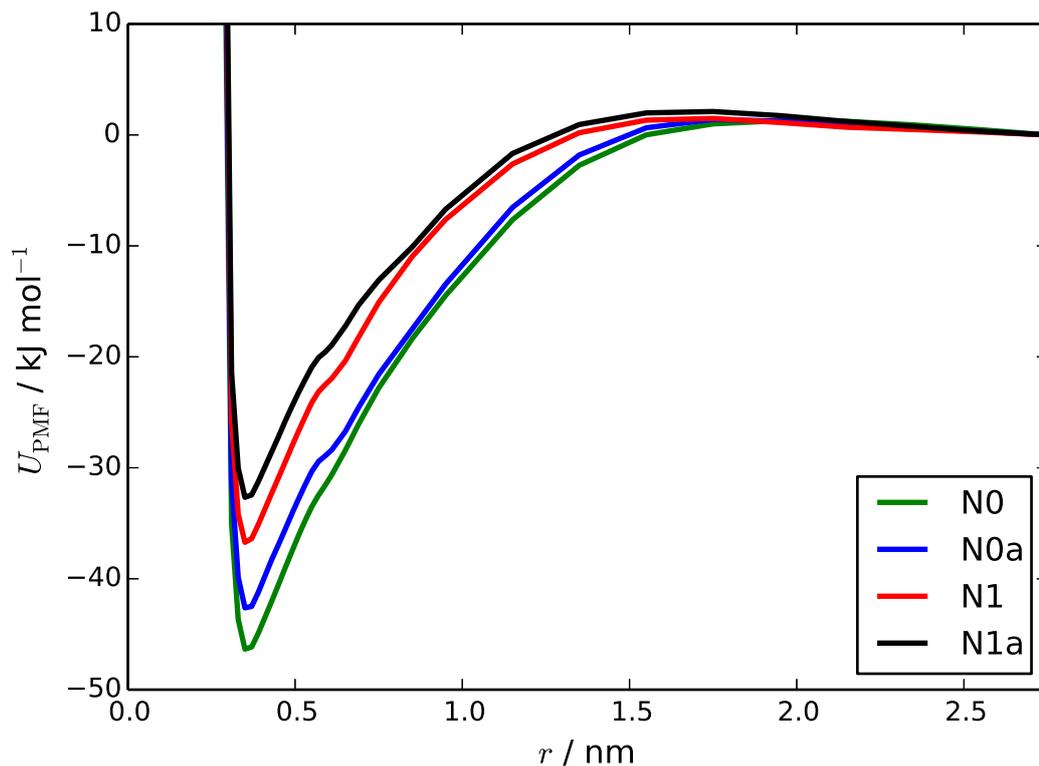**Figure 5.20:** PMFs for the separation of a TP6EO2M dimer, calculated using the four MARTINI 3 models.
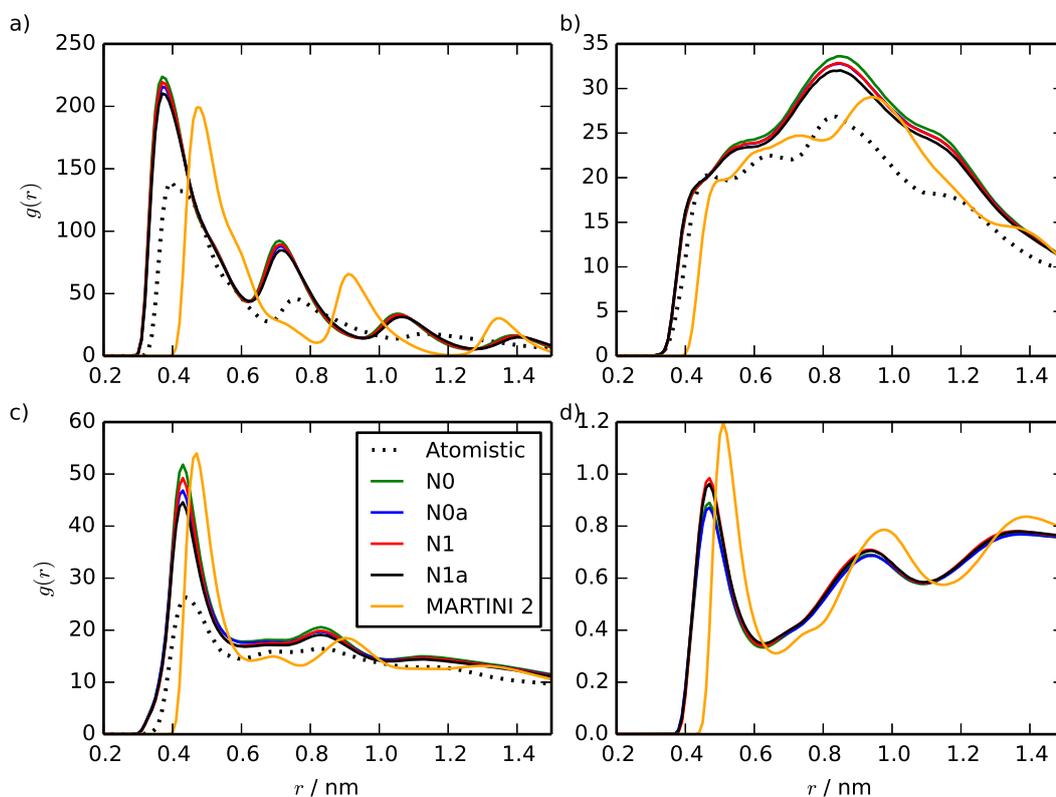

Figure 5.21: Site-site intermolecular RDFs calculated from a stack of 10 TP6EO2M molecules, using the MARTINI 2 and 3 models and compared to the atomistic results: a) C-C, b) C-A, c) A-A, d) A-W. No atomistic arm-water RDF is shown due to the difficulty in directly mapping 4 molecules to a single bead.
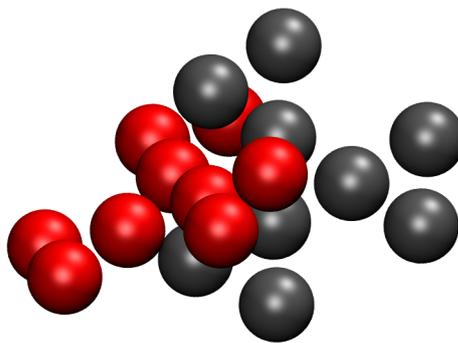
Figure 5.22: Simulation snapshot of a dimer using original SAFT-γ Mie model with unadjusted bond lengths, with the core–core distance constrained at 0.55 nm during a PMF calculation. Only the aromatic cores are shown.

more accurate stacking distances and free energies calculated by the MARTINI 3 forcefield, when compared to MARTINI 2, suggest that significant improvement has indeed been achieved. Although the exact strategy used for the re-parametrisation has not yet been published and is not available, it seems that the main strength of the new force field is in the accurate parameters for a range of bead-sizes. For a molecule like TP6EO2M, where smaller mappings are required, the improvement this brings is clear.

5.4.3 Insights from the SAFT-γ Mie approach

The MARTINI 2 model of TP6EO2M was initally compared to a coarse-grained forcefield developed using the SAFT-γ Mie approach. The parametrisation and much of the analysis was carried out by Jos Tasche. Full details can be found in the original publication, but the main points are summarised below.[40]

The parametrisation strategy for the TP6EO2M/water model was similar to that used in Chapter 3 for the octane/benzene mixture, where like-like interactions from separate models of the different components of the system were combined; in this case, pre-existing models of benzene[98], polyethylene oxide[209] and water[96] (in this case, a 2-to-1 mapping was used for water). The unlike interactions were calculated using the standard SAFT mixing rules (see Equations 3.6–3.9), with the k_{ij} parameter fitted to match experimental mixing data. Enthalpies of mixing were used for the C-A and A-W interactions, and liquid-liquid equilibrium compositions for the C-W interaction.

The first main finding from this model was that the use of tangential spheres,

$k_{Ar/EO}$	$k_{Ar/W}$	$k_{EO/W}$	Morphology
↑	↑	↑	conglomerate
↑	↑	=	chromonic stack
↑	=	↑	chromonic stack with additional aggregation
=	↑	↑	conglomerate of short stacks
↑	=	=	monomers only
=	↑	=	chromonic stack
=	=	↑	chromonic stack with additional aggregation

Figure 5.23: The influence of unfavourable cross-interactions on chromonic stacking. Results are shown for the final revised SAFT model with the same bond lengths as used in the MARTINI model. Up arrows indicate interactions that are less favourable than mixing rules with $k_{ij} = 0.2$. Equals signs indicate $k_{ij} = 0$. For the case where pure combining rules are used ($k_{ij} = 0$ for all cross-interactions), only monomers and transient self-assembled dimers are seen.

which leads to the same value being used for σ in the Mie potential and r_0 in the bond potential, causes significant issues for structural representability. For TP6EO2M, this led to a bond length which was too large, allowing unphysical dimer structures like the one seen in Figure 5.22 to be sampled. Simply altering the bond length to the one used in the MARTINI model was found to prevent the formation of such structures; this is not a universal solution to the issue, however, as it would be expected to cause issues in capturing bulk properties such as density.

