Integrating polarisation effects into non-polarisable models to better model the self-assembly of mesoporous silica nanomaterials

PhD Thesis

Andrew Milne

Department of Chemical and Process Engineering

University of Strathclyde, Glasgow

April 8, 2020

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Abstract

Mesoporous silica nanomaterials are a class of materials of rapidly growing importance. Despite this, many details of their synthesis are still poorly understood. For this reason, computational studies have been widely used to further our understanding of the processes involved at the molecular level. Unfortunately, many of these models are either too generic or specialised for general use and neglect phenomena that play an important role in the synthesis process, such as polarisation. For this reason, this thesis had two main goals: 1) to develop new methodologies to integrate polarisation effects into fixed-charge force fields; 2) to employ these new insights into a generic, transferable force field for organosilica molecules. In this work, we demonstrated that the current modelling paradigm for water, the solvent in which mesoporous silica synthesis takes place, is fundamentally flawed when it comes to accounting for the cost of polarisation in the free energy. We have therefore proposed a new, more detailed analytical correction that can obviate this current failing of classical non-polarisable force fields. Subsequently, we carried out a detailed analysis of different methodologies for obtaining point charges suitable for the liquid phase from quantum-mechanical calculations. Finally, we employed quantum-mechanical calculations, molecular dynamics simulations and polarisation corrections to parametrise models for organosilica molecules. This was then used to propose a molecular model for silicic acid, the principal precursor in the synthesis of mesoporous silica, via transferability. In future, we plan to expand our proposed transferable force field to a wider range of organosilica molecules, thereby facilitating the simulation of the silica synthesis with a much wider range of organosilica precursors. This would be advantageous for the discovery of new, *in situ* functionalised mesoporous silica nanomaterials.

Contents

List of Figures

List of Tables

Acknowledgements

I am deeply thankful to a great many people who have made this thesis possible. In the first instance, I would like to thank my family, whose continual support has propped up my morale even when the work seemed never-ending. I am particularly grateful to my parents for giving me everything I ever needed to reach this far; without them, my academic adventure would have ended long before this point. I would also like to acknowledge all my wonderful colleagues throughout my doctoral studies who have made the last four years enriching personally, as well as academically. All our complaining about our awful results over the years made a great many things bearable when the light at the end of the tunnel seemed a long way off. In my own research group, I am particularly indebted to Maria Cecilia Barrera, whose help at times was critical. Last, but very much not least, I would like to thank my supervisor, Miguel Jorge. A great deal of what I have learnt during my time as a doctoral researcher derives directly from his enthusiasm and dedication, whether that has been in the office or over a glass of whisky.

Chapter 1

Introduction

Nanotechnology is a rapidly developing field that has long since transcended science fiction to become a reality; perusing *Nature Nanotechnology* for review papers only published in the first half of this year yields applications as diverse as fertilisers [1], sensors [2] and materials [3, 4]. One such class of materials that have drawn a lot of attention are mesoporous silica nanomaterials (MSN), whose discovery is generally credited to employees of Mobil [5, 6]. These are materials whose basic chemical structure is SiO_2 and have an amorphous structure with pore sizes of 2-50 nm. These materials are formed by a complex self-assembly process, whereby a solvent (principally water), an amphiphilic surfactant and an organosilica source interact to spontaneously form hexagonal structures, at the surface of which the silica source polymerises via a condensation reaction [7, 8]. The polymerisation locks the structure in place and, after the removal of the surfactant and solvent, a porous silica phase remains. The attention that these materials have garnered is largely due to the capability to functionalise, i.e. specialise, them [9–11]. Tuning the surfactant molecule, e.g. by shortening or lengthening the hydrophobic tail, can lead to control of the material's pore size and different organosilica sources can be used to embed different functional groups in the pore walls, whereby the chemical specificity can be altered [12]. As such, if the physics of the interactions involved in the synthesis of MSNs were to be well understood, then it would provide countless opportunities for the discovery of new materials for a wide range of applications [13–15].

Since the dawn of modern computing during World War II, computers have been employed to solve chemical problems; the first quantum-mechanical calculations were carried out as long ago as the 1950s, with the first Hartree-Fock calculations being carried out in 1956 at MIT [16]. The first molecular simulations followed not long after [17, 18]. Computational methods have the advantage over experimental studies that the properties, to the accuracy of the calculation or simulation, are enumerated precisely for every atom, thereby providing us with a high level of molecular-level insight. For example, it is impossible to know the velocity of every molecule in a physical experiment, while this is trivial in a computer simulation. Such knowledge of the underlying physics potentially allows us to predict and therefore control molecular-level, i.e. nanoscale, behaviour. With enough such knowledge, it would be possible to create materials with specific characteristics to a high degree of accuracy. However, the power of computational studies goes beyond purely characterising and directing the synthesis of an already existing material; ideally, *in silico* experiments would also have predictive power to search for new materials. This would allow for a large number of potential candidate materials to be screened before even setting foot in the lab, funnelling the practical synthesis of a desired nanomaterial to the most promising candidates. This has the potential to greatly expedite the discovery of new, targeted materials, making the process cheaper and faster.

Nevertheless, despite the monumental increases in computing power over the last few decades (in the two decades leading to 2010 a 10,000-fold increase in individual processor power was observed [19]), the ability to represent nature accurately with the highest level of theory available to us, namely quantum mechanics, is constrained to length and time scales that are vanishingly small by material-science standards. Density functional theory (DFT - itself a simplification of and quicker than *ab initio* approaches) studies are generally restricted to hundreds [20,21] up to a few thousand [22,23] atoms. This is a paltry quantity considering that a gram of a given material will generally have a number of atoms on the order of Avogadro's constant (10^{23}). For this reason, Newton's classical mechanics are very much alive and well, its relative simplicity making it, when certain caveats are accounted for, ideal for decreasing the computational burden and

therefore opening the door to larger systems and/or longer time scales.

Against this background, the term *multiscale modelling* [24–26] has become popular - this describes the use of higher levels of theory, e.g. quantum mechanics, to better inform the development of models at lower levels, e.g. molecular-level approaches such as molecular dynamics (MD) [18] and Monte Carlo (MC) [17]. Despite the heavy limitations on the application of quantum-mechanical approaches, they nevertheless provide snapshots of rigorous detail into molecular-level behaviour, which, when interpreted appropriately, can then be integrated into molecular models to provide them with a solid theoretical basis. Alternatively, molecular models can be improved via analytical calculations [27,28] that separate out some computationally arduous component of a calculation or simulation and replace it with an approximate analytical correction, thereby ameliorating the computational burden. As such, molecular simulation is a technique that can nowadays accommodate simulations up to hundreds of thousand of atoms and millisecond time scales [29], although such extreme conditions necessitate high-performance computing (HPC) facilities. In **Chapter 2**, the principles behind quantum-mechanical and molecular-simulation methods will be elucidated. This will then be followed by a literature review of computational approaches to studying silica self-assembly in **Chapter 3**.

One such physical phenomenon that presents complexities for molecular simulation models is polarisation [30–32]; this describes how the electron cloud of a molecule reacts to its environment, resulting in momentarily induced dipole and higher-multipole moments. Polarisation effects also provide a pertinent case study for how the computational cost of a simulation increases sharply as its mathematical representation tends to reality. Some water models do explicitly account for polarisation effects [33], albeit in a simplified manner, and said models are slower than non-polarisable models by a factor of 3-10 times [34]. Thus it can be seen how adding an extra physical phenomenon into a classical model introduces a large decrease in computational tractability. Whether this extra computational burden is disproportionate or not depends on the system in question; if the properties of interest are strongly dependent on the effect in question then it might be necessary. However, for the majority of systems, this will not be the

case.

Large systems, of the type necessary to study silica self-assembly processes, are generally too large to explicitly incorporate such polarisation, especially for water, which makes up the bulk of the molecules for many systems. This could be problematic in the case of silica synthesis, where charge screening, i.e. an electrostatic phenomenon, is a key driving force [7, 35]. Additionally, since self-assembly is a phase-equilibrium problem [7, 36], free energy quantifies this driving force and the degree of polarisation will directly affect this. As such, this work intends to investigate the following themes, in order to deepen our understanding of the fundamental physics of the self-assembly of porous silica materials:

- The effect of polarisation on the self-solvation free energy of water and therefore also on the free energy of the phase equilibria involved in self-assembly
- Deriving charges from quantum-mechanical methods that correspond to an appropriate level of polarisation in pure liquids and solutions
- Using the insights developed in the first two sections, to begin to parametrise a transferable organosilica force field for use in studying the functionalisation of mesoporous silica nanomaterials

These areas will be dealt with in **Chapters 4**, **5** and **6**, respectively. While the first two are relevant to mesoporous silica synthesis, the principles are also widely applicable to any modelled system where polarisation is an important factor and where charges need to be derived for liquid-phase models. As such, this work has the potential to be influential in the study of liquid systems in general, i.e. well beyond the scope of silica self-assembly.

There are two publications associated with this work:

- Jorge, Miguel; Milne, Andrew W.; Sobek, Olivia N.; Centi, Alessia; Perez-Sanchez, German; Gomes, Jose R. B. (**2018**), "*Modelling the Self-Assembly of Silica-Based Mesoporous Materials*", Molecular Simulation 44, 435-452

- Milne, Andrew W.; Jorge, Miguel (**2019**), "*Polarization Corrections and the Hydration Free Energy of Water*", J. Chem. Theory Comput. 15, 2, 1065-1078

Chapter 2

Theory and Methods

2.1 Classical and Quantum Theories of Matter

At the beginning of the 20th century, a new theory of matter began to emerge, namely that of quantum mechanics (QM). It was shown that atoms were comprised of even smaller components, so-called subatomic particles. Until that point, it had been believed that the atom was the smallest fundamental unit of matter; the word atom derives itself from the Greek word for "uncuttable". Since QM describes the fundamental smallest energy scale (so-called "quanta"), it allows for the most accurate description of matter, albeit at great computational cost. Due to this complexity, even with modern computing power, it remains a near impossible task to describe even the simplest macroscopic properties using QM. For this reason, the much simpler, therefore more computationally tractable, classical (i.e., atomic) mechanics still enjoys widespread use. The reason for this is that, despite its deficiencies, classical theory does a good job of describing matter at the atomic level, where quantum effects can generally be considered negligible.

In multiscale modelling, multiple levels of theory are applied together in order to bridge different time and length scales. In this work, we are mainly interested in using detailed QM calculations to accurately parametrise classical molecular models. As such, both theories will be important in the context of this work and the next sections will hence look at them more closely.

2.2 Classical Mechanics and Molecular Simulation Techniques

Classical mechanics is largely attributed to Isaac Newton, hence it is sometimes referred to as Newtonian mechanics. Classical mechanics generally treats objects as point particles, i.e. of negligible size, an assumption that holds exactly when using an object's centre of mass. The classical approximation tends to break down for hydrogen (i.e., the smallest possible molecule) at low temperatures but is otherwise an excellent proxy for molecular-level particle behaviour.

Molecular simulation techniques are usually split into two categorie: molecular dynamics (MD) and Monte Carlo (MC). In the first section, we will discuss MD, which is a numerical technique whereby Newton's (classical) equations of motion are solved by numerical integration. We will then describe MC simulations, a stochastic technique, i.e. relying on probabilities.

2.2.1 Molecular Dynamics

Classical mechanics can be described succinctly by Newton's equations of motion. Newton's first law states that if the net force ($\mathbf{F}$) acting on a particle is equal to zero, then the acceleration ($\mathrm{d}\mathbf{V}/\mathrm{dt} = \mathbf{a}$) of the particle will also be equal to zero, i.e. its velocity ($\mathbf{V}$) will remain constant. Mathematically,

$$\mathbf{F} = 0 \Longleftrightarrow \frac{d\mathbf{V}}{dt} = 0 \tag{2.1}$$

If a net force is applied, then this is described by Newton's second law

$$\mathbf{F} = \frac{d\mathbf{p}}{dt} = m\mathbf{a} \tag{2.2}$$

where $\mathbf{p}$ is momentum ($=m\mathbf{V}$), which is a conserved property for any closed system. This shows that the acceleration of the object is directly proportional to the net force, where the proportionality constant is the mass of the object. Finally, the third law describes how every force produces an equal and opposite reaction:

$$\mathbf{F_i} = -\mathbf{F_j} \tag{2.3}$$

In MD simulations, these equations of motion are solved numerically. This means that they are not solved as a single set of equations over the desired time frame (i.e., continuously) but as many related sets of equations at different points in time. The breaking up of time from a continuous function into a finite set of points is called discretisation. The collection of results at different discretised points in time then approximate the true (continuous) result. The advantage of such a method is that it replaces a single extremely complex problem, whose analytical solution in such a formulation is infeasible, with many much simpler problems. Mathematically, for a generic property (x)

$$x = \int_{t=0}^{T} f(t)dt \approx \sum_{t=0}^{T} f(t)\Delta t \tag{2.4}$$

Δt represents the difference between two points in time, i.e. how much time is incremented between points; in MD simulations, this is known as the time step. This method, numerical integration, is in practice an approximation of the continuous solution; with an infinite number of points, the two solutions would be identical. Since an infinite number of time steps is obviously impossible, a compromise has to be made. Smaller time steps are more accurate but larger time steps allow greater time scales to be simulated, upon which the observation of relevant properties is often dependent. Millisecond time scales are the current state-of-the-art [29].