Using the SAFT-γ model with fitted unlike interactions, chromonic stacking was observed. The investigation of a range of different SAFT models using different k_{ij} parameters revealed a wide range of behaviour as shown in Table 5.23, ranging from no aggregation, to chromonic stacks, to aggregation into other structures. This confirms that association into chromonic stacks is extremely sensitive to the balance of interactions in the model. In particular, it was shown that the hydrophobic/hydrophilic balance of the aromatic and ethylene oxide groups is crucial; this was also highlighted by the results of the different MARTINI 3 models.

5.5 Simulating liquid crystal phases with MARTINI 3

5.5.1 Self-assembly

It is always important to rigorously test coarse-grained models on a small system. However, the point behind coarse-grained modelling is to run larger simulations which are not easily accessible to atomistic models. Therefore, simulations of larger systems of aqueous TP6EO2M were carried out. Given that the MARTINI 3 models outperformed the MARTINI 2 models in terms of structural and thermodynamic representability, it was decided to use the MARTINI 3 models for these larger scale simulations.

To test the ability of these models to self-assemble chromonic phases, a large simulation box was set up containing 1000 TP6EO2M molecules with random positions and orientations, solvated by 36876 MARTINI water beads; this corresponds to a concentration of 26.1% by weight, which is in the nematic region of the TP6EO2M phase diagram at 280 K. An atomistic representation of this system would consist of 580512 atoms, the simulation of which would be not be feasible without very large computational resources; however, using 72 CPU cores, speeds of around 35 ns per day were achieved in simulations of the MARTINI 3 model.

These large scale simulations provide further insight into the representability of the four MARTINI 3 models. All of the models produced chromonic stacks during the simulations. However, the N0 and N0a models, which performed well in the small-scale simulations, phase separated, forming distinct liquid TP6EO2M and water regions. The N1 and N1a models, on the other hand, did not phase separate, forming long chromonic stacks which were fully solvated in water. Representative simulation snapshots from the N0 and N1 models are shown in Figure 5.24.

The difference in behaviour between the models can be attributed to the solubility of the beads used for the hydrophilic arms. The N0 and N0a beads are insufficiently soluble, so that when multiple long stacks are modelled, they preferentially aggregate together. The higher solubility of the N1 and N1a beads prevents this from happening. These results highlight the difficulties in parametrising models for hierarchical systems like chromonic liquid crystals; a model which is representative at one scale may not perform as well at another scale.
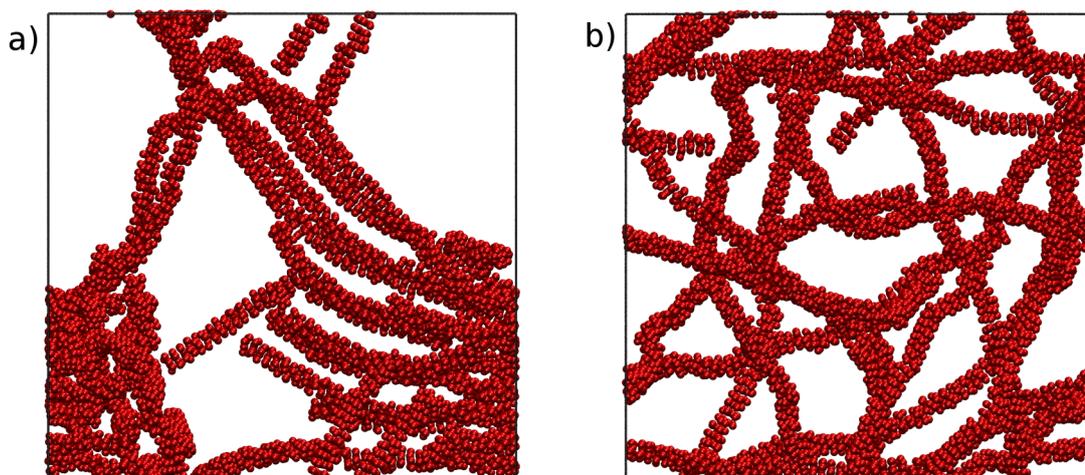
Figure 5.24: Final configurations after 500 ns of simulation from a starting point of 1000 TP6EO2M molecules in water, using the a) N0 and b) N1 MARTINI 3 models.
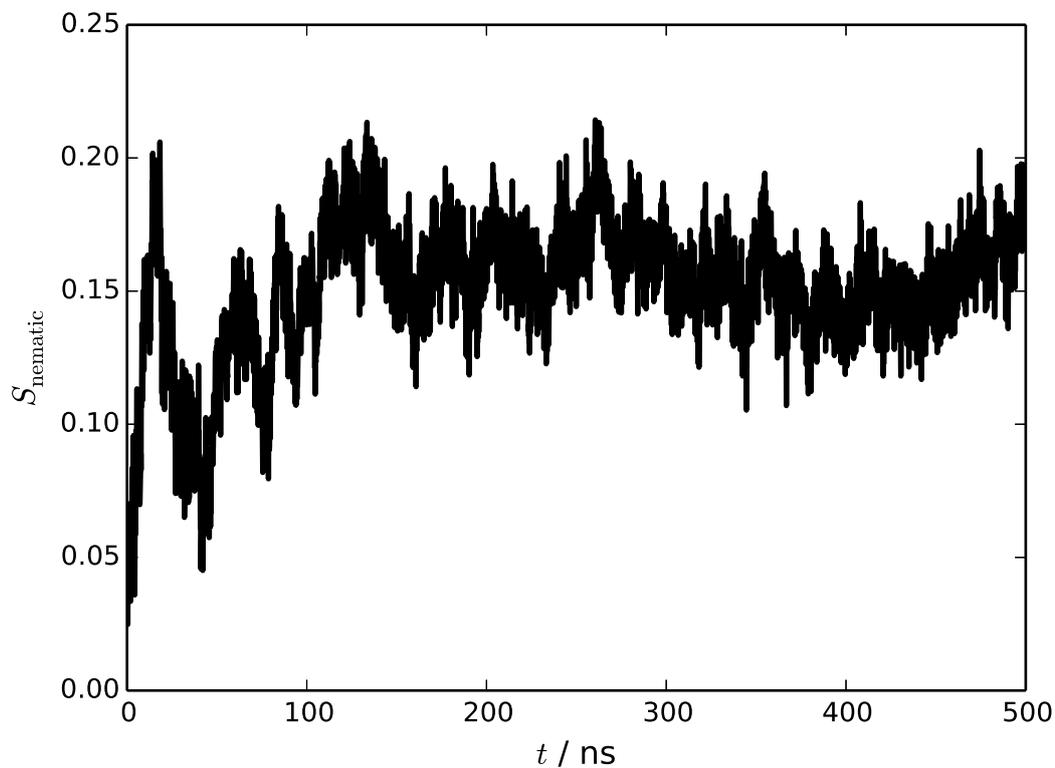

Figure 5.25: Evolution of S_{nematic} over 500 ns of simulation using the N1 model, from a randomly dispersed starting configuration of 1000 TP6EO2M molecules in water.

Even though the formation of chromonic stacks was observed in the first few nanoseconds of simulation of the N1 and N1a models, the self-assembly of these stacks into the nematic phase was not seen over 500 ns of simulation. The evolution of the nematic order paramer, S_{nematic}, over 500 ns of simulation using the N1 model is plotted in Figure 5.25. This mirrors the behaviour seen in DPD simulations of chromonic liquid crystals, where the introduction of a pseudo-magnetic field was required for the spontaneous formation of the nematic phase.[41] In this case, it is unlikely that simply running the simulation for longer would result in the expected phase behaviour. Many of the chromonic stacks formed are continuous over periodic boundary conditions, and are tangled up with other stacks. For the nematic phase to be formed, these stacks would first need to break and disentangle themselves. Given the speed with which chromonic stacks are formed, and the high free energy of association for molecular stacking, it is extremely unlikely that this will be observed within a reasonable simulation timescale using standard molecular dynamics, even for a coarse-grained model.

The use of alternative simulation techniques to simulate the formation of the nematic phase may be required. For example, an aligning field, while not supported in the Gromacs package, could be used to encourage the formation of the nematic phase as it was the DPD study.[41] Monte Carlo simulations may also help to overcome the energy barrier associated with breaking up and disentangling stacks which extend over periodic boundary conditions.