Although it is known to be easier to solve MD simulations in a discretised manner, this does not tell us how exactly this should be achieved. In practice, the forces are calculated at every time step, which then yield velocities via Newton's laws of motion. Next, the velocity and the time step are multiplied to give us the displacement for that period of time. This is repeated over many time steps in order to produce a trajectory of the movement of the particles/molecules in question [37–39].

The precise method chosen to carry out this numerical integration is called an *integrator*. The Euler integrator is the most basic example [40] and correspondingly has

several flaws. It is simply a Taylor expansion of the position vector $\mathbf{x_i}$

$$\mathbf{x_i}(t+\Delta t) = \mathbf{x_i}(t) + \frac{d\mathbf{x_i}(t)}{dt}\Delta t + \frac{1}{2}\frac{d^2\mathbf{x_i}(t)}{dt^2}(\Delta t)^2 + O((\Delta t)^3) \tag{2.5}$$

where O gives the truncation error in the Taylor expansion, while $O((\Delta t)^3)$ informs us of the order of the error term. Since $\frac{d\mathbf{x_i}(t)}{dt} = \mathbf{V_i}$ and $\frac{d^2\mathbf{x_i}(t)}{dt^2} = \mathbf{a_i} = \frac{\mathbf{F_i}}{m}$, we can rewrite it as

$$\mathbf{x_i}(t+\Delta t) = \mathbf{x_i}(t) + \mathbf{V_i}\Delta t + \frac{1}{2}\frac{\mathbf{F_i}}{m}(\Delta t)^2 + O((\Delta t)^3) \tag{2.6}$$

Correspondingly, the velocity is

$$\mathbf{V_i}(t+\Delta t) = \mathbf{V_i}(t) + \frac{\mathbf{F_i}}{m}\Delta t + O((\Delta t)^2) \tag{2.7}$$

First of all, its result diverges from the real solution in finite time when unbounded potentials are used, e.g. a potential that only tends to zero at infinite distance but will never equal it, and hence bounding it within limits will cause the loss of part of the true result. This is the case for the prevailingly used Lennard-Jones and electrostatic intermolecular potentials (which will be discussed shortly) and hence is unfavourable for molecular systems. Another disadvantage is that the Euler integrator does not exhibit time reversibility since it has no knowledge of previous steps; this means that reversing the velocities of all the molecules in a simulation will not necessarily lead back to the starting configuration, which is clearly not consistent with the deterministic nature of an ideal MD simulation. Finally, the Euler integrator does not demonstrate phase-space conservation. This means that the relative density of states in phase space does not stay constant over the system's trajectory, as is required by Liouville's theorem [41].

A major improvement on the above method is the Verlet algorithm [42]. It does this by solving the future state of the system with the previous (not the current) state. Consider **Equations 2.6** and **2.8**

$$\mathbf{x_i}(t-\Delta t) = \mathbf{x_i}(t) - \mathbf{V_i}\Delta t + \frac{1}{2}\frac{\mathbf{F_i}}{m}(\Delta t)^2 - O((\Delta t)^3) \tag{2.8}$$

The addition of these two equations gives us

$$\mathbf{x_i}(t+\Delta t) = 2\mathbf{x_i}(t) + \mathbf{x_i}(t-\Delta t) + \frac{\mathbf{F_i}}{m}(\Delta t)^2 + O((\Delta t)^4) \tag{2.9}$$

The consequences of this formulation are twofold: 1) the two $O((\Delta t)^3)$ terms cancel making the error term a magnitude lower than before; 2) since the past positions of the system are incorporated, the algorithm is time reversible. However, the cancelling of the velocity term leads to some difficulties, principally that the velocity of the current state is only known in the subsequent step. Since the velocity is necessary for calculating properties of the system, such as kinetic energy and therefore temperature, it is necessary to introduce approximations to account for this. A way to circumvent this is to use the so-called velocity-Verlet algorithm.

The position form in velocity Verlet [43] is the same as in **Equation 2.6**; the past state of the system is accounted for in the velocities instead

$$\mathbf{V_i}(t+\Delta t) = \mathbf{V_i}(t) + \frac{\mathbf{F_i}(t) + \mathbf{F_i}(t+\Delta t)}{m}\Delta t + O((\Delta t)^3) \tag{2.10}$$

This algorithm is algebraically the same as the Verlet algorithm but with instantaneous velocities also being calculated. As such, it has become one of the most widely used algorithms in MD simulations.

Another very popular algorithm is the leapfrog algorithm [39]. It has the following form

$$\mathbf{V_i}(t+\frac{1}{2}\Delta t) = \mathbf{V_i}(t-\frac{1}{2}\Delta t) + \frac{\Delta t}{m_i}\mathbf{F_i}(t) \tag{2.11}$$

$$\mathbf{x_i}(t+\frac{1}{2}\Delta t) = \mathbf{x_i}(t) + \mathbf{V_i}(t+\frac{1}{2}\Delta t)\Delta t \tag{2.12}$$

The leapfrog algorithm is the default integrator for all versions of GROMACS [44], the chosen MD package for this work, up to the present version (version 2019.3). Nevertheless, the velocity-Verlet algorithm is generally preferred when high accuracy in temperature- or pressure controlled-simulations is necessary [39]. Otherwise, given

equivalent starting coordinates and velocities, the velocity-Verlet and leapfrog algorithms will produce identical trajectories and have the same benefits - they are symplectic (phase-space conserving) and time reversible.

2.2.2 Molecular Interactions

The way that different chemical behaviour is incorporated into molecular simulations is via the method for the computation of the forces (in molecular dynamics) or energies (in Monte Carlo). For this, a forcefield is required; this is a set of equations that determine the forces or energies between any two interaction centres (atoms/group of atoms). This can be of any form as long as it captures the correct physical behaviour between any two interaction centres for any displacement. These functions are usually given in terms of the potential energy, which is then related to the overall force by

$$\mathbf{F_i} = \sum_j \frac{dU_{ij}(\mathbf{r_{ij}})}{\mathbf{r_{ij}}} \tag{2.13}$$

where i and j are two interaction centres and $\mathbf{r_{ij}}$ is the displacement between them. Note the sum instead of the integral, since MD is a numerical technique.

Intermolecular interactions can be split into two categories, namely short- and long-range interactions [45], and the various types of each are listed in **Table 2.1**. Short-range interactions are those that take place when there is significant overlap between wave functions. As such, they decay sharply as the displacement between the two interacting particles increases, adhering to a mathematical relationship akin to $\exp(-\alpha r)$, where α is a constant and r is the displacement. Due to this exponential form, such contributions are negligible at a short distance away from a particle. In contrast to this, long-range interactions have the form r^{-n}, where n is a constant. This means that they decay much more slowly and their impact is often non-negligible even at a respectable distance from the particle.

Of these intermolecular interactions, some are generally much more important for classical, molecular simulations. For example, magnetic interactions are very small and exchange-repulsion strongly dominates over the other short-range interactions at short

Table 2.1: Types of intermolecular interactions between atoms/molecules. Adapted from Stone [45]

Interaction	Sign	Type
Electrostatic	$\pm$	Long-range
Induction	$-$	
Dispersion	$-$	
Resonance	$\pm$	
Magnetic	$\pm$	
Exchange-repulsion	$+$	Short-range
Exchange-induction	$-$	
Exchange-dispersion	$-$	
Charge transfer	$-$	

range. As such, we are principally interested in the electrostatic, induction, dispersion and exchange-repulsion contributions.

Electrostatic contributions are those that are caused by the interaction of two permanent (i.e., that of the isolated state) charge distributions. Induction (often referred to as "polarisation" - we will adopt this term for the rest of this work) is the distortion of a molecule's charge distribution by an electric field. This can either be external, i.e. that applied in an experiment, or internal, i.e. an electric field precipitated by the presence of a collection of molecules [33]. It is the second case that is more common and also most relevant for this work. There are three modes by which this can take place: electronic, geometrical and orientational polarisation. In electronic polarisation, the electron density itself is altered; in the geometrical variant, the molecular geometry (i.e., the nuclear distribution) changes and in orientational polarisation, the alignment of a molecule with a non-zero multipole moment (i.e., one that is polar) relative to an external field [33] occurs. Next, the London dispersion (or simply "dispersion") interaction arises from the correlation of instantaneous deformations to the charge distribution. The electrons of a molecule are constantly in motion, meaning that the electron distribution is also in a state of constant flux. This then causes a knock-on effect in other molecules, whose own electronic distribution will deform to align itself more favourably with its new environment. Finally, exchange-repulsion (often shortened to simply "repulsion") arises when the electron clouds of two molecules overlap at short range, incurring a strong energy penalty due to the need to deform the wavefunction in order to not violate the

Pauli exclusion principle [45].

The relative strength of all these interactions depends on the properties of the molecule in question. For example, electrostatics are very important in systems where the constituent atoms have large differences in electronegativity. This leads to electrons being more likely to be located in certain areas of a molecule. For this reason, these types of interactions are particularly important in ionic (e.g., $NaCl$) and polar covalent systems (H_2O). Polarisation also tends to be important in similar systems, where the collection of molecules with anisotropic distributions of electrons produces an electric field, polarising further the electron distribution of other molecules. For example, the permanent dipole of water (i.e., in the gas phase) is 1.85 D [46], whereas in the liquid phase, i.e. after entering the electric field induced by an assembly of water molecules, this is polarised to approximately 2.9 D [47, 48].

Dispersion is always present since all electrons are in constant motion, thereby creating instantaneous multipoles in all molecules. However, these interactions are usually weaker than electrostatic interactions in polar/ionic molecules and tend to have a similar magnitude to that of induction. Dispersion therefore dominates in non-polar systems such as hydrocarbons. Finally, since it only occurs at short range, repulsion would only dominate for condensed systems under high pressure.

2.2.3 Implementation of Molecular Interactions in Simulations

In the classical theory of matter, the atom is the smallest unit, precluding the inclusion of the motion of individual electrons. Since the aforementioned interactions all arise in one way or another from the electron distribution of an atom or molecule, how are they represented in classical simulations? In MD, molecular interactions are split into two categories - intermolecular ("non-bonded") and intramolecular ("bonded"). Bonded potentials are those that only occur within a molecule, whereas non-bonded ones arise both between different molecules and, despite the implication following from the term "intermolecular interactions", between different parts of the same molecule. This distinction is fundamentally artificial, since all interactions arise from the same phenomena, however it allows functional forms to be simplified, i.e. their computational

tractability to be improved. Intermolecular interactions are therefore excluded between neighbouring atoms (so-called 1-2 interactions) and those in a chain separated by only one other atom (1-3). Additionally, 1-4 interactions may also be excluded depending on the chosen force-field implementation. These excluded interactions are then replaced by bonded interactions.

Bonded Interactions

There are three types of bonded interactions that are commonly employed in molecular simulations: bond stretching, angle bending and dihedral torsion. Bond stretching (B) and angle bending (θ) are normally represented by

$$U_B = \frac{k_B}{2}(r - r_0)^2 \tag{2.14}$$

where U_B is the potential energy penalty arising from bond stretching, k_B is the force constant, while r and r_0 are the actual and equilibrium values of the bond length between two atoms;

$$U_\theta = \frac{k_\theta}{2}(\theta - \theta_0)^2 \tag{2.15}$$

where U_θ is the potential energy penalty from angle bending, k_θ is the force constant, while θ and θ_0 are the actual and equilibrium values of the angle between three bonded atoms. These formulations are mainly used due to their computational tractability rather than their accuracy. For example, in real molecules, at a certain distance, the bond joining two atoms would break, distributing the electron density between the two atoms. However, as can be seen in **Equations 2.14** and **2.15**, this is not allowed to occur in standard MD simulations. The reason for this is that MD is not intended to sample very high energy states, meaning that being able to describe such chemical behaviour is generally not necessary [137]. Additionally, the above potential would make it seem that bond stretching and angle bending are symmetric. In reality, this is not the case - the potential will be more repulsive when the bond shrinks and less so when it expands. For bond stretching, more physically realistic behaviour is captured by the

Morse potential, which has the form

$$U_B = W(1 - \exp(-a(r - r_0)))^2 \tag{2.16}$$

where W is the well depth and a controls the width of the potential. At very high values of r, U_B will tend to W, i.e. moving the atoms further apart will effectively no longer incur a stretching penalty because the atoms will have dissociated. At this point, the attractive force between the two atoms is decaying extremely slowly. At the other end of the scale, i.e. when the atoms come very close together, the term grows exponentially due to repulsion. Despite the higher level of physical realism of the Morse potential, its harmonic counterpart has one less fitting parameter and is usually adequate for small deviations around the equilibrium value. Due to the strength of chemical bonds, small deviations around the equilibrium value will comprise the vast majority of those found in molecular simulations and the harmonic potential is therefore routinely preferred to the Morse potential.

Dihedral torsion (T) is the most complicated of the three bonded interactions and describes twisting around a bond. A common form of this potential, used in the GROMACS MD package [39], is

$$U_T = \sum_{n=0}^{5} k_{T,n}(cos(\psi))^n \tag{2.17}$$

where U_T stands for the potential energy penalty, ψ is the dihedral angle and $k_{T,n}$ are a set of up to six constants, although this is restricted to the first four terms in most cases. For a chain of atoms $i-j-k-l$, the dihedral angle is that between the planes formed by $i-j-k$ and $j-k-l$. Dihedrals are especially important in larger molecules, such as proteins, where the correct conformation of the molecule may be essential for describing its behaviour accurately. The dihedral potentials penalise moving away from the equilibrium conformation, thereby enforcing it. As noted above, the number of bonded exclusions depends on the force field, but non-bonded interactions are always included for chain separations of three atoms and above. The main question, therefore, is how to deal with 1-4 interactions. For some force fields, they are excluded or included entirely. For others, they are present but scaled by a factor, e.g. 0.5, to smooth the

transition between included and excluded interactions. There is also a second type of dihedral function, called an improper dihedral, whose function is to enforce chirality. However, their application is relatively rare and we do not them employ them in this work.

Non-Bonded Interactions

Dispersion and repulsion are usually grouped together into one potential, the most widely used of which is the Lennard-Jones (LJ) potential. It takes the form

$$U_{ij} = 4\varepsilon_{ij}\left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right) \tag{2.18}$$

where ε is the well depth (the maximum attractive interaction between the two interaction centres) and σ is the van der Waals radius (the "size" of the atoms in question). The r^{-6} term represents dispersion and is isotropic here. This is acceptable since the motion of electrons is a lot faster than the nuclei, so they will only feel the average effect of these oscillations. If the oscillations are assumed to be isotropic, i.e. the probability function of finding an electron is spherically symmetric further from the centre of mass, then the dispersion potential will also be isotropic. Since this is the case for the noble gases, where dispersion dominates, this is generally taken to be a valid assumption [49]. Using this assumption, London [49] also showed that the magnitude of dispersion had a spatial dependence of r^{-6}, providing the basis for this form in the LJ potential. The r^{-12} term, on the other hand, was chosen for computational reasons, rather than due to any theoretical underpinnings; as mentioned above, it should be of the form $\exp^{-\alpha r}$. However, the r^{-12} term can be factorised, as shown in **Equation 2.19**, simplifying the calculation. This made it cheaper than the Buckingham potential, which uses a more realistic exponential form for repulsion. However, while this was relevant in previous eras, where computational power was at a premium, the LJ potential has remained the most popular intermolecular potential due to its continued success over the years and the large quantity of previously parametrised models employing it.

$$U_{LJ,ij} = 4\varepsilon_{ij}\left(\left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^6\right)^2 - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^6\right) \tag{2.19}$$

The behaviour and form of the LJ potential are best illustrated by considering **Figure 2.1**, where it is plotted for two carboxylate oxygen atoms according to the parameters of the OPLS force field [50]. The ε value is 0.878 kJ/mol, while σ corresponds to 0.296 nm. Looking at **Figure 2.1**, it can be seen that the minimum corresponds to 0.878 kJ/mol, i.e. the well depth (ε). 0.296 nm (σ) is the x-intersect; this is because the electron clouds of the two atoms start to overlap around σ, hence this is the part of the potential that corresponds to repulsion. Meanwhile, the negative part of the potential corresponds to the attractive dispersion interaction, arising from the formation and interaction of instantaneous induced multipoles. From **Figure 2.1**, the reader may notice a similarity between the LJ potential and the Morse potential introduced in the previous section. Both increase exponentially at distances below that of the minimum, whereas in the opposite case the function slowly reaches a plateau. This a good demonstration of the fact that bonded and non-bonded interactions nevertheless describe the same types of interactions. However, unlike for bond stretching, for which the variation in bond lengths in a simulation will be relatively minor, the LJ potential must be accurate over a much larger range of distances between interaction centres and therefore necessitates a more complex functional form.

It can be seen that the tail of the function in **Figure 2.1** gradually and monotonically tends to zero as the atom-atom displacement tends to infinity. In practice, this is undesirable, since it means that it is impossible to calculate exactly, since the simulated system cannot be of infinite size and hence some part of the dispersion term is always lost. In practice, a cut-off term is applied, i.e. interactions that occur further than a certain displacement apart are neglected. This cut-off can either be abrupt (instantaneously becoming equal to zero at the cut-off) or can be a smoothed function. Since, as can be seen in **Figure 2.1**, the tail quickly becomes a small value, this is often sufficient. However, in situations where this might nevertheless be important (such as in free energy calculations [39]), a long-range dispersion correction is applied. For systems that are homogeneous past the cut-off value, the value of the LJ tail can be

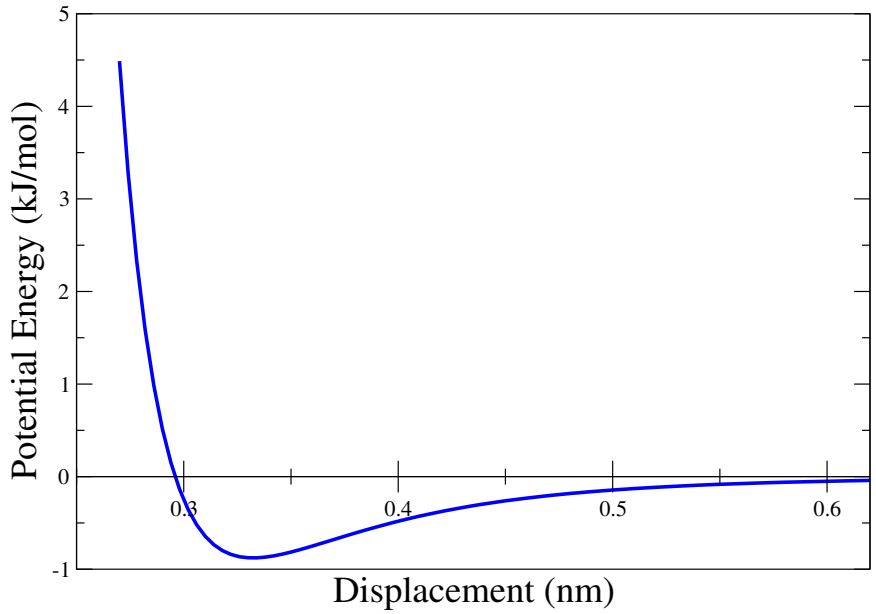


Figure 2.1: The Lennard-Jones potential for oxygen in the carboxylate functional group in the OPLS-AA forcefield ($\sigma = 0.296\,;nm$, $\varepsilon = 0.878\ kJ/mol$) [50]. It shows the potential energy between two oxygen atoms in different molecules derived from dispersion and repulsion

calculated analytically [39]. However, for systems for which this criterion is not met, more advanced techniques may be applied. These will be discussed in the next section.

In all the equations in this section, the constants are labelled "ij", i.e. each interaction between any two atom types has its own constants. In practice, it would be extremely arduous to parametrise every single possible pairwise interaction. Instead, so-called combining rules are employed that estimate the interaction constants for unlike atom types from those for like interactions. These combining rules are specific to the potential for which they are used. For example, the most common combining rules for the LJ potential are the Lorentz-Berthelot rules [51]:

$$\sigma_{ij} = \frac{\sigma_{ii} + \sigma jj}{2} \tag{2.20}$$

$$\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}} \tag{2.21}$$

Despite their ubiquity, these rules are not exact; several works have been published highlighting their limitations [52–54]. Indeed, other combining rules have been suggested for the LJ function, including the Waldman-Hagler [55], Fender-Halsey [56] and Kong [57] rules. Nevertheless, any set of such rules is fundamentally an approximation, however good.

The final intermolecular potential that is usually taken into account in MD simulations is electrostatics. This describes the interaction between permanent molecular multipoles, i.e. static distributions of electrons in space. These are usually approximated as partial charges located on interaction sites. By convention, charges of the same sign interact repulsively (positive potential energy), whereas those of opposite sign attract each other (negative potential energy). The usual way of calculating these energies is by using Coulomb's Law

$$U_{ES,ij} = \frac{q_i q_j}{4\pi\epsilon r_{ij}} \tag{2.22}$$

where q is the charge of the interaction centre, r is the displacement between the two and ϵ is the permittivity of the medium. Normally ϵ is made equal to the vacuum permittivity, ϵ_0 $(= 5.726 \times 10^{-4})$. The electrostatic contribution decays much more slowly than that of the LJ potential, leading to significantly higher cut-off errors. We will now move on to discussing two innovations that are often implemented to negate these errors.

2.2.4 Periodic Boundary Conditions and Ewald Summation

Due to computational limitations, it is only possible to simulate a small subset of molecules compared to the whole macroscopic system. This means that the system will be of finite size and that there will have to be a boundary somewhere in the system. Two problems arise from this decision: 1) it introduces strong boundary effects into bulk systems where these effects are negligible; 2) for long-range interactions, a considerable part of the potential may lie outside our chosen system boundary. In this section, we will demonstrate how these problems are overcome practically in molecular simulations.

The solution to the first part is the implementation periodic boundary conditions

(PBCs) [37]. The idea behind PBCs is simple; the system is surrounded by copies of itself, known as images. This effectively creates an infinitely repeating lattice, with the system acting as the unit cell. This will create periodic effects into the simulation; for continuously repeating solids, e.g. crystals, this is desired since the material is itself periodic. However, for liquids this is undesirable due to them having no long-range structure. Nevertheless, these effects are much smaller than the consequences of having no or free boundaries [39]. Furthermore, it is relatively easy to test the magnitude of this bias by running simulations with unit cells of different sizes. These are known as finite-size effects and larger unit cells will decrease their influence on a given simulation.

Naturally, the particles cannot be allowed to interact with every particle in every other image since this would multiply the desired number of interactions. For this reason, every particle is only allowed to interact with one image of every other particle. However, this doesn't tell us which image of another particle it should interact with. Normally, the minimum-image convention is used, under which a particle only interacts with the closest image of any other. See **Figure 2.2** for a two-dimensional example of PBCs and the minimum-image convention.

A further potential problem is that in a small system a particle could interact with its own image, which is obviously highly unphysical. Since intermolecular interactions decay very slowly, it would be impossible to create a simulation where the unit cell was sufficiently large to avoid this phenomenon. In practice, however, the tail intermolecular interactions are truncated after a certain range (usually on the scale of a nanometre). In this way, as long as the unit cell is longer than this truncation, particles will not interact with their own images. As noted before, PBs are not boundaries as such since particles can freely move in and out of the unit cell; when this happens its image on the other side enters the box and the simulation continues.