5.5.2   Seeding chromonic columns

Due to the difficulty of achieving spontaneous self-assembly of liquid crystal phases, the ability of the MARTINI 3 model to represent chromonic liquid crystal phases was examined by starting from a pre-assembled hexagonal columnar system, and then observing the evolution of the system. Starting structures were constructed by placing 30 columns with 30 TP6EO2M molecules each (for a total of 900 molecules) in a hexagonal arrangement in a simulation box. These columns were then solvated with varying amounts of MARTINI 3 water, giving the following range of TP6EO2M/water concentrations: 27.1, 39.4, 55.8 and 71.4 wt%. All starting structures had the same initial volume, so the volumes of the boxes were allowed to equilibrate before any data was gathered. A semi-isotropic pressure coupling was
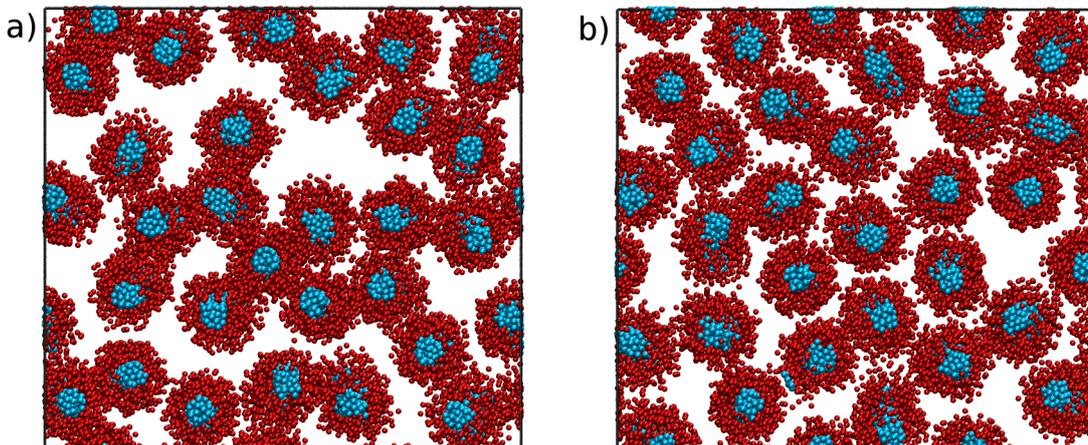
Figure 5.26: Configurations taken after 200 ns of simulations of the 55.8 wt% TP6EO2M/water system. a) The N1 model results in clustering of the columns, while b) the N1a shows the C_H phase. Showing only cores (cyan) and arms (red).

Table 5.6: TP6EO2M/water concentrations studied, and their densities, nematic order parameters and the chromonic phases formed.

Concentration / wt%	ρ / $kg\,m^{-3}$	$S_{nematic}$	LC phase
27.1	1093	0.914	N
39.4	1129	0.915	N
55.8	1181	0.922	C_H
71.4	1238	0.931	C_H

used, to allow the cross-section of the nematic phase and the lengths of the columns to change independently. The system densities equilibrated within 200 ps for all of the concentrations, and remained stable throughout the simulations; equilibrated densities are given in Table 5.6. Each system was simulated for a total of 500 ns.

These simulations, of course, have the opposite issue to that faced when attempting to self-assemble the nematic phase: the phase transition between nematic and isotropic is unlikely to be computationally accessible. This means that the nematic or hexagonal phase will be stable over the simulation time-scale, even if the isotropic phase would be energetically preferred. However, transitions between the hexagonal and nematic phase are feasible; this means it is possible to study the degree of hexagonal ordering present in the system, and how that changes with concentration. The ability of the model to represent the concentration dependence of the TP6EO2M phase diagram is an important test of the representability and transferability of the MARTINI 3 force field in modelling this system.

Running these simulations provided further insights into which bead type best represents the arm beads. The N1 and N1a models were both tested, and exhibited
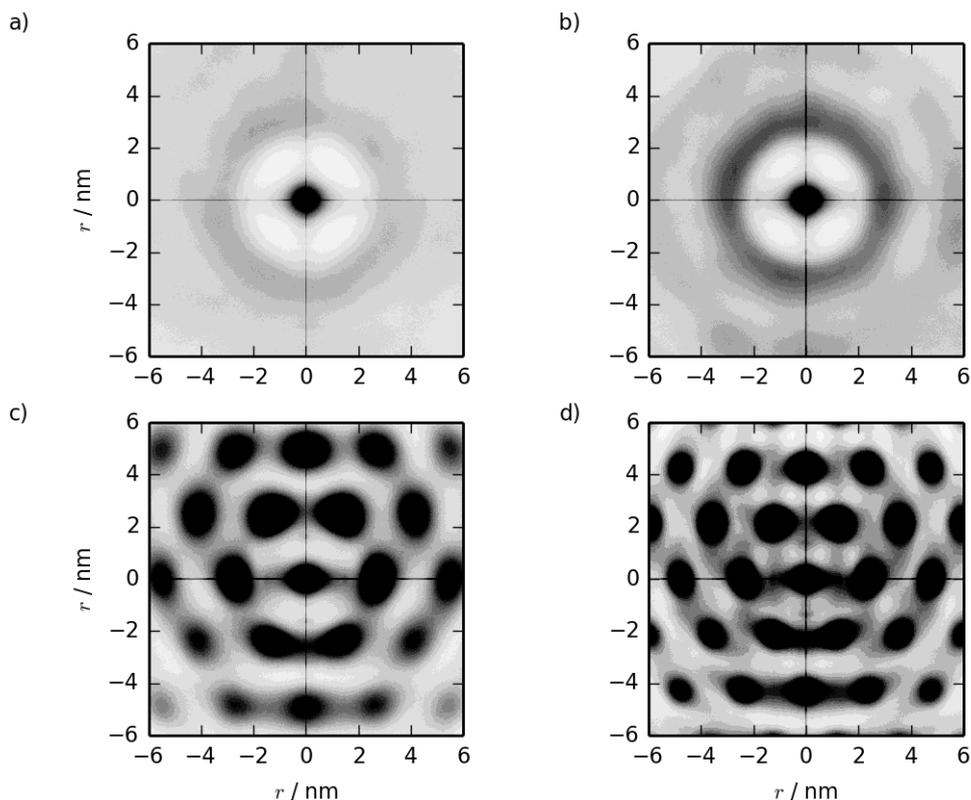
Figure 5.27: Two dimensional $g(u,v)$ from simulations of pre-assembled TP6EO2M columnar phases at concentrations of a) 27.1, b) 39.4, c) 55.8 and d) 71.4 wt%.

significantly different behaviour. Using the N1 model, there was a tendency for the stacks to form small clusters attached by the ethylene oxide arms, where the stacks were not fully solvated. The N1a model did not show this clustering. The different structures seen for the two models are shown in Figure 5.26 for the 55.8 wt% system. At this concentration, the N1a model shows the C_H phase as predicted by the phase diagram, while the formation of clusters by the N1 model disrupts the formation of hexagonal structures. The N1 model performed better in simulating the self-assembly of short stacks and has a free energy of association closer to the atomistic and experimental values, but when applied to a large system at higher concentrations it is far less representative.