In some cases, when we are not dealing with a bulk system, PBCs are not desired, since surface effects are part of the behaviour that we wish to study. For example, we may have a surface which is not repeated vertically, i.e. the bulk should not be sandwiched between surfaces but open. In the context of this work, we are only interested in the bulk behaviour of physical systems and hence we will only apply PBCs in all three

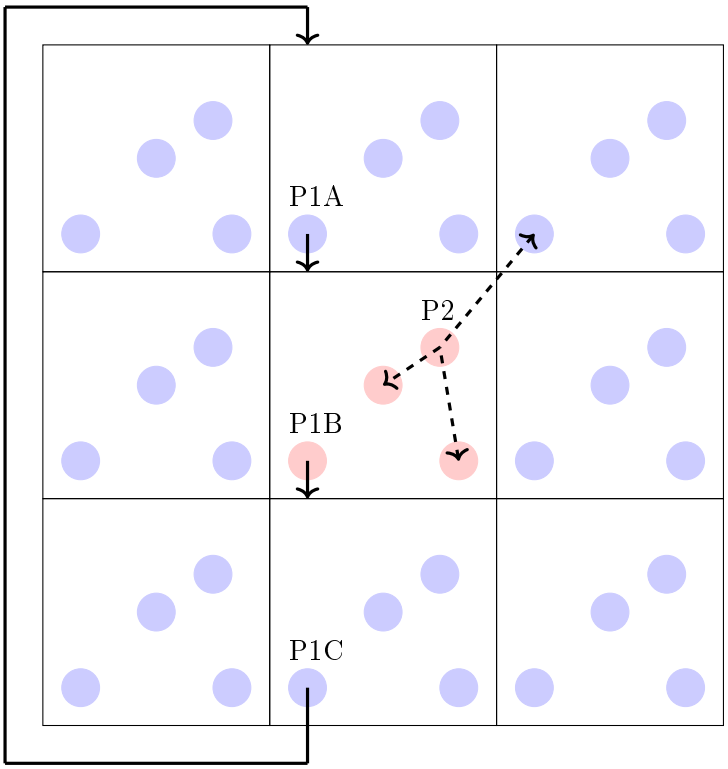


Figure 2.2: Two-dimensional example of periodic boundary conditions. The central square with pink particles represents the unit cell of the simulation; the surrounding squares with the blue particles depict copies of the unit cell. The particles labelled P1 illustrate the periodic nature of PBCs. All three P1 particles have the same velocity (unbroken lines; that of P1C loops around for illustrative purposes only) and therefore, as P1B leaves through one side of the unit cell, its copy in the cell behind it (P1A) enters the abandoned cell to replace it. P1A will then itself be replaced by P1C. The particle labelled P2 demonstrates the minimum-image convention; it interacts directly, i.e. in the short-range part of the potential, only with the closest image of any other particle, as illustrated by the dashed lines

dimensions.

As stated above, if an interaction potential is simply truncated at the box boundary, then for longer-range potentials a non-negligible part of the interaction may be unaccounted for. The Coulomb potential decays very slowly (r^{-1}), much more so than dispersion (in the LJ formulation: r^{-6}) or repulsion (r^{-12}), and therefore the calculation of long-range potentials in the literature has primarily focussed on this interaction. The simple cut-off method for electrostatic contributions has been shown to lead to inaccuracies in calculating properties such as the internal energy, heat of adsorption, pressure and surface tension [58]. This means that electrostatic interactions also require corrections in order to avoid producing artefacts due to truncation.

Ewald summation [59] was developed to study long-range electrostatic forces in periodic crystals. On the surface, this may seem far removed from a widely applicable solution for molecular simulations in general. Its applicability comes from the use of PBCs in molecular simulations. The replication of the simulation box causes it to become the "unit cell" of a periodic system, i.e. as in a crystal.

Ewald summation works by splitting one slowly converging function into two that are rapidly convergent and one constant term [58,60,61]. Formally,

$$U_{ES} = U_{SR} + U_{LR} + U_0 \tag{2.23}$$

where U_{SR} and U_{LR} are the short-range and long-range parts of the electrostatic potential, respectively, whereas U_0 is the constant term, commonly known as the self-term [60]. It has the form

$$U_0 = \frac{-f\beta}{\sqrt{\pi}} \sum_{i=1}^{N} q_i^2 \tag{2.24}$$

and cancels out any unphysical interactions arising from splitting the potential function [60]. $f = \frac{1}{4\pi\epsilon_0}$ and β is a tunable parameter that determines the weighting of how much of the total electrostatic potential should be calculated in the two sums, U_{SR} (SR = short-range) and U_{LR} (LR = long-range); the importance of this parameter will be explained in more detail shortly. The remaining two terms can be represented mathematically as [58]

$$U_{SR} + U_{LR} = \frac{f}{2} \sum_{\mathbf{L}} \sum_{ij}^{N} \frac{q_i q_j}{|\mathbf{r_{ij}} + \mathbf{L}|} = \frac{f}{2} \sum_{\mathbf{L}} \sum_{ij}^{N} \frac{q_i q_j \phi(r_{ij})}{|\mathbf{r_{ij}} + \mathbf{L}|} + \frac{f}{2} \sum_{\mathbf{L}} \sum_{ij}^{N} \frac{q_i q_j (1 - \phi(r_{ij}))}{|\mathbf{r_{ij}} + \mathbf{L}|} \tag{2.25}$$

where $\mathbf{L}$ are the lattice vectors and $\phi(r_{ij})$ is a convergence function that splits the total energy between the two functions so that it is conserved. The reason that this approach is so effective is that one of these functions is evaluated in real space, whereas the other is computed in Fourier, also known as reciprocal, space. As vectors get longer in real space, they become smaller in reciprocal space, as the name implies. This means that the function in real space converges quickly at small distances but slowly at longer distances; for the function in Fourier space, the case is exactly the opposite. In this way, each part of the Coulomb potential is evaluated in that space that allows it to be computed most efficiently, enabling a high level of accuracy to be obtained via a significantly lower computational burden.

The switch from one function to the other is controlled via a cut-off at a pre-specified boundary. This is usually taken to be equal to that of the unit cell, i.e. the cut-off for the direct part is half the box size for the simulation. In this way, the real-space function represents the electrostatic energy arising from interactions between interaction centres that both lie in the unit cell. The reciprocal-space potential represents the long-range component of the electrostatic potential, i.e. that produced by interactions between the interaction centres in the unit cell and those in the periodic images. The classical convergence function for Ewald summation is the error function, rendering the Coulomb potential as [62]

$$U_{SR} + U_{LR} = \frac{f}{2} \sum_{\mathbf{L}} \sum_{ij}^{N} \frac{q_i q_j}{|\mathbf{r_{ij}} + \mathbf{L}|} erfc(\beta \mathbf{r_{ij}}) + \frac{f}{2} \sum_{\mathbf{L}} \sum_{ij}^{N} \frac{q_i q_j}{|\mathbf{r_{ij}} + \mathbf{L}|} erf(\beta \mathbf{r_{ij}}) \tag{2.26}$$

where erf represents the error function, which has the form

$$erf(x) \equiv \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt \tag{2.27}$$

and erfc has the property

$$erfc(x) \equiv 1 - erf(x) \tag{2.28}$$

thereby fulfilling the criterion for the convergence function. For rapid convergence, it is necessary to choose β parameter carefully. The optimum strategy is to choose β so that each component takes on the same amount of computational effort [58]. In this way, the scalability of Ewald summation can be reduced from N^2 to $N^{\frac{3}{2}}$ [60]. In reciprocal space, the long-range part (the second term in **Equation 2.26**) becomes [62]

$$U_{LR} = \frac{1}{2}\frac{4\pi f}{V_{cell}}\frac{q_i q_j}{k^2}\exp(\frac{-k^2}{4\beta^2}) \tag{2.29}$$

where V_{cell} is the volume of the unit cell are k and plane waves. Ewald summation has two requirements, namely that the overall charge of the system is zero and that the system is periodic. Usually the periodicity is three-dimensional, but the technique has also been generalised to one- and two-dimensional cases [63].

Despite its mathematical elegance, the original Ewald approach leaves somewhat to be desired computationally. As stated above, the scalability of the Ewald sum with perfect parametrisation is $N^{\frac{3}{2}}$. For most molecular simuations, the computational effort is approximately proportional to the number of interaction centres, i.e. is of order N. This means that as the number of interaction centres increases, the Ewald sums rapidly begin to require the majority of the computational effort to carry out the simulation. For example, for scaling a simulation from 10 to 1,000 atoms, the Ewald sums would take 1,000 times as long, whereas the computational time for the rest of the simulation would only rise by a factor of 100. In this way, it is obvious that this approach will be incredibly inefficient for many applications. For this reason, Ewald sums are not usually implemented in the classical manner.

The breakthrough for the use of the Ewald method arrived in the implementation of fast Fourier transforms (FFTs). These are algorithms that reduce the scalability of Fourier transforms from N^2 (like that seen above) down to $N \log N$ [60, 64, 65]. This is favourable because the rate at which the $\log N$ part grows decreases as N increases. For example, in our previous case of increasing the system size of a simulation from 10

to 1,000 atoms, the increase in computational cost for the Ewald sum was 1,000, but this decreases to a 200-fold increase for an approach employing FFTs.

The first approach applying FFTs to long-range interactions was the particle-particle particle-mesh (P3M) method of Hockney and Eastwood [66]. They split the electrostatic energy into local and long-range contributions. The local contribution is calculated explicitly (i.e., particle-particle), whereas the long-range part was approximated by the interaction of the particles with a mesh of the same charge density. Although this method was a breakthrough in computational efficiency, it was initially difficult to achieve a high level of accuracy with it [65]. Darden et al. [64] subsequently used a similar, i.e. particle-mesh, approach but changed the switching function of the P3M method to that of Ewald sums, thereby uniting Ewald summation and FFTs. This approach is therefore termed the particle-mesh Ewald (PME) method. Finally, by changing the method chosen for interpolation of the mesh, Essmann et al. [65] developed an algorithm in which the long-range electrostatic potential was smoothly rather than piecewise differentiable, leading to the moniker smooth PME (SPME). This property of the new formulation allowed the forces arising from the potential of the reciprocal sum to be calculated analytically. Previously, the forces and energy had to be calculated separately, leading to the conservation of energy being violated [65]. The cost of this improvement are small local violations in the conservation of momentum [61]. Nevertheless, the SPME scheme provides better accuracy in fewer computational steps than P3M [61]. This new approach for the interpolation of the mesh was later also applied to P3M to create P3M - analytical differentiation (P3M-AD) [39]. It has been shown by Ballennegger et al. [61] that SPME is actually just a special case of the more general P3M-AD; if the splitting function for P3M-AD is chosen to be the error function (see **Equation 2.27**) then the two approaches are strictly equivalent.

SPME is today the most popular way of calculating long-range interactions in molecular simulations, being standard in programs such as GROMACS (the simulation package of choice in this work), AMBER and NAMD [39,61]. PME is noticeably superior to classical Ewald summation in terms of computational speed for medium to large systems. For small systems, classical Ewald summation may be more efficient due to

scaling not being a significant issue and it obviating the need to set up a mesh and subject it to a Fourier transform [39]. Classical Ewald summation may also be more efficient for MC simulations; only a few atoms are usually perturbed per simulation step, meaning that recalculating the whole particle mesh at every time step is inefficient [58]. Classical Ewald summation, on the other hand, can simply retain and reuse large parts of the reciprocal-space sum. Although SPME is relatively computationally efficient, it is nonetheless not trivial, which is why it is not (yet) the norm for correcting dispersion [58]. Attempts to implement pairwise (i.e., of order N scalability) long-range corrections have not yet been shown to be as accurate or more efficient than SPME [58]. SPME, therefore, remains the gold standard for correcting cut-off errors in electrostatic interactions.