The N1a model was therefore used for more detailed analysis of the system across the concentration range. The nematic order parameter, S_{nematic}, was calculated for each concentration to quantify the orientational order of each system, and the values are shown in Table 5.6. In each case, S_{nematic} is greater than 0.9, showing long-range orientational order. The value increases slightly with TP6EO2M concentration; this is expected, given that increased concentration leads to closer packing of the
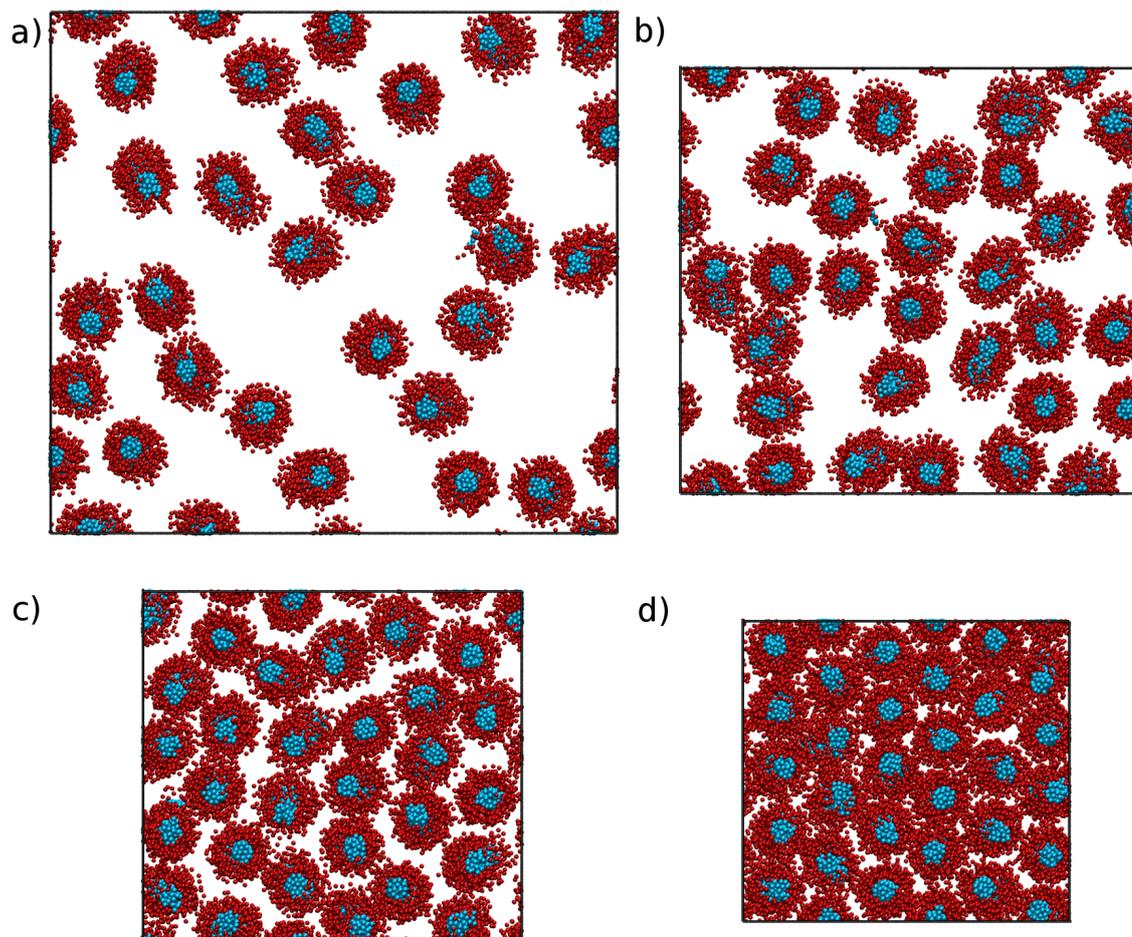
Figure 5.28: Configurations taken after 500 ns of simulations of the N1a model, starting from pre-assembled columnar structures at a) 27.1, b) 39.4, c) 55.8 and d) 71.4 wt%.

columns.

Two dimensional pair distribution functions were calculated for each system, and are shown in Figure 5.27, along with the configurations after 500 ns showing the cross-section of the nematic/hexagonal phase in Figure 5.28. The two lower concentrations show clear liquid behaviour in two dimensions; combined with the S_{nematic} values, this indicates the nematic phase. Between 39.4% and 55.8%, there is a clear transition from liquid-like to hexagonal ordering. This is in line with the experimental phase diagram in Figure 5.3, although experimentally the 39.4% system is close to the phase boundary but on the hexagonal columnar side. The phase boundary is slightly offset between experiment and the simulations, and in principle, the concentration where phase transition occurs could be narrowed down by carrying out simulations at some intermediate concentrations.

These simulations are effectively modelling a series of infinitely long stacks. However, in the real system, self-assembly is quasi-isodesmic (ie. the binding energy is

the same for the formation of all stack sizes apart from a dimer).[86], and the supply of unstacked monomers is large. This leads to an exponential distribution of stack sizes, where there is a large number of short stacks and a smaller number of long stacks.[41] Such a system would be difficult to model here due to the issues discussed in Section 5.5.1 with the rapid formation of infinite stacks. The size of system required to sample the size distribution well would also likely be higher, increasing the computational cost of the simulations. A system with a distribution of sizes might be expected to behave slightly differently. In particular, the presence of short stacks in the system may disrupt hexagonal packing, resulting in a slightly more disordered C_H phase. Despite this, the results here show that the MARTINI 3 force field is able to capture the concentration dependence of the TP6EO2M/water phase diagram reasonably well.

Finally, further refinement of the force field could be beneficial. A model which is able to model the formation of a chromonic stack, and which is also able to model liquid crystal phases accurately, would be a definite improvement over the representability and transferability of the MARTINI 3 models studied here, all of which are more representative at one scale than the other. This could be done by testing combinations of existing bead types within the MARTINI 3 framework, or by reparametrising the interactions so that they represent the system better.

5.6   Conclusions

The parametrisation of coarse-grained models of TP6EO2M brings with it a number of challanges not present in simpler systems. The behaviour of the system relies on a very delicate balance between the hydrophobic and hydrophilic parts of the molecule, which must be captured by a coarse-grained model. Moreover, within a coarse-grained representation, the coarse-graining procedure changes the balance between enthalpic and entropic interactions. Coarse-grained water has a lower gain in entropy, in comparison with real water, when freed from interaction with a chromonic molecule. Consequently, the free energy balance within a coarse-grained model must be represented by more favourable enthalpic interactions favouring association.

HFM was used to parametrise a number of coarse-grained models of TP6EO2M, highlighting a number of issues with the representability of *bottom-up* coarse grain-

ing of more complex systems. A model which captures chromonic assembly was successfully parametrised; however, thermodynamic properties calculated using this model were not representative of the real system; this perhaps relates to the difficulty of sampling rarer configurations in an atomistic reference system. The mapping of multiple solvent molecules to a bead within the HFM framework was also addressed. However, the models produced did not result in stable chromonic stacks, because the loss of entropy from the coarser mapping was not balanced by the enthalpic contribution to the free energy.

The MARTINI force field was also successful in capturing chromonic assembly. The non-bonded interactions within MARTINI are designed to model the partitioning (and hence the transfer free energy) between common solvents. MARTINI is therefore able to do a reasonably good job of capturing both the overall hydrophilic-lipophilic balance for a chromonic molecule in water, and the relative hydrophilic-lipophilic balance between the different parts of the molecule.

The MARTINI 3 simulations have further highlighted the subtle balance of interactions required when modelling liquid crystals. Even a small change in the solubility of the arm beads in the TP6EO2M model results in qualitatively different behaviour in simulations. Crucially, these differences may not manifest themselves in, and cannot be predicted from, the small-scale, low-concentration, simulations which were carried out validate the models, but are very clear when larger systems are studied. The key point here is that the results of coarse-grained simulations must be examined critically at all stages of a study.

Coarse-graining using SAFT-γ also provides insights. Across a small parameter range for the unlike interactions, one can achieve a range of aggregation behaviour, with chromonic self-assembly only occurring at the right balance of hydrophilic and lipophilic interactions for the aromatic and ethylene oxide components. If the balance of interactions is wrong, the self-assembly process may not be observed within a reasonable simulation time, either due to a lack of self-assembly, or the formation of other conglomerates.

Coarse-graining allows for a huge speed-up in observing chromonic assembly, with savings arising from a faster exploration of phase space in addition to a large reduction in the number of sites simulated. TP6EO2M reduces from 138 sites to 27 (a factor of $\sim 5\times$ and, in the *top-down* models, four water molecules (12 sites) reduce to 1 site. This speed-up has allowed the simulation of larger systems than those

accessible to atomistic systems to investigate the phase behaviour of TP6OE2M at a range of concentrations. The MARTINI 3 force field has been used to model the nematic and hexagonal phases, and the concentrations at which these phases occur are consistent with the experimental phase diagram. The spontaneous assembly of these phases was not observed, however, Despite the computational speed-up, there are still processes which are not easily accessible to systematic coarse-grained models.

Overall, the *top-down* and *bottom-up* models described in this chapter have been successful in modelling the self-assembly of chromonic stacks. The first simulations of the chromonic N and C_H phases using a systematic coarse-grained model have also been carried out with the MARTINI 3 force field, showing good agreement with experimental results. However, there is still room for improvement in both approaches. The ultimate goal of this work is to be able to predict the phase behaviour of a liquid crystal molecule using systematic coarse-graining. This would require addressing the issues described in this chapter, and others, and would be very valuable to the fields of coarse-grained modelling and liquid crystals.

Chapter 6

Conclusions

6.1 Summary

This thesis has presented an investigation of different methods for parametrising coarse-grained models for two contrasting chemical systems. Insights have been obtained about how well these methods are able to produce models which are both representative and transferable. Each method has its own strengths and weaknesses, but work is still needed to achieve coarse-grained models which exhibit all of the desired characteristics.

Chapter 3 presented a comparison between three different coarse graining methods, hybrid force matching (HFM), iterative Boltzmann inversion (IBI) and SAFT-γ Mie, applied to liquid octane/benzene mixtures at a range of concentrations. The models were tested by calculating radial distribution functions (RDFs), densities at a range of temperatures and concentrations and solvation free energies; these were all compared to the same quantities calculated for the atomistic system. The two *bottom-up* methods, HFM and IBI, were both able to represent the structure of the underlying atomistic system very well, and after the application of a linear pressure correction were also shown to be thermodynamically representative. However, neither technique produced models which were particularly transferable. The coarse-grained interaction potentials were highly state-point dependent, differing significantly depending on the concentration at which they were parametrised. This state-point dependence resulted in poor transferability across both concentrations and temperatures, particularly in calculating densities. The poor density/temperature relationship highlighted the main issue with the linear pressure

correction: it corrects the pressure at one state point, but does not address the state-point dependence which causes the poor representation of pressure in coarse-grained models. The *top-down* SAFT-γ Mie model, by contrast, had good chemical and temperature transferability, as well as good thermodynamic representability. However, the structural accuracy was poor, with the model unable to reproduce atomistic RDFs accurately; this arises from the tangential sphere constraint in the version of SAFT used in this study.

Chapter 4 built on the results obtained in Chapter 3, and investigated how the generally poor transferability of *bottom-up* coarse-grained models may be improved. The MS-IBI methodology was very successful in parametrising a coarse-grained model of octane/benzene with good transferability across concentrations without much loss of representability compared to standard IBI. However, even using MS-IBI with references at multiple temperatures, a temperature transferable coarse-grained model of octane was not achieved. Coarse-grained pair potentials are known to be state-point dependent, and this may prevent temperature transferability in coarse-grained models based on pair potentials without the loss of representability. Attempts were therefore made to address this by including local-density dependent interactions in the model. Firstly, a simpler *bottom-up* model of the octane/benzene mixture was parametrised based on Morse potentials, using the simplex minimisation algorithm; this model was structurally representative and chemically transferable, as well as being very efficient to parametrise. Using these Morse potentials, initial work to parametrise local density potentials to match pressure was not successful. However the combination of the simplex approach for pair potentials and local density dependent potentials to address the state-point dependence of the pair potentials is one which is worth pursuing in future work.

Chapter 5 went on to investigate a different system: the chromonic liquid crystal TP6EO2M in water. The feasibility of modelling this system using *bottom-up* and *top-down* coarse-grained models was tested by applying HFM and different versions of the MARTINI forcefield. Each coarse-grained model was tested by examining its ability to simulate chromonic self-assembly, the structure of a short chromonic stack (by calculating RDFs and stacking distances) and the potential of mean force (PMF) for the formation of a dimer. The first finding for the HFM method was that separating electrostatic and VdW interactions in the coarse-grained model significantly improved the self-assembly behaviour of the model. The HFM model with separate

electrostatics had fairly good structural representability, capturing the main features of the atomistic RDFs and showing accurate stacking distances. However, the PMF showed a free energy of association around five times too large, which highlights the difficulties in *bottom-up* models capturing the properties of structures not present in the reference system. Attempts to use a more computationally efficient mapping for the HFM model did not result in models which allowed stable chromonic stacks. The MARTINI models all exhibited chromonic self-assembly. The models using the MARTINI 3 force field were more structurally and thermodynamically representative than the MARTINI 2 force field which suggests that significant efforts have been made in the reparametrisation of the force field. Finally, the MARTINI 3 force field was applied to a large system to test how well the model captures liquid crystal phases. The self-assembly of the nematic phase was not observed over the simulation time scales, partially due to the rapid formation of infinite stacks which extended over periodic boundary conditions. Simulations from a pre-assembled hexagonal lattice, however, showed a concentration dependent transition between the nematic and hexagonal phases, which is in line with the experimental phase diagram.

6.2   Outlook

Overall, the work in this thesis has highlighted a number of open questions in the field of coarse-grained modelling. It is clear that, while coarse-grained models are a powerful tool for investigating systems at scales inaccessible to higher resolution models, there are a number of difficulties which can make their application problematic.

The biggest issue for *bottom-up* coarse grained models remains their transferability between state points. This can partially be solved by ensuring that the references system used represent all of the chemical and configurational space of interest, perhaps requiring multiple reference trajectories. MS-IBI was shown here, and in other studies,[117,118] to be useful for this purpose. Multiple reference systems could also be utilised in the force matching method which could, for example, help improve the HFM model of TP6EO2M by including both stacked and unstacked reference states. This would in principle allow the orientational dependence of the interactions between molecules to be captured. Both references would need to be sufficiently well sampled, but the relative weighting of each state-point would be

particularly important as it would determine the relative stability of a chromonic stack.

The larger issue of transferability caused by the state-point dependence of coarse-grained pair potentials may be solved by the introduction of local density dependent potentials. Such a potential could act as a pressure correction with more of a basis in underlying thermodynamics than the linear pressure correction commonly used in *bottom-up* coarse graining. Future research efforts should go towards how these potentials can be efficiently parametrised. This may involve the use of an analytical form which could be parametrised using the simplex approach used in Chapter 4 for parametrising Morse pair potentials.

Top-down coarse-grained models generally have good chemical and thermodynamic transferability compared to *bottom-up* models because they were parametrised using data from a range of thermodynamic state points. However, this appears to come from a trade-off between transferability and structural representability, which is often poor for *top-down* models. The SAFT-γ Mie approach currently suffers from the use of tangential spheres, although the use of shape factors to solve this appears promising.[162] The MARTINI 3 model seems to have improved on the structural representability of MARTINI 2. When the force field is formally published it should become clear whether this has come at the cost of transferability of thermodynamic accuracy compared to the earlier model.

The issue of whether a coarse-grained model can adequately represent the overall free energy of a system is fundamental to its representability, since the entropy/enthalpy balance is altered by the coarse graining process. The formation of complex structures in chemical systems often arises from both enthalpic and entropic contributions, and so addressing this issue is required for a coarse-grained model to be successful. *Top-down* models which are parametrised to match experimental free energies, such as the MARTINI force field,[32] are therefore usually able to model complex structures with at least qualitative accuracy. This is not always the case with *bottom-up* coarse graining, particularly when coarser mappings are used.

Finally, it is worth discussing how much of an advantage is gained by coarse-grained modelling in terms of speed-up. Running a coarse-grained simulation is undeniably faster than running an atomistic simulation of the same length and system size, both due to the reduction in the number of particles and the more

efficient equilibration and sampling of phase space. This advantage is of course dependent on the level of coarse graining, and the complexity of the interactions used in the model. However, the time taken to actually parametrise a coarse-grained model has not been discussed. This issue particularly applies to *bottom-up* coarse graining, where the parametrisation is often very lengthy. Of course, if a system cannot be modelled atomistically then the time spent parametrising a good coarse-grained model is still worth it. However, *bottom-up* coarse graining is still at a massive disadvantage compared to, for example, the MARTINI model, where the coarse graining process is often much faster. The time spent parametrising a *bottom-up* model for one system could alternatively be spent modelling multiple systems using a MARTINI approach. Therefore, the speed of parametrisation, the speed of the models and their relative accuracies must all be taken into account when choosing a model to use.

Coarse-grained modelling has already been used to provide insights into the behaviour of systems which cannot be easily modelled using other methods. Solving the issues presented in this thesis will therefore be extremely valuable, allowing for more trustworthy results from coarse-grained simulations. Ultimately, coarse-grained models fill an important niche between atomistic and mesoscale models, and so will become a more valuable tool all across soft matter science.

Bibliography

[1] F. Jensen, *Introduction to Computational Chemistry*, John Wiley & Sons, 2007.

[2] K. Burke, *J. Chem. Phys.*, 2012, **136**, 150901.

[3] C. D. Sherrill, *J. Chem. Phys.*, 2010, **132**, 110902.

[4] R. Car and M. Parrinello, *Phys. Rev. Lett.*, 1985, **55**, 2471.

[5] M. Karplus and J. A. McCammon, *Nat. Struct. Biol.*, 2002, **9**, 646–652.

[6] J. Durrant and J. A. McCammon, *BMC Biol.*, 2011, **9**, 71.

[7] P. Español and P. B. Warren, *J. Chem. Phys.*, 2017, **146**, 150901.

[8] S. Chen and G. D. Doolen, *Annu. Rev. Fluid Mech.*, 1998, **30**, 329–364.

[9] E. Brini, E. A. Algaer, P. Ganguly, C. Li, F. Rodríguez-Ropero and N. F. A. van der Vegt, *Soft Matter*, 2013, **9**, 2108–2119.

[10] H. M. Senn and W. Thiel, *Angew. Chem., Int. Ed.*, 2009, **48**, 1198–1229.

[11] M. Praprotnik, L. Delle Site and K. Kremer, *J. Chem. Phys.*, 2005, **123**, 224106.

[12] A. V. Sinitskiy and G. A. Voth, *J. Chem. Phys.*, 2018, **148**, 014102.

[13] T. A. Wassenaar, K. Pluhackova, R. A. Böckmann, S. J. Marrink and D. P. Tieleman, *J. Chem. Theory Comput.*, 2014, **10**, 676–690.