Polarisation

As mentioned previously in **Section 2.2.2**, polarisation is the response of a molecule's charge distribution to an electric field [33]. A charge distribution for a space can be represented conveniently by a multipole expansion. This is an expansion in terms of distance from an origin of the potential energy due to Coulomb's Law. This results in [33]

$$U_T(\mathbf{r}) = U_1(\mathbf{r}) + U_2(\mathbf{r}) + U_3(\mathbf{r}) + ... \tag{2.30}$$

where U_T is the total potential energy and the numbered subscripts refer to the power of the inverse r term in each part. For example, the leading term is

$$U_1 = \frac{1}{4\pi\epsilon_0}\frac{q}{r} \tag{2.31}$$

where the term q represents the total charge of the system and is the *monopole*, i.e. the first moment, of the system. This is only important for systems that are not charge neutral. In this work, all systems employed are charge neutral. The second term is more instructive for our purposes and is

$$U_2 = \frac{1}{4\pi\epsilon_0 r^3}\mathbf{r}\mu_i \tag{2.32}$$

where μ is the dipole-moment tensor, i.e. a vector describing the distribution of charge in all three spatial dimensions. Namely,

$$\mu_i = \begin{pmatrix} \mu_x \\ \mu_y \\ \mu_z \end{pmatrix} \tag{2.33}$$

μ_i is calculated as

$$\mu_i = \sum_l r_{i,l} q_l \tag{2.34}$$

where i refers to the coordinate of the atom along the i-axis and l refers to the atom in the molecule. Normally, it is assumed that the leading term in the expansion will dominate. Since the monopole is usually equal to zero, this means that the dipole term is usually the most important term and in many cases is where the expansion is cut off [33]. One of the points that we will touch upon in this work is if this assumption is sufficient in the calculation of correction terms for free energy calculations. As such, we will also include quadrupoles. These are of the form

$$U_3 = \frac{1}{4\pi\epsilon_0 r^5}\mathbf{r^2}\Theta \tag{2.35}$$

where the quadrupole tensor is a 3×3 matrix since it describes a distribution on a plane. Its components are given by

$$\Theta_{ij} = \sum_l q_l(3r_{i,l}r_{j,l} - ||\mathbf{r_l}||^2\delta_{ij}) \tag{2.36}$$

where i and j refer to the coordinates of atom l along the i- and j-axes, respectively. $\delta_i j$ is the Kronecker delta and $||\mathbf{r_l}||$ is the magnitude of the vector $\mathbf{r_l}$.

Polarisation is by far the most complicated of the interactions mentioned thus far to include explicitly. This is because the polarisation of a molecule is a product of its environment, i.e. of the aggregate of the molecules surrounding it. This means that it is

a many-body interaction. Since the number of many-body interactions scales in a much less computationally favourable way than pairwise interactions, this means that calculating polarisation "on-the-fly" is usually not carried out. This means that the charges on the individual atoms are fixed and do not respond to their environment. This is not a major problem when the model is used in only one relatively homogeneous state, since parameters can be tweaked to cancel out this neglected effect. This approach is known as the mean-field approach [33] and models employing this methodology are commonly known as *non-polarisable* models because the effect of polarisation is taken into account implicitly rather than explicitly, i.e. rather than carrying out any calculations along the trajectory based on the current state of the system. This is achieved in practice by increasing the dipole of the model from that of the gas-phase dipole moment to one that represents the effective dipole (μ_{eff}) of the aggregate state. For example, most commonly used water models (e.g., SPC, SPC/E, TIP3P and TIP4P) have a dipole moment between 2.2 and 2.4 D (usually arrived at by optimising the charges against experimental data), cf. that of the gas phase is 1.86 D [46]. However, the liquid-phase dipole (μ_{liquid}) of water has been shown to be around 2.9 D [47,48], so the model values would seem to be at odds with the liquid-phase values. However, it has been proposed that this apparent inconsistency is actually a result of most molecular simulations being carried out with a dielectric constant equal to that of vacuum, thereby ignoring purely electronic screening in the medium [30]. Leontyev and Stuchebrukhov were able to approximately reconcile the model and liquid dipoles using the simple scaling relation

$$\mu_{eff} = \frac{\mu_{liquid}}{\sqrt{\epsilon_\infty}} \tag{2.37}$$

where ϵ_∞ is the high-frequency dielectric constant. Therefore, non-polarisable models use an effective, i.e. mean, dipole value so as to approximate the aggregate effect of polarisation. Due to its simplicity, this is the most commonly used way of dealing with polarisation in condensed phases since it doesn't involve any extra calculations. Instead, it simply scales up the electrostatic interactions that are already present. For water, this generally works well for ambient conditions but less well for other phases and for binding to larger molecules [33]. This leads to problems when properties are

computed involving a change between multiple states, e.g. heats of vaporisation. For example, a water molecule in the gas and liquid phases will have the same point charges in a non-polarisable model, unlike in reality, where the liquid water molecule's electron distribution will be distorted by its neighbours. Hence, when comparing the energies of the simulations for each phase, the classical MD simulation will not be able to describe this difference. In order to take this into account in classical force fields, a correction term is usually applied to account for this. The most commonly used correction is that of Berendsen [31]

$$\Delta E_{pol} = \frac{(\mu_M - \mu_G)^2}{2\alpha} \tag{2.38}$$

where μ_M is the dipole magnitude of the fixed-charge model, μ_G that of the gas and α is the isotropic polarisability of the molecule. This correction assumes that the model dipole is a good proxy for the liquid dipole and that the energy to change the dipole from one appropriate to the gas to the liquid phase dominates the total of the polarisation energy. This type of correction, along with recent proposed improvements, will be discussed further in **Chapter 4**.

Despite the additional computational expense, (explicitly) polarisable force fields do exist [67]. They are mostly used for biomolecular simulations, since many biomolecules' structures can only be simulated by using explicit polarisation. This is because biomolecules usually exist in water, a highly polarising solvent, therefore the necessary structure is usually also polarised [68]. There are three types of polarisable models available in widely-used MD codes [33,67]:

1. Fluctuating Charges
2. Drude Oscillators (Charges on Springs)
3. Inducible Dipoles/Higher Multipoles

Fluctuating charge models are based on electronegativity, which is defined as the change in energy as the charge in a system changes [69]. From an expansion of the

energy of an atom with respect to charge, i.e. the energy change in creating partial charge on an atom, it follows that [69–71]

$$U(q) = U(0) + q(\frac{\partial U}{\partial q})_{q=0} + \frac{q^2}{2}(\frac{\partial^2 U}{\partial q^2})_{q=0} + ... \tag{2.39}$$

where $U(0)$ is a constant, termed C_i. The first derivative is equal to Mulliken's definition of electronegativity (χ_i^0) and the second is twice the hardness (J_i^0) of the atom [70]. The total electrostatic potential energy for a fluctuating charge system is therefore [33]

$$U_{el} = \sum_i (C_i + \chi_i^0 q_i + \frac{1}{2} J_i^0 q_i^2) + \sum_i \sum_{j>i} \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}} \tag{2.40}$$

The second piece of electronegativity theory drawn upon is electronegativity equalisation, i.e. all atoms in a molecule will equalise their electronegativity so as to achieve the same value [69–71]. Since electronegativity is defined as

$$\chi_i = \frac{\partial U}{\partial q} \tag{2.41}$$

the charges that minimise the electrostatic potential energy, while subject to overall charge neutrality, are those that also satisfy the principle of electronegativity equalisation [70]. This is the basis of fluctuating charge models. The TIP4P-FQ model is an example of this subclass [70]. The advantages of this method are its simplicity (it doesn't add any extra interactions compared to standard MD simulations), whereas its main drawback is that it only allows polarisation in the direction of bonds, which is therefore constrained by the geometry of the molecule [33, 67].

Drude-oscillator-type models attach an additional massless charge centre (representing the electron cloud) via a spring to the principal atom centre (representing the nucleus) [33, 67]. Once again, this method is relatively simple, but it also doubles the amount of electrostatic interactions. The reward for this extra complexity is that it incorporates a degree of anisotropy. The equation for the force on the spring is simply

$$\mathbf{F_S} = k_S \Delta \mathbf{r} \tag{2.42}$$

where $\mathbf{F_S}$ is the force on the spring, k_S is the force constant of the spring and $\Delta\mathbf{r}$ is the distance between the virtual polarisable site and the fixed atom to which it is attached. If the spring is at rest, then [33]

$$\mathbf{F_S} = \mathbf{F_{el}} \tag{2.43}$$

where F_{el} is the force due to the electronic interactions. This is equal to

$$\mathbf{F_{el}} = q_p\mathbf{E} \tag{2.44}$$

where q_p the charge on the moveable additional charge. Therefore the distance of the virtual charge must be equal to

$$\Delta\mathbf{r} = \frac{q_p}{k_S}\mathbf{E} \tag{2.45}$$

where $\mathbf{E}$ is the electric field. The only unknown left is the value for k_S. This can be found via the relationship for the induced dipoles, namely

$$\Delta\mu = q\Delta\mathbf{r} = \alpha\mathbf{E} \tag{2.46}$$

where $\Delta\mu$ is the induced dipole and α is the isotropic polarisability. Using these two identities, it follows that

$$k_S = \frac{q_p^2}{\alpha} \tag{2.47}$$

The addition of the virtual site does not add any new types of interactions; it only complicates existing ones. Adding an extra, for water a fourth, massless interaction site increases the number of Coulombic terms. Additionally, the Berendsen polarisation correction has to be calculated for every molecule in every frame due to the changing moments, rather than just applying that for the average dipole moment value at the end.

Force fields employing inducible dipoles add these alongside the partial charges according to **Equation 2.46**. Unlike the previous two approaches, this adds new kinds of energy terms, since the multipoles are represented explicitly, rather than modifying

existing partial charges or adding new ones. Therefore, not all of the electrostatics can be described by Coulombic potentials. The advantage of this is that it is fairly easy to parametrise, although it is more difficult to implement in standard MD codes (although it is present in AMBER). Methods using multipole electrostatics are the most advanced since they employ not just monopoles (point charges), but also dipoles and quadrupoles. In this way, they can describe highly anisotropic behaviour without any extra parameters (e.g., Drude oscillators include extra charge sites for lone pairs). The most well-known model of this type is the AMOEBA package [72], which builds in polarisation by using induced dipoles at all points. While polarisable models have made great progress in the last few years [73, 74], it remains to be shown that the decrease in sampling (i.e., the shorter simulation time) compared to their non-polarisable brethren is worth the extra chemical detail [34].

2.2.5 Statistical Mechanics

The full set of all possible configurations for any real material would have a gargantuan number of degrees of freedom, thus making it impossible to sample any appreciable collection thereof, even with all the computing power in the world. In practice, this means that computers can only deal with systems up to millions of interaction centres, i.e. microscopic systems. This then begs the question - how can we link the information provided by our microscopic systems to macroscopic ones, which are useful for real-world applications? The link turns out to be *statistical mechanics*.

Statistical mechanics, as the name implies, deals with a statistical treatment of matter exposed to forces/undergoing motion. In a molecular dynamics simulation we have the advantage, relative to experiment, of being able to exactly tabulate the velocities and positions of all the matter being simulated. These velocities and positions are not constant, i.e. they are constantly fluctuating, yet they give rise to aggregate behaviour which is often static, i.e. equilibrium properties. Statistical mechanics applies statistics and probabilities to show that these seemingly static states arise from averages of the dynamic behaviour of collections of molecules.

One of the first things that has to be defined in statistical mechanics is how the states

of a system are formally enumerated using positions and momenta. For this purpose, a concept called *phase space* is used. The position and momentum of a particle can each be considered as a three-dimensional space. If the two spaces are independent, i.e. the momentum is not dependent on the position, then our particle can be exactly defined on a six-dimensional hypercube [75]. This unified formulation of momentum and position is known as phase space. However, QM tells us that position and momentum are in fact not independent, according to Heisenberg's uncertainty principle [76]. Mathematically [77],

$$\sigma_x \sigma_p \geq \frac{\hbar}{2} \tag{2.48}$$

where σ is the standard deviation in the position and momentum, respectively, and $\hbar$ is the reduced Planck constant. This means that the standard deviation, i.e. the uncertainty, of both position and momentum cannot be simultaneously zero; in other words, we cannot precisely define a point on our hypercube. As such, when we talk of a point in phase space we really mean a volume large enough that it is outside this quantum limit. This is all very well for a single particle, but we are mainly interested in systems of many bodies. If these N particles are statistically independent we then have a $6N$-dimensional phase space, a.k.a. Γ-space. All systems in statistical mechanics are defined in this phase space [75].

In total, there are three "flavours" of statistical mechanics: Maxwell-Boltzmann (based on classical mechanics), Fermi-Dirac and Bose-Einstein (both quantum-mechanical) [75]. In the classical case, the particles under study can be assumed to be distinguishable because the degree of interparticle spacing is deemed to be large enough so that quantum effects are negligible. Bosons and fermions, contrary to classical particles, are indistinguishable, i.e. identical, and hence their path cannot be traced. This is due to Heisenberg's uncertainty principle that states you cannot know the position and momentum of a quantum particle simultaneously. This means that if two particles were to swap places we could not distinguish them since we have incomplete knowledge of their paths. Classical particles can be considered statistically independent (we can state the probability of a single particle being in a given state without reference to the rest of the individual particles) since their wave functions, i.e. electronic orbitals, are localised,

meaning that the particles are comparatively weakly interacting.

The difference between the two quantum variants is that Fermi-Dirac statistics deals with Fermions (particles with half-integer spins) and Bose-Einstein with Bosons (integer spins). The practical relevance of this distinction is that the Pauli exclusion principle is applied to the former and not the latter; this states that two particles cannot share exactly the same quantum state, i.e. be in the same position with the same value of spin. Since this will limit the number of possible microstates, it will naturally have an effect on the statistical accounting of these states. For the rest of this brief introduction, we will restrict ourselves to classical statistical mechanics since it is the form that will be applied in this work.

One of the key concepts in statistical mechanics are statistical ensembles, which are probability distributions of all possible states that a system could occupy. This is then comprised of the degeneracy of a system (i.e., the number of possible independent states) and an appropriate weighting (how likely it is that this specific state will occur compared to others). In order to define this, it is necessary to fix sets of conserved macroscopic variables, the combinations thereof often lending the individual ensembles their names. Examples can be seen in **Figure 2.3**.

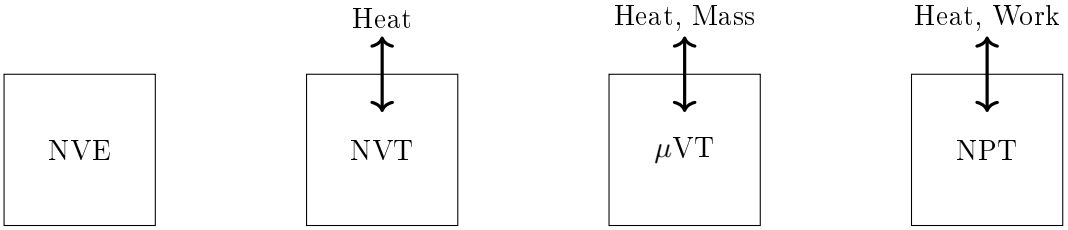


Figure 2.3: Summary of common thermodynamic ensembles, where the edge of the box represents the system boundary. From left to right: the microcanonical, canonical, grand canonical and isothermal-isobaric ensembles

The simplest ensemble is the microcanonical, where the composition (number of molecules - N), volume (V) and energy (E) define the state of the system. This ensemble corresponds to an isolated system, i.e. a system in which no transfer of mass or energy into or out of the system occurs. This means that the thermodynamic variable that defines the equilibrium of an isolated system is the entropy, since for a closed system

$$dU + PdV \leq TdS \tag{2.49}$$

meaning that if U, V and N (since V is extensive) are held constant, S will also be constant. Since it is also known that S is at a maximum for an isolated system, we can therefore say that S completely defines such a system at equilibrium [75].

The reason for the microcanonical ensemble being treated as the most fundamental example is that one of the principle tenets of statistical mechanics is that all states in a system with fixed N, V and E are equally probable. This is known as the assumption of *a priori probabilities* and follows from the *principle of indifference* [37]. That is to say that

$$p_i = \frac{1}{\Omega} \tag{2.50}$$

where p_i is the probability of state i occurring and Ω is the degeneracy of the system. This means that if we were to enumerate all the possible states of a system to find the macroscopic value, A, it would be calculated as

$$\langle A \rangle = \frac{\sum_j A_i}{\Omega} \tag{2.51}$$

where $\langle A \rangle$ is the *ensemble average* and A_i is the property in state i. If every possible state were to be enumerated, A and $\langle A \rangle$ would be identical. However, this is rarely the case since the number of microstates is astronomically large. As such, methods that employ statistical mechanics have the task of representatively sampling ensembles as efficiently as possible. This is made simpler by the fact that in most practical situations not all microstates are equally likely as in the microcanonical ensemble. In real life, the energy of a system is almost never constant and hence, despite its theoretical purity, the NVE ensemble is seldom of practical relevance.

An ensemble that is more experimentally relevant is the *canonical ensemble*. In this ensemble, composition, volume and temperature are the conserved properties (NVT). This corresponds to a system where no mass transfer over the system boundary may occur, but weak heat transfer may, so long as it does not bring the system out of

equilibrium. For this formulation, the probability of a given state is

$$p_i = \frac{1}{Z} exp(\frac{-E_i}{K_B T}) \tag{2.52}$$

where E_i is the energy of state i, K_B is the Boltzmann constant and T is the absolute temperature. Z represents the *canonical partition function* and has the form

$$Z = exp(\frac{-F}{K_B T}) \tag{2.53}$$

where F is the Helmholtz free energy. F is chosen since it is the thermodynamic variable that defines the closed system, much like S for the isolated system; a closed system is at equilibrium if F is at a minimum and at fixed N, V and T it will be a constant, defining the system. The term $exp(\frac{-E}{K_B T})$ is known as the Boltzmann factor, which acts as a weighting factor, while the partition function is the sum of all the individual weighting factors and hence acts as a normalisation factor. Therefore, we could also write

$$p_i = exp(\frac{F - E_i}{K_B T}) \tag{2.54}$$

and

$$exp(\frac{-F}{K_B T}) = \sum_j exp(\frac{-E_i}{K_B T}) \tag{2.55}$$

Two other common statistical ensembles are the grand canonical (μVT) and the isobaric-isothermal (NPT). The former is equivalent to an open system where both heat and mass can cross the system boundary, whereas the latter represents a closed system where energy can be exchanged by heat or work. In the grand canonical ensemble, the appropriate thermodynamic potential is the grand potential (Ψ) and for the isothermal-isobaric the Gibbs free energy is used (G).

Earlier, the concept of an ensemble average was briefly introduced, being denoted by angular brackets. However in MD simulations the possible ensembles are not sampled directly but are rather inferred by the time evolution of the system. The assumption that the collection of states visited during the trajectory is equivalent to the collection of

ensembles is called the *ergodic hypothesis*. Stated simply, if an infinite trajectory of one initial state was simulated for an ergodic system, then relative states in the trajectory would be visited in the same proportion as predicted from the relative ensemble. In reality, this is sometimes not the case. For this reason, it has been shown in recent years that the errors in some properties calculated from MD, such as the solvation free energy [78,79], can be significantly reduced by running several shorter simulations from different initial conditions, rather than one long simulation. This theory is related to Liouville's theorem, stating that the phase-space density is constant in time, i.e. along a trajectory the relative probabilities of configurations do not change [41]. Due to these discrepancies, time-averaged properties are denoted differently to ensemble-averaged ones; instead of angle brackets, time-averaged properties are signified by a bar over the top, i.e. $\bar{A}$.

2.2.6 Free Energy Calculations

One of the most useful properties that can be obtained from molecular simulations is the free energy change for a given process. The free energy change of a proposed process is useful because it informs us as to whether a thermodynamic transformation will occur spontaneously, i.e. without the need for any external energetic input. A negative value of the free energy change, indicating that the final state of the process has a lower free energy than the initial state, shows that this change can spontaneously occur. However, this might not necessarily happen in practice; it assumes that there is not another state with an even lower free energy and that the pathway is not kinetically hindered. This is because the state with the lowest free energy is the configuration that will be assumed in the limit of infinite time. Therefore, it may not be possible to reach the true thermodynamic equilibrium state within a reasonable amount of computational time.

Two applications of this class of calculations are ligand-binding and, as will be used in this work, solvation processes. In ligand-binding calculations, we want to know to which part of a given biomolecule a particular molecule will bind to. This is done by calculating the free energy of different complexes, i.e. when the molecule binds

to different sites on the biomolecule, in order to see which has the lowest energy free energy and hence would be most likely to be formed. This is commonly used in the pharmaceutical domain to investigate whether proposed molecules can target a specific biological receptor. Solvation calculations are similar; the free energy of an ensemble of solvent molecules is calculated with and without the presence of a solute molecule. If the introduction of the solute results in a system with a lower free energy, then the solute is soluble (at least to some degree). As the fall in the free energy becomes more marked, the degree of solubility increases. This can be important when solubility is competitive.

Above in **Section 2.2.5**, the role of the free energy in a statistical-mechanical system was touched upon. Indeed, we can trivially derive the following expression for a system in the canonical (NVT) ensemble from **Equation 2.53**

$$F = -\beta^{-1} \ln(Z) \tag{2.56}$$

where β is the thermodynamic beta ($= \frac{1}{K_B T}$). This demonstrates that calculating free energies from a statistical-mechanical perspective is equivalent to calculating partition functions. Since calculating the partition function directly is infeasible for all but the simplest systems, this would suggest that there is a problem with this approach. However, this difficulty can be circumvented by not calculating partition functions, but rather by estimating the difference between two of them. Formally, for a process beginning with state A and ending with state B

$$\Delta F_{BA} = F_B - F_A = -\beta^{-1} \ln(\frac{Z_B}{Z_A}) \tag{2.57}$$

Having rewritten the problem as finding a ratio of partition functions, rather than a single partition function, **Equation 2.57** is the basis of all free energy calculations. Estimators differ in the way that they calculate this ratio. Like all molecular simulations, free energy calculations are comprised of three components: a Hamiltonian (i.e., a force field in MD simulations), a sampling strategy and the application of theory from which we can derive the relevant property of interest from our simulation outputs. We will

now briefly discuss the two free energy estimators used in this work so as to give the reader a brief introduction to free energy calculations.

Force Estimation - Thermodynamic Integration

One simple way to calculate that shown in **Equation 2.57** is to integrate along a path between the two end states. The path may correspond to a real or imaginary, i.e. alchemical, process since the free energy is a state function, i.e. is path independent. The availability of the alchemical approach is advantageous because it allows the simulator the flexibility to choose the best approach. To control the mutation of the system from state A to B, it is necessary to choose a parameter that defines the progress along that path, i.e. a coupling parameter (λ), such that [80]

$$\Delta F_{BA} = F(\lambda_B) - F(\lambda_A) = \int_{\lambda_A}^{\lambda_B} \frac{dF}{d\lambda} d\lambda = \int_{\lambda_A}^{\lambda_B} \langle \frac{dH}{d\lambda} \rangle_\lambda d\lambda \tag{2.58}$$

where H is the Hamiltonian of the system. λ is usually set to zero at one end point and unity at the other. What λ does is to join the two individual Hamiltonians of A and B so as to create a combined Hamiltonian. The simplest way to do this would be a simple linear combination, i.e.

$$H_{combined} = \lambda H_B + (1 - \lambda) H_A \tag{2.59}$$

This will mean that at $\lambda = 0$ the system will be equivalent to our starting state, A, whereas at $\lambda = 1$ we will have achieved our terminal state, B. In the case of the solvation free energy that we wish to calculate in this work, the addition of the solute is simulated by having the intermolecular interactions of the solute molecule made equal to zero at one end of the λ-values, i.e. "absent", and gradually increasing the interaction strength via the path of the λ-values so that the molecule in question has been "dissolved". In practice, this is achieved by running simulations at various λ-values and then integrating over this range. This approach is referred to as "staging" and will be the sampling protocol that we will employ in this work. However, care has to be taken with this approach so as to avoid numerical instabilities due to the singularity in

the commonly used LJ potential as the molecule is created or annihilated. This is due to the $\langle \frac{dH}{d\lambda} \rangle_\lambda$ being very large due to the repulsive tail of the LJ potential [81]. Since numerical integration is used for evaluating the free energy change, having a similar λ-spacing for the whole range would artificially cause the repulsive tail section to dominate. For this reason non-linear approaches (e.g., usually a much smaller spacing in the tail region to minimise the integration error) or soft-core potentials are often employed to prevent this [39,80,81]. Soft-core potentials for free energy calculations were pioneered by Beutler et al. and their original formulation was [82]

$$U_{LJ,ij} = 4\epsilon_{ij}(1-\lambda)([\alpha_{soft}\lambda^2 + (\frac{r_{ij}}{\sigma_{ij}})^6]^{-2} - [\alpha_{soft}\lambda^2 + (\frac{r_{ij}}{\sigma_{ij}})^6]^{-1}) \tag{2.60}$$

where α_{soft} controls the "softness" of the repulsive tail. Other soft-core potentials are available but function similarly. As can be seen from **Equation 2.60**, the LJ function in this formulation will disappear when $\lambda = 1$ due to the $1-\lambda$ term, while the addition of the $\alpha_{soft}\lambda^2$ component will strongly damp the function when $\frac{r_iJ}{\sigma_{ij}}$ is small. However, this also brings further complications since it may introduce instabilities in the Coulomb part of the potential [81]. This can be mitigated by also using a soft-core formulation for the electrostatic part of the intermolecular potential. Another approach is to split the calculation into two parts: in the first part, one potential is gradually switched on or off; in the second stage, this is then done for the remaining component. The first approach is more economical, whereas the latter is simpler.