[14] J. F. Rudzinski and W. G. Noid, *J. Phys. Chem. B*, 2014, **118**, 8295–8312.

[15] M. Chakraborty, C. Xu and A. D. White, *J. Chem. Phys.*, 2018, **149**, 134106.

[16] M. G. Martin and J. I. Siepmann, *J. Phys. Chem. B*, 1998, **102**, 2569–2577.

[17] L. Yang, C. Tan, M. Hsieh, J. Wang, Y. Duan, P. Cieplak, J. Caldwell, P. A. Kollman and R. Luo, *J. Phys. Chem. B*, 2006, **110**, 13166–13176.

[18] W. L. Jorgensen, J. D. Madura and C. J. Swenson, *J. Am. Chem. Soc.*, 1984, **106**, 6638–6646.

[19] J. F. Dama, A. V. Sinitskiy, M. McCullagh, J. Weare, B. Roux, A. R. Dinner and G. A. Voth, *J. Chem. Theory Comput.*, 2013, **9**, 2466–2480.

[20] S. Izvekov and G. A. Voth, *J. Chem. Phys.*, 2005, **123**, 134105.

[21] D. Reith, M. Pütz and F. Müller-Plathe, *J. Comput. Chem.*, 2003, **24**, 1624–1636.

[22] J. McCarty, A. J. Clark, I. Y. Lyubimov and M. G. Guenza, *Macromolecules*, 2012, **45**, 8482–8493.

[23] Z. E. Hughes, M. R. Wilson and L. M. Stimson, *Soft Matter*, 2005, **1**, 436–443.

[24] G. Saielli, G. A. Voth and Y. Wang, *Soft Matter*, 2013, **9**, 5716–5725.

[25] J. Zhang, J. Su, Y. Ma and H. Guo, *J. Phys. Chem. B*, 2012, **116**, 2075–2089.

[26] S. J. Marrink and D. P. Tieleman, *Chem. Soc. Rev.*, 2013, **42**, 6801–6822.

[27] M. P. Allen, *Molecular Physics*, 2019, DOI:10.1080/00268976.2019.1612957.

[28] M. Walker and M. R. Wilson, *Soft Matter*, 2016, **12**, 8588–8594.

[29] V. Rühle, C. Junghans, A. Lukyanov, K. Kremer and D. Andrienko, *J. Chem. Theory Comput.*, 2009, **5**, 3211–3223.

[30] V. Rühle and C. Junghans, *Macromol. Theory Simul.*, 2011, **20**, 472–477.

[31] N. J. H. Dunn, K. M. Lebold, M. R. DeLyser, J. F. Rudzinski and W. Noid, *J. Phys. Chem. B*, 2018, **122**, 3363–3377.

[32] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman and A. H. de Vries, *J. Phys. Chem. B*, 2007, **111**, 7812–7824.

[33] C. Avendaño, T. Lafitte, A. Galindo, C. S. Adjiman, G. Jackson and E. A. Müller, *J. Phys. Chem. B*, 2011, **115**, 11154–11169.

[34] K. A. Maerzke and J. I. Siepmann, *J. Phys. Chem. B*, 2011, **115**, 3452–3465.

[35] H. I. Ingólfsson, C. A. Lopez, J. J. Uusitalo, D. H. de Jong, S. M. Gopal, X. Periole and S. J. Marrink, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2013, **4**, 225–248.

[36] M. S. Shell, *J. Chem. Phys.*, 2008, **129**, 144108.

[37] A. A. Louis, *J. Phys.: Condens. Matter*, 2002, **14**, 9187.

[38] M. E. Johnson, T. Head-Gordon and A. A. Louis, *J. Chem. Phys.*, 2007, **126**, 144509.

[39] J. W. Wagner, J. F. Dama, A. E. P. Durumeric and G. A. Voth, *J. Chem. Phys.*, 2016, **145**, 044108.

[40] T. D. Potter, J. Tasche, E. L. Barrett, M. Walker and M. R. Wilson, *Liq. Cryst.*, 2017, **0**, 1–11.

[41] M. Walker, A. J. Masters and M. R. Wilson, *Phys. Chem. Chem. Phys.*, 2014, **16**, 23074–23081.

[42] E. C. Allen and G. C. Rutledge, *J. Chem. Phys.*, 2009, **130**, 034904.

[43] N. J. H. Dunn, T. T. Foley and W. G. Noid, *Acc. Chem. Res.*, 2016, **49**, 2832–2840.

[44] A. Soper, *Chem. Phys.*, 1996, **202**, 295–306.

[45] P. Carbone, H. A. K. Varzaneh, X. Chen and F. Müller-Plathe, *J. Chem. Phys.*, 2008, **128**, 064904.

[46] H. J. Qian, P. Carbone, X. Chen, H. A. Karimi-Varzaneh, C. C. Liew and F. Müller-Plathe, *Macromolecules*, 2008, **41**, 9919–9929.

[47] K. Prasitnok and M. R. Wilson, *Phys. Chem. Chem. Phys.*, 2013, **15**, 17093–17104.

[48] C. Peter and K. Kremer, *Soft Matter*, 2009, **5**, 4357.

[49] C. Hijón, P. Españ̃ol, E. Vanden-Eijnden and R. Delgado-Buscalioni, *Faraday Discuss.*, 2009, **144**, 301–322.

[50] A. Davtyan, J. F. Dama, G. A. Voth and H. C. Andersen, *J. Chem. Phys.*, 2015, **142**, 154104.

[51] L. Saiz and M. L. Klein, *Account. Chem. Res.*, 2002, **35**, 482 – 489.

[52] S. J. Marrink, A. H. de Vries and D. P. Tieleman, *Biochim. Biophys. Acta, Biomembr.*, 2009, **1788**, 149–168.

[53] A. Catte, M. R. Wilson, M. Walker and V. S. Oganesyan, *Soft Matter*, 2018, **14**, 2796–2807.

[54] P. Setny and M. Zacharias, *Nucleic Acids Res.*, 2011, **39**, 9118–9129.

[55] R. Pool and P. G. Bolhuis, *J. Phys. Chem. B*, 2005, **109**, 6650–6657.

[56] E. N. Brodskaya, *Colloid J.*, 2012, **74**, 154–171.

[57] W. Shinoda, R. DeVane and M. L. Klein, *Mol. Sim.*, 2007, **33**, 27–36.

[58] X. He, W. Shinoda, R. DeVane, K. L. Anderson and M. L. Klein, *Chem. Phys. Lett.*, 2010, **487**, 71–76.

[59] R. D. Groot and T. J. Madden, *J. Chem. Phys.*, 1998, **108**, 8713 – 8724.

[60] A. A. Gavrilov, Y. V. Kudryavtsev and A. V. Chertovich, *J. Chem. Phys.*, 2013, **139**, 224901.

[61] J. S. Lintuvuori and M. R. Wilson, *Phys. Chem. Chem. Phys.*, 2009, **11**, 2116–2125.

[62] C. Zannoni, *J. Mater. Chem.*, 2001, **11**, 2637–2646.

[63] M. Walker and M. R. Wilson, *Soft Matter*, 2016, **12**, 8876–8883.

[64] R. W. Hockney, S. P. Goel and J. W. Eastwood, *J. Comput. Phys*, 1974, **14**, 148–158.

[65] W. C. Swope, H. C. Andersen, P. H. Berens and K. R. Wilson, *J. Chem. Phys.*, 1982, **76**, 637–649.

[66] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins: Struct., Funct., Bioinf.*, 2006, **65**, 712–725.

[67] C. J. Dickson, L. Rosso, R. M. Betz, R. C. Walker and I. R. Gould, *Soft Matter*, 2012, **8**, 9617–9627.

[68] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.

[69] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.

[70] W. L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1988, **110**, 1657–1666.

[71] B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463–1472.

[72] T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.

[73] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.

[74] G. Lamoureux, A. D. MacKerell and B. Roux, *J. Chem. Phys.*, 2003, **119**, 5185–5197.

[75] J. W. Ponder, C. Wu, P. 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, *J. Phys. Chem. B*, 2010, **114**, 2549–2564.

[76] L. R. Pratt and S. W. Haan, *J. Chem. Phys.*, 1981, **74**, 1864–1872.

[77] E. S. Severin and D. J. Tildesley, *Mol. Phys.*, 1980, **41**, 1401–1418.

[78] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.

[79] S. Nosé, *J. Chem. Phys.*, 1984, **81**, 511–519.

[80] W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695–1697.

[81] M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182–7190.

[82] B. Smith, T. Hauschild and J. M. Prausnitz, *Molecular Physics*, 1992, **77**, 1021–1031.

[83] G. S. Rushbrooke, *Transactions of the Faraday Society*, 1940, **36**, 1055–1062.

[84] C. H. Bennett, *J. Comput. Phys*, 1976, **22**, 245–268.

[85] J. G. Kirkwood, *J. Chem. Phys.*, 1935, **3**, 300–313.

[86] A. Akinshina, M. Walker, M. R. Wilson, G. J. T. Tiddy, A. J. Masters and P. Carbone, *Soft Matter*, 2015, **11**, 680–691.

[87] S. J. Marrink, A. H. de Vries and A. E. Mark, *J. Phys. Chem. B*, 2004, **108**, 750–760.

[88] L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman and S. J. Marrink, *J. Chem. Theory Comput.*, 2008, **4**, 819–834.

[89] D. H. de Jong, G. Singh, W. F. D. Bennett, C. Arnarez, T. A. Wassenaar, L. V. Schäfer, X. Periole, D. P. Tieleman and S. J. Marrink, *J. Chem. Theory Comput.*, 2013, **9**, 687–697.

[90] C. A. López, A. J. Rzepiela, A. H. de Vries, L. Dijkhuizen, P. H. Hünenberger and S. J. Marrink, *J. Chem. Theory Comput.*, 2009, **5**, 3195–3210.

[91] H. Lee, A. H. de Vries, S. J. Marrink and R. W. Pastor, *J. Phys. Chem. B*, 2009, **113**, 13186–13194.

[92] G. Rossi, L. Monticelli, S. R. Puisto, I. Vattulainen and T. Ala-Nissila, *Soft Matter*, 2011, **7**, 698–708.

[93] S. O. Yesylevskyy, L. V. Schäfer, D. Sengupta and S. J. Marrink, *PLoS Comput. Biol.*, 2010, **6**, e1000810.

[94] W. G. Chapman, K. E. Gubbins, G. Jackson and M. Radosz, *Ind. Eng. Chem. Res.*, 1990, **29**, 1709–1721.

[95] T. Lafitte, D. Bessieres, M. M. Piñeiro and J. Daridon, *J. Chem. Phys.*, 2006, **124**, 024509.

[96] O. Lobanova, C. Avendaño, T. Lafitte, E. A. Müller and G. Jackson, *Mol. Phys.*, 2015, **113**, 1228–1249.

[97] C. Avendaño, T. Lafitte, C. S. Adjiman, A. Galindo, E. A. Müller and G. Jackson, *J. Phys. Chem. B*, 2013, **117**, 2717–2733.

[98] T. Lafitte, C. Avendaño, V. Papaioannou, A. Galindo, C. S. Adjiman, G. Jackson and E. A. Müller, *Mol. Phys.*, 2012, **110**, 1189–1203.

[99] O. Lobanova, A. Mejía, G. Jackson and E. A. Müller, *J. Chem. Thermodyn.*, 2016, **93**, 320–336.

[100] W. Tschöp, K. Kremer, J. Batoulis, T. Bürger and O. Hahn, *Acta Polym.*, 1998, **49**, 61–74.

[101] R. L. Henderson, *Phys. Lett. A*, 1974, **49**, 197–198.

[102] VOTCA development team, *VOTCA 1.3.1 User Manual*, 2016.

[103] R. L. C. Akkermans and W. J. Briels, *J. Chem. Phys.*, 2001, **114**, 1020–1031.

[104] G. Milano and F. Müller-Plathe, *J. Phys. Chem. B*, 2005, **109**, 18609–18619.

[105] Q. Sun and R. Faller, *J. Chem. Theory Comput.*, 2006, **2**, 607–615.

[106] E. Brini, V. Marcon and N. F. A. van der Vegt, *Phys. Chem. Chem. Phys.*, 2011, **13**, 10468–10474.

[107] H. Wang, C. Junghans and K. Kremer, *Eur. Phys. J. E*, 2009, **28**, 221–229.

[108] P. Ganguly, D. Mukherji, C. Junghans and N. F. A. van der Vegt, *J. Chem. Theory Comput.*, 2012, **8**, 1802–1807.

[109] P. Ganguly and N. F. A. van der Vegt, *J. Chem. Theory Comput.*, 2013, **9**, 5247–5256.

[110] A. Lyubartsev, A. Mirzoev, L. Chen and A. Laaksonen, *Faraday Discuss.*, 2009, **144**, 43–56.

[111] A. P. Lyubartsev, *Eur. Biophys. J.*, 2005, **35**, 53–61.

[112] T. Murtola, E. Falck, M. Karttunen and I. Vattulainen, *J. Chem. Phys.*, 2007, **126**, 075101.

[113] S. Y. Mashayak, M. N. Jochum, K. Koschke, N. R. Aluru, V. Rühle and C. Junghans, *PLoS One*, 2015, **10**, e0131754.

[114] J. A. Nelder and R. Mead, *Comput. J.*, 1965, **7**, 308–313.

[115] N. Hansen, A. Ostermeier and A. Gawelczyk, Proceedings of the 6th International Conference on Genetic Algorithms, San Francisco, CA, USA, 1995, pp. 57–64.

[116] K. Farah, A. C. Fogarty, M. C. Böhm and F. Müller-Plathe, *Phys. Chem. Chem. Phys.*, 2011, **13**, 2894–2902.

[117] T. C. Moore, C. R. Iacovella and C. McCabe, *J. Chem. Phys.*, 2014, **140**, 224104.

[118] T. C. Moore, C. R. Iacovella, R. Hartkamp, A. L. Bunge and C. McCabe, *J. Phys. Chem. B*, 2016, **120**, 9944—-9958.

[119] W. G. Noid, J.-W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das and H. C. Andersen, *J. Chem. Phys.*, 2008, **128**, 244114.

[120] L. Lu, S. Izvekov, A. Das, H. C. Andersen and G. A. Voth, *J. Chem. Theory Comput.*, 2010, **6**, 954–965.

[121] W. G. Noid, P. Liu, Y. Wang, J.-W. Chu, G. S. Ayton, S. Izvekov, H. C. Andersen and G. A. Voth, *J. Chem. Phys.*, 2008, **128**, 244115.

[122] S. Izvekov and G. A. Voth, *J. Phys. Chem. B*, 2005, **109**, 2469–2473.

[123] S. Izvekov and G. A. Voth, *J. Chem. Theory Comput.*, 2006, **2**, 637–648.

[124] Y. Wang, S. Izvekov, T. Yan and G. A. Voth, *J. Phys. Chem. B*, 2006, **110**, 3564–3575.

[125] Z. Cao and G. A. Voth, *J. Chem. Phys.*, 2015, **143**, 243116.

[126] A. Das and H. C. Andersen, *J. Chem. Phys.*, 2012, **136**, 194113.

[127] A. Das and H. C. Andersen, *J. Chem. Phys.*, 2012, **136**, 194114.

[128] A. Das and H. C. Andersen, *J. Chem. Phys.*, 2010, **132**, 164106.

[129] S. Izvekov, P. W. Chung and B. M. Rice, *J. Chem. Phys.*, 2010, **133**, 064109.

[130] A. Das, L. Lu, H. C. Andersen and G. A. Voth, *J. Chem. Phys.*, 2012, **136**, 194115.

[131] L. Lu and G. A. Voth, *J. Phys. Chem. B*, 2009, **113**, 1501–1510.

[132] J. W. Mullinax and W. G. Noid, *Phys. Rev. Lett.*, 2009, **103**, 198104.

[133] J. F. Rudzinski and W. G. Noid, *Eur. Phys. J. Spec. Top.*, 2015, **224**, 2193–2216.

[134] C. R. Ellis, J. F. Rudzinski and W. G. Noid, *Macromol. Theory Simul.*, 2011, **20**, 478–495.

[135] J. F. Rudzinski and W. G. Noid, *J. Phys. Chem. B*, 2012, **116**, 8621–8635.

[136] K. Lu, J. F. Rudzinski, W. G. Noid, S. T. Milner and J. K. Maranas, *Soft Matter*, 2014, **10**, 978–989.

[137] E. Brini and N. F. A. van der Vegt, *J. Chem. Phys.*, 2012, **137**, 154113.

[138] S. P. Carmichael and M. S. Shell, *J. Phys. Chem. B*, 2012, **116**, 8383–8393.

[139] N. J. H. Dunn and W. G. Noid, *J. Chem. Phys.*, 2015, **143**, 243148.

[140] N. J. H. Dunn and W. G. Noid, *J. Chem. Phys.*, 2016, **144**, 204124.

[141] D. Rosenberger and N. F. A. v. d. Vegt, *Phys. Chem. Chem. Phys.*, 2018, **20**, 6617–6628.

[142] I. Pagonabarraga and D. Frenkel, *Mol. Simul.*, 2000, **25**, 167–175.

[143] T. Sanyal and M. S. Shell, *J. Chem. Phys.*, 2016, **145**, 034109.