Energy Difference Methods - Bennett Acceptance Ratio

The other free energy estimator that we will use in this work is the Bennet Acceptance Ratio (BAR) [83], which is based on calculating energy differences between states, rather than integrating forces as in TI. The simplest of this class of methods is free energy perturbation (FEP). This applies the Zwanzig equation [84] that relates the free energy change to the exponential average of the energy difference between said states

$$\Delta F_{BA} = -\frac{1}{\beta} \ln\langle \exp[-\beta \Delta V_{BA}] \rangle_A \tag{2.61}$$

The exponential term here is dominant and hence the configurations with the lowest values of V_{BA} will strongly outweigh other contributions. Therefore if we sample two very different states we are unlikely to sufficiently sample those states where the energy difference is small, i.e. important. Once more, staging is often employed as the sampling strategy (as in this work), i.e. the calculation into broken into many smaller transitions (i.e., perturbations) where the difference between the two states is also smaller. There will therefore be greater overlap in the energy difference distribution, thereby optimising the sampling [80]. Other common sampling protocols are *importance sampling* and *adiabatic decoupling*. In importance sampling, a biasing term is introduced to the system so as to increase the chance of states with a high statistical weight being visited. The most-well known method of this type is *umbrella sampling*, as introduced by Torrie and Valleau [85], which employs a bias in the potential energy. In adiabatic decoupling, the path along λ is decoupled from all other degrees of freedom by using fictitious masses [80].

BAR also employs perturbations but, instead of using just one from A to B, it perturbs both A and B to an intermediate state I, i.e. according to a modified version of the Zwanzig equation

$$\Delta F_{BA} = \Delta F_{BI} - \Delta F_{AI} = -\frac{1}{\beta} \ln \frac{\langle \exp[-\beta \Delta V_{IA}] \rangle_A}{\langle \exp[-\beta \Delta V_{IB}] \rangle_B} \tag{2.62}$$

The optimal intermediate state, i.e. that at which the energy difference distributions, ρ_A and ρ_B, overlap, is then found in a postprocessing step. This follows from the identity

$$\frac{\rho_A}{\rho_B} = \frac{\exp(-\beta \Delta H)}{\exp(-\beta \Delta F_{BA})} \tag{2.63}$$

that states that the difference in the Hamiltonians is equal to the difference in free energy at the point of intersection, i.e. $\rho_A = \rho_B$ [80]. Nevertheless, between the starting and final states there is often minimal overlap, making adequate sampling prohibitively expensive. For this reason, BAR calculations are also usually staged and broken down into a chain of simulations that gradually mutate state A into B, as in TI. Non-linear sampling or soft-core potentials are also usually necessary.

2.2.7 Thermostats and Barostats in Molecular Dynamics

An MD simulation carried out with the algorithms presented earlier would be in the microcanonical ensemble; they are only solving the equations of motion and consequently do not include energy transfer from the surroundings, i.e. the simulation is as a closed system. As already noted, the microcanonical (NVE) ensemble is often impractical for real-world applications and therefore its canonical (NVT) or isothermal-isobaric (NPT) cousins are generally preferred. However, this necessitates new components of the algorithms that regulate temperature and/or pressure - these are named thermostats and barostats, respectively.

Thermostats

Due to the obvious practical demand for MD simulations to be carried out under constant temperature conditions, thermostats are ubiquitous and it is therefore unsurprising that several variants have been developed. The simplest is velocity rescaling [86]

$$\mathbf{V_{new}} = \lambda \mathbf{V} \tag{2.64}$$

Microscopically, the macroscopic temperature is a product of the individual velocities of the particles. Mathematically, for an ideal gas

$$\frac{3}{2} K_B T = \langle K \rangle = \frac{1}{2} m \langle V^2 \rangle \tag{2.65}$$

where $\langle K \rangle$ is the ensemble average of kinetic energy. From this we can also say

$$\lambda = \frac{\mathbf{V_{new}}}{\mathbf{V}} = \sqrt{\frac{T_{avg}}{T}} \tag{2.66}$$

where T_{avg} is the desired average temperature and T is the instantaneous temperature. The problem with this simple approach is that it does not capture the dynamics properly; by rescaling the velocities at each time step, the kinetic energy would be constant, which is obviously unphysical. Such fluctuations do occur in natural systems - equilibrium is dynamic and properties oscillate around their equilibrium values.

As noted before, the canonical ensemble is equivalent to a closed system in contact with an extremely large heat bath. As such, many thermostats attempt to imitate this. One such technique is the Berendsen thermostat [87]. In this method, the temperature of the simulation is coupled weakly to a heat bath by

$$\frac{dT}{dt} = \frac{1}{\tau}(T_{avg} - T) \tag{2.67}$$

In this way, the velocities are gradually rescaled so as to upset the dynamics as little as possible. The exact rescaling coefficient is

$$\lambda^2 = 1 + \frac{\Delta t}{\tau}(\frac{T_{avg}}{T} - 1) \tag{2.68}$$

An important detail is the term τ, which controls how fast this occurs. Large values of τ will lead to poor temperature control since the correction to the temperature will occur very slowly. On the other hand, at very low values of τ, namely as it tends to Δt, the dynamics will tend to that of the velocity rescaling algorithm.

The Andersen thermostat [88] is a stochastic technique whereby random particles are hit by virtual collisions. After a collision, the momentum of the chosen particle is updated instantaneously by choosing a random momentum from a Boltzmann distribution of momenta at the desired temperature. The advantage of this technique is that it rigorously conforms to the canonical ensemble, unlike the previous two thermostats. The problem with this approach is similar to that of the velocity rescaling - it leads to unphysical dynamics. The updates to the momenta are stochastic and hence the trajectory is no longer deterministic but a Markov chain.

A more sophisticated algorithm is the Nosé-Hoover thermostat [89–91]. It keeps the temperature constant by incorporating the heat bath into the dynamics of the system, i.e. extending the Hamiltonian of the system. The Hamiltonian of a system is a function that gives the energy of that system for any state. For a classical system this is

$$H = K + U = \sum_{i=1}^{N} \frac{\mathbf{p_i}}{2m_i} + U \tag{2.69}$$

i.e., the total energy of the system is the sum of the kinetic and potential energies of

the system. In the Nosé-Hoover interpretation, the Hamiltonian of the system becomes the Hamiltonian of both the system and the heat bath; in this way, the total energy is conserved but the energy of the system is variable. The extended Hamiltonian is [92]

$$H = \sum_{i=1}^{N} \frac{\mathbf{p_i}}{2m_i} + U + \frac{p_s^2}{2Q} + N_f K_B T \ln s \tag{2.70}$$

The new terms are the kinetic and potential energy of the heat bath, respectively, which is at the necessary temperature and can interact with our system. p_s is the momentum associated with the heat bath, whereas Q is a virtual mass used to control the fictitious dynamics. s is a degree of freedom used to define the heat bath and N_f is the number of degrees of freedom, which is 3N + 1, since the presence of s adds an additional degree of freedom. The scaling is then

$$\mathbf{p_i} = m_i \mathbf{v_i} s \tag{2.71}$$

The advantage of the Nosé-Hoover interpretation is that it preserves the dynamics of the system but a disadvantage is that it is not always ergodic. The risk of this is minimised by using a second Nosé-Hoover thermostat to control the first and so forth [39]; with increasing length of the thermostat chain, the accuracy of the overall thermostat is improved. Another important point is that the damping is oscillatory rather than exponential, cf. Berendsen thermostat, i.e. much slower for the same time constant.

One final technique is the Langevin thermostat [86, 93]. It functions by adding a noise term to the equations of motion, which is conceptualised as a bath of much smaller particles constantly undergoing small collisions with the system particles. A second additional term is the friction term, which arises from these collisions. The interaction between the noise and friction terms is used to control the temperature of the system. Due to the random nature of the noise term, this set-up is often called stochastic dynamics. Advantages include ergodicity and being able to use a longer time step. However, momentum transfer is not preserved, meaning it is often not suitable for calculations of dynamic properties, e.g. diffusion coefficients [93].

Barostats

Barostats are used to control the pressure in an MD simulation. They do this by manipulating the size of the system, in much the same fashion as thermostats control velocities, i.e. momenta, to control temperature. In MD simulations, there are three commonly used barostats: 1) The Berendsen barostat [87], which is analogous to the Berendsen thermostat; 2) The Andersen barostat [88], which is similar to the Nosé-Hoover thermostat; 3) The Parrinello-Rahman barostat [94], which is also similar to the Nosé-Hoover thermostat.

The Berendsen barostat weakly couples the system to a "pressure bath", using the relation [39]

$$\frac{dP}{dt} = \frac{P_{avg} - P}{\tau_P} \tag{2.72}$$

where τ_p is the pressure coupling constant. The scaling factor (μ) for the box vectors is then

$$\mu^3 = 1 - \frac{\beta \Delta t}{\tau_P}(P_{avg} - P) \tag{2.73}$$

where β is the isothermal compressibility, however it is not necessary to know this value exactly since it becomes part of the overall term β/τ_P, where τ_P is a specified input. As such, β is often taken as the isothermal compressibility of water [39]. This relation is analogous to **Equation 2.68** for temperature, except now the sign is reversed since too high a pressure leads to a larger cell (to spread the force over a larger area). Conversely, too high a temperature necessitates a reduction in the velocities. Although this algorithm does produce the correct pressure, it does not replicate the NPT ensemble correctly, but rather an approximation of it [39]. For this reason, the Andersen and Parrinello-Rahman thermostats are generally preferred.

The Andersen and Parrinello-Rahman thermostats are analogous to the Nosé-Hoover thermostat, employing an extension of the system Hamiltonian, i.e. adding kinetic and potential terms. These terms relate to a theoretical piston that acts on the system and thereby regulates the pressure. Anisotropic scaling can be implemented for both.

Unlike the Berendsen formulation, they do preserve the NPT ensemble.

2.2.8 Monte Carlo

As noted at the beginning, another common computational technique are Monte Carlo simulations. The underlying difference between MC and MD, generally speaking, is that MC directly samples the ensemble under investigation rather than relying on a time evolution of the system. There is, however, some overlap between the two techniques since ensembles of MD simulations from different initial conditions can be calculated and there also exist so-called kinetic Monte Carlo (KMC) techniques that do follow a time evolution. Although an approach using an evolution of ensembles might seem more rigorous (since we do not have to consider ergodicity), it also has the disadvantage of requiring non-equilibrium statistical mechanics for computing dynamic properties such as viscosity and diffusion constants. Due to much higher computational costs, this is often not feasible. For this reason, MD and KMC are preferred for calculating such properties.

Unlike MD simulations, MC is a stochastic technique; this means that MC simulations rely on random sampling to draw results. Consequently, two MC simulations of the same system with the same initial conditions will not be identical, since the paths in the simulations are random. This is a fundamental difference to MD, where the such simulations would produce identical results since it is a deterministic method - the result only depends on the initial conditions. This is because MC simulations employ a Markov chain, i.e. the future result is independent of past results.