[144] M. R. DeLyser and W. G. Noid, *J. Chem. Phys.*, 2017, **147**, 134111.

[145] T. Sanyal and M. S. Shell, *J. Phys. Chem. B*, 2018, **122**, 5678–5693.

[146] J. Jin and G. A. Voth, *J. Chem. Theory Comput.*, 2018, **14**, 2180–2197.

[147] K. R. Hadley and C. McCabe, *Mol. Simul.*, 2012, **38**, 671–681.

[148] S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A. E. Dawid and A. Kolinski, *Chem. Rev.*, 2016, **116**, 7898–7936.

[149] T. D. Potter, J. Tasche and M. R. Wilson, *Phys. Chem. Chem. Phys.*, 2019, **21**, 1912–1927.

[150] N. J. Boyd and M. R. Wilson, *Phys. Chem. Chem. Phys.*, 2015, **17**, 24851–24865.

[151] N. J. Boyd and M. R. Wilson, *Phys. Chem. Chem. Phys.*, 2018, **20**, 1485–1496.

[152] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess and E. Lindahl, *Bioinformatics*, 2013, **29**, 845–854.

[153] V. Papaioannou, T. Lafitte, C. Avendaño, C. S. Adjiman, G. Jackson, E. A. Müller and A. Galindo, *J. Chem. Phys.*, 2014, **140**, 054107.

[154] A. Mejía, C. Herdes and E. A. Müller, *Ind. Eng. Chem. Res.*, 2014, **53**, 4131–4141.

[155] L. Morávková, Z. Wagner and J. Linek, *J. Chem. Thermodyn.*, 2008, **40**, 607–617.

[156] J. H. Dymond and K. J. Young, *Int. J. Thermophys.*, 1981, **2**, 237–247.

[157] J. A. Barker and D. Henderson, *Molecular Physics*, 1971, **21**, 187–191.

[158] W. G. Noid, J.-W. Chu, G. S. Ayton and G. A. Voth, *J. Phys. Chem. B*, 2007, **111**, 4116–4127.

[159] C. C. Chappelow, P. S. Snyder and J. Winnick, *J. Chem. Eng. Data*, 1971, **16**, 440–442.

[160] L. Lu, J. F. Dama and G. A. Voth, *J. Chem. Phys.*, 2013, **139**, 121906.

[161] N. F. Carnahan and E. A. Müller, *Phys. Chem. Chem. Phys.*, 2006, **8**, 2619–2623.

[162] C. C. Walker, J. Genzer and E. E. Santiso, *J. Chem. Phys.*, 2019, **150**, 034901.

[163] Y. An, K. K. Bejagam and S. A. Deshmukh, *J. Phys. Chem. B*, 2018, **122**, 7143–7153.

[164] E. F. Sabattié, J. Tasche, M. R. Wilson, M. W. Skoda, A. Hughes, T. Lindner and R. L. Thompson, *Soft Matter*, 2017, **13**, 3580–3591.

[165] J. Tasche, *Ph.D. thesis*, Durham University, 2018.

[166] P. Ren and J. W. Ponder, *J. Phys. Chem. B*, 2003, **107**, 5933–5947.

[167] P. Warren, *Phys. Rev. E*, 2003, **68**, 066702.

[168] J. W. Wagner, T. Dannenhoffer-Lafage, J. Jin and G. A. Voth, *J. Chem. Phys.*, 2017, **147**, 044113.

[169] P. Mereghetti, G. Maccari, G. L. B. Spampinato and V. Tozzini, *J. Phys. Chem. B*, 2016, **120**, 8571–8579.

[170] L. Yang, L. Sun and W.-Q. Deng, *J. Phys. Chem. A*, 2018, **122**, 1672–1677.

[171] J. Tasche, unpublished work.

[172] L. Larini, L. Lu and G. A. Voth, *J. Chem. Phys.*, 2010, **132**, 164107.

[173] J. F. Dama, J. Jin and G. A. Voth, *J. Chem. Theory Comput.*, 2017, **13**, 1010–1022.

[174] W. G. Noid, *J. Chem. Phys.*, 2013, **139**, 090901.

[175] I. T. Todorov, W. Smith, K. Trachenko and M. T. Dove, *J. Mater. Chem.*, 2006, **16**, 1911–1918.

[176] V. Krishna, W. G. Noid and G. A. Voth, *J. Chem. Phys.*, 2009, **131**, 024103.

[177] L. Lu and G. A. Voth, *J. Chem. Phys.*, 2011, **134**, 224107.

[178] S. Chandrasekhar, *Liquid Crystals*, Cambridge University Press, 1992.

[179] M. J. Stephen and J. P. Straley, *Rev. Mod. Phys.*, 1974, **46**, 617–704.

[180] L. Onsager, *Ann. N. Y. Acad. Sci.*, 1949, **51**, 627–659.

[181] J. P. F. Lagerwall and G. Scalia, *Curr. Appl. Phys.*, 2012, **12**, 1387–1412.

[182] W. H. de Jeu, in *Phase Transitions in Liquid Crystals*, ed. S. Martellucci and A. N. Chester, Springer US, Boston, MA, 1992, pp. 3–16.

[183] L. Maibaum, A. R. Dinner and D. Chandler, *J. Phys. Chem. B*, 2004, **108**, 6778–6781.

[184] E. E. Meyer, K. J. Rosenberg and J. Israelachvili, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 15739–15746.

[185] S. T. Hyde, *Handb. Appl. Surf. Colloid Chem.*, 2001, **2**, 299–332.

[186] T. K. Attwood, J. E. Lydon, C. Hall and G. J. T. Tiddy, *Liq. Cryst.*, 1990, **7**, 657–668.

[187] J. Lydon, *Curr. Opin. Colloid Interface Sci.*, 1998, **3**, 458–466.

[188] J. Lydon, *J. Mater. Chem.*, 2010, **20**, 10071–10099.

[189] T. K. Attwood, J. E. Lydon and F. Jones, *Liq. Cryst.*, 1986, **1**, 499–507.

[190] D. F. Bradley and M. K. Wolf, *PNAS*, 1959, **45**, 944–952.

[191] J. Lydon, *Curr. Opin. Colloid Interface Sci.*, 2004, **8**, 480–490.

[192] M. Walker and M. R. Wilson, *Mol. Cryst. Liq. Cryst.*, 2015, **612**, 117–125.

[193] N. Boden, R. J. Bushby and C. Hardy, *J. Phyique Lett.*, 1985, **46**, 325–328.

[194] N. Boden, R. J. Bushby, C. Hardy and F. Sixl, *Chem. Phys. Lett.*, 1986, **123**, 359–364.

[195] N. Boden, R. J. Bushby, L. Ferris, C. Hardy and F. Sixl, *Liq. Cryst.*, 1986, **1**, 109–125.

[196] Z. H. Al-Lawati, B. Alkhairalla, J. P. Bramble, J. R. Henderson, R. J. Bushby and S. D. Evans, *J. Phys. Chem. C*, 2012, **116**, 12627–12635.

[197] R. Hughes, A. Smith, R. Bushby, B. Movaghar and N. Boden, *Mol. Cryst. Liq. Cryst. Sci. Technol., Sect. A*, 1999, **332**, 547–557.

[198] T. Bast and R. Hentschke, *J. Phys. Chem.*, 1996, **100**, 12162–12171.

[199] J. Hubbard, *PhD Thesis*, University of Leeds, 1997.

[200] F. Chami and M. R. Wilson, *J. Am. Chem. Soc.*, 2010, **132**, 7794–7802.

[201] R. G. Edwards, J. R. Henderson and R. L. Pinning, *Mol. Phys.*, 1995, **86**, 567–598.

[202] P. K. Maiti, Y. Lansac, M. A. Glaser and N. A. Clark, *Liq. Cryst.*, 2002, **29**, 619–626.

[203] R. Thind, *Phd Thesis*, Durham University, 2018.

[204] Martini 3 open-beta, `http://cgmartini.nl/index.php/martini3beta`, accessed July 2018.

[205] R. Hashim, G. R. Luckhurst and S. Romano, *Molecular Physics*, 1985, **56**, 1217–1234.

[206] K. R. Hadley and C. McCabe, *J. Phys. Chem. B*, 2010, **114**, 4590–4599.

[207] T. C. Moore, C. R. Iacovella and C. McCabe, in *Foundations of Molecular Modeling and Simulation: Select Papers from FOMMS 2015*, ed. R. Q. Snurr, C. S. Adjiman and D. A. Kofke, Springer Singapore, Singapore, 2016, pp. 37–52.

[208] B. van Hoof, A. J. Markvoort, R. A. van Santen and P. A. J. Hilbers, *J. Phys. Chem. B*, 2011, **115**, 10001–10012.

[209] O. Lobanova, *Ph.D. thesis*, Imperial College London, 2014.