The simplest form of MC is brute-force MC. In this technique, trial moves are made and from the collected results of the trial moves some information can be determined for the system. For example, consider a circle inside a square whose edge length is equal to that of the diameter of the circle, which will be unity for our example. If we were to drop a very large amount of marbles on this arrangement and tabulate what distribution fell within the circle, this would be equal to

$$\frac{N_{circle}}{N_{total}} = \frac{A_{circle}}{A_{square}} = \frac{\pi D_{circle}^2}{4D_{square}^2} = \frac{\pi}{4} \tag{2.74}$$

This is a simple example of how to calculate the value of π using MC. Generically, we are solving

$$I = \int_a^b f(\boldsymbol{x}_i) = \frac{b-a}{N} \sum_{i=1}^{N} f(\boldsymbol{x}_i) \tag{2.75}$$

where b and a are the upper and lower limits of the integral (1 and 0 in our case) and $f(\boldsymbol{x}_i)$ for the Monte Carlo integral is

$$f(\boldsymbol{x}_i) = f(x_i, y_i) = \begin{cases} 1 & \text{if } x^2 + y^2 < 1 \\ 0 & \text{if } x^2 + y^2 \geq 1 \end{cases} \tag{2.76}$$

The problem with applying this to thermodynamic systems is that there are an overwhelming number of possible states available to the system, i.e. there are simply too many states for us to achieve a representative sample by brute force. Fortunately, we know that most possible states of the system will have a vanishingly small probability and hence will have a negligible contribution to the total. The question is therefore, how can we bias our system towards sampling the states of the system that are the most likely?

The solution to this was developed by Metropolis and was formalised in the Metropolis-Hastings algorithm. The key point is that, instead of randomly sampling and *then* weighting the results with the Boltzmann factor, they sampled according to the Boltzmann factor and then weighted evenly [17]. This is not a trivial rearrangement; in this way, likely states, i.e. those with a non-vanishing probability, are much more likely to be sampled than highly unlikely states. This neatly circumvents the problem with uniform sampling, namely that there are far more states that hardly contribute to the thermodynamic average than those that do. This results in a very efficient algorithm for sampling high-dimensional spaces, as chemical systems are.

This then raises the question as to how exactly to generate configurations that are proportional to the Boltzmann constant. We start, using the example from Frenkel [37], from the condition of detailed balance. At equilibrium, the distribution of states is constant and therefore, since equilibrium is dynamic, any configuration at equilibrium should be destroyed and created in equal measure. If this is not true, as the number

of configurations tends to infinity there will be a bias leading away from equilibrium. Mathematically,

$$\nu_i \pi_{ij} = \nu_j \pi_{ji} \tag{2.77}$$

where ν represents a given microstate and π represents a transition probability distribution describing the transition from state i to j. π is essentially comprised of two components, P_{move} and P_{acc}, which describe the probability of an MC move being attempted and accepted, respectively. Hence

$$\pi_{ij} = P_{move,ij} P_{acc,ij} \tag{2.78}$$

In the original Metropolis algorithm, $P_{move,ij}$ was chosen to be a symmetric matrix, i.e. moves are tried in both directions with equal likelihood. This is not a necessary condition for detailed balance, since it is the product of P_{move} and P_{acc} that should be equal in both directions, but we will see that setting one as symmetric allows us to define the other more easily. If $P_{move,ij}$ is symmetric, then we can also state

$$\nu_i P_{acc,ij} = \nu_j P_{acc,ji} \tag{2.79}$$

then we can also state

$$\frac{P_{acc,ij}}{P_{acc,ji}} = \frac{\nu_j}{\nu_i} = \frac{\frac{1}{Z} exp(\frac{-E_j}{K_B T})}{\frac{1}{Z} exp(\frac{-E_i}{K_B T})} = exp(\frac{-(E_j - E_i)}{K_B T}) \tag{2.80}$$

Here one can see the elegance of the Metropolis algorithm; we can eliminate the need to calculate the partition function, Z, because we are not interested in the overall probability of a state but rather their relative probabilities, hence their cancellation here. Since the probability cannot exceed 1, P_{acc} is given by

$$P_{acc,ij} = \begin{cases} exp(\frac{-(E_j - E_i)}{K_B T}) & \text{if } E_i > E_j \\ 1 & \text{if } E_j \geq E_i \end{cases} \tag{2.81}$$

The practical implementation of this algorithm can be summarised as follows

- Select a particle and calculate its energy (E_i)
- Perform a trial move, i.e. enact some change on the system, and calculate the new energy of the particle (E_j)
- Calculate the probability of accepting the trial move using $P_{acc,ij}$
- Generate a random number (R) in the range [0,1]
- If $P_{acc,ij} > R$, then the move is accepted. Otherwise, the system reverts back to its pre-trial-move state.

The term trial move encompasses a multitude of possible processes that are strongly dependent on the system, as well as the thermodynamic ensemble used. Simple moves include translation, rotation and reptation, and more complicated examples would be insertion/deletion of particles (i.e., in the grand canonical ensemble where the composition of the system is not fixed) and reactive moves (such moves are utilised in so-called reactive Monte Carlo (RMC) simulations) [95]. Various trial moves are not necessarily tried with equal frequency; rarer processes (e.g., reactions) are usually attempted much less frequently than common ones (such as translation). The customisability of trial moves is both a blessing and a curse: on one hand, it allows for flexibility when studying a system because they do not have to be physically realisable; on the other hand, this lack of a physical basis for the moves means that MC models are rarely transferable, i.e. they can less often be applied to other similar systems than MD models. Another advantage of the sometimes unphysical path in MC is that it allows the simulation to escape local minima since even unfavourable transitions have a non-zero chance of occurring. For simple schemes this will make the simulation ergodic.

Another important aspect of MC algorithms is that they employ random numbers to ensure that the moves are accepted in proportion to the Boltzmann factor. For this reason, care has to be taken to ensure that the numbers chosen are equally distributed between zero and unity. Computationally speaking, there is no such thing as a perfect random number generator; computers can only carry out a list of instructions and hence cannot act in a random manner. For this reason, random number generators

are frequently referred to as pseudorandom - they are essentially functions which, upon inputting a seed number, produce a large sequence of numbers that conform to a random distribution. However, this sequence will eventually repeat itself and hence the choice of random number generator is an integral part of an MC algorithm, especially since it is not uncommon for millions or billions of trial moves to be attempted in a simulation. If some numbers occur more often than others, this can lead to bias, which will shift the distribution of states away from its true form. This occurs since a bias towards larger numbers will mean physical states are rejected more often than is proper, and vice versa.

An extension of MC is dynamic Monte Carlo (DMC), sometimes also known as kinetic Monte Carlo (KMC). Equilibrium MC can be thought of as a Markov chain with the same distribution as the system at equilibrium. As stated already in our discussion of detailed balance, at equilibrium the rate of change of any one state will be zero. In DMC, this is no longer the case. Instead of **Equation 2.77**, we can write [96]

$$\frac{\partial P_j(t)}{\partial t} = \sum_i P_{ij} P_i(t) - \sum_j P_{ji} P_j(t) \tag{2.82}$$

where P_i and P_j are the probabilities of states i and j. **Equation 2.82** is known as the master equation. Detailed balance in this context is merely the master equation under the constraint of equilibrium. MC, like all simulation methods, takes place in discrete time and hence to link the results to scientific processes it is necessary to derive a correspondence between continuous (real) time and the simulation steps. In DMC, the link are rate constants for the physical process that one wishes to simulate, which can be welded to MC through a Poisson process. For a Poisson process, the likelihood of an integer number of events (N_e) occurring in a given time interval (t) is

$$P_{N_e}(t) = e^{-rt} \frac{(rt)^{N_e}}{N_e!} \tag{2.83}$$

where k is the rate constant. The ensemble of independent forward-reverse processes, which is a precondition, is well-behaved in that it can be treated as a single, large Poission process [96]. Together with detailed balance, DMC can trace a path from a

non-equilibrium state to an equilibrium one. The downside to this technique is that the applicable rate constants for the desired process must be known *a priori* since DMC is not based upon a physical process itself. This could be difficult, for example, in systems with many chemical reactions occurring simultaneously whose rates constants would be difficult to decouple.

In this work, we have not employed MC simulations. This section is intended to provide a cursory overview of the technique so that a reader is better equipped to understand the literature on silica simulations, which will be presented in **Chapter 3**, where MC and DMC simulations are regularly employed. For a more detailed discussion of equilibrium MC, the reader is directed to Frenkel [37]. Regarding a more in-depth discussion of DMC, we recommend the paper of Fichthorn and Weinberg [96].

2.3 Quantum-Mechanical Calculations

For a long time, Newton's formulation of mechanics was considered complete. However, this began to change in the 19^{th} century and an altogether new approach began to supersede it by the beginning of the 20^{th} century. The name of this new discovery was coined as "quantum mechanics" by Max Born in his 1925 paper "*Zur Quantenmechanik*" [97]. This radical new theory was brought about by mounting examples of natural phenomena that were incompatible with Newton's mechanics. Some notable discoveries were

- The dual wave-particle nature of light
- The existence of the electron
- Black-body radiation
- The photoelectric effect

These discoveries gradually revealed the fallacy that was the assumed indivisibility of the atom. Its division into constituent components, namely electrons, protons and neutrons, and then later into further subatomic particles, fundamentally altered our understanding of physics and chemistry.

One of the fundamental postulates of QM, such so that it lent the theory its name, is that the energy of a physical system exists in discrete packets, i.e. is quantised. This idea was so radical that Max Planck initially only introduced the idea, which was first proposed by Ludwig Boltzmann, into his work as a mathematical trick to describe black-body radiation, rather than considering it theoretically rigorous. Indeed, he later described his use of quantum theory as an "act of desperation" (*"Akt der Verzweiflung"*). He did this by assuming that the energy (E) radiated by a photon was proportional to its frequency (ν)

$$E \propto \nu \tag{2.84}$$

the constant of proportionality is known as the Planck constant ($h = 6.626 \times 10^{-34} J.s$). As the field progressed, further relationships were defined and the one most important for QM calculations is the time-dependent Schrödinger equation (TDSE). In the classical approximation, the movement and momentum of a particle are described by Newton's second law, as discussed previously; the TDSE is its quantum analogue.

$$i\hbar \frac{d}{dt}|\Psi(t)\rangle = \hat{H}|\Psi(t)\rangle \tag{2.85}$$

where $\hbar$ is the reduced Planck constant ($\hbar = \frac{h}{2\pi}$), Ψ is the system's state vector and $\hat{H}$ is the Hamiltonian operator. However, the solution of this equation is extremely complex; as such, it is desirable to simplify it. One way to do this is to use the time-independent form, i.e. one that describes a stationary wave instead of one in motion. The time-independent Schrödinger equation (TISE) is hence

$$\hat{H}|\Psi\rangle = E|\Psi\rangle \tag{2.86}$$

where E is the total energy of the system. In this form, the wave function is an eigenfunction of the Hamiltonian operator. The reason that the constraint of time independence can be applied is that molecular orbitals can be described as stationary waves. The wave function is important generally, since it is the mathematical function describing the quantum state, i.e. the probability distribution of any observable of

an isolated system. Although this approach seems very different to that of classical mechanics, they nevertheless share some similarities. For example, they are both based upon the law of the conservation of energy. Indeed, the non-relativistic form of the TISE for a single particle is

$$\left[\frac{-\hbar^2}{2\mu}\nabla^2 + V(\mathbf{r})\right]\Psi(\mathbf{r}) = E\Psi(\mathbf{r}) \tag{2.87}$$

where $\frac{-\hbar^2}{2\mu}\nabla^2$ is the kinetic energy (with μ being the reduced mass of a particle) of the system, whilst $V(\mathbf{r})$ is its potential energy, i.e. similar to the approach used in molecular dynamics.

When we speak of the non-relativistic form, we are referring to Einstein's theories of special and general relativity [98,99]. General relativity describes how two objects with different uniform (i.e., constant) velocities experience space and time differently relative to each other, hence the name. As the difference in velocity between the observer's frame of reference and the object nears the speed of light, this effect becomes increasingly appreciable until Newtonian mechanics can no longer describe the situation effectively. If we can however neglect such effects, then we can describe our system as non-relativistic, as above. A particle can be considered relativistic when its kinetic energy is comparable to its rest energy (E_0), which is given by Einstein's famous equation

$$E_0 = m_0 C^2 \tag{2.88}$$

where m_0 is the particle's rest mass, i.e. that which is invariable and does not depend on the particle's motion and is hence the same in all frames of reference. This implies that the same object will have different masses at different frames of reference. These are hence called relativistic masses since these changes are brought about by relativistic effects. Therefore, relativistic effects only need to be accounted for for particles with extremely small masses or extremely high kinetic energy, i.e. photons. For the domain that we are interested in, namely that of atomic particles and above under standard conditions, these relativistic effects can be ignored. The QM calculations used in this work therefore focus on solving the non-relativistic TISE.

2.3.1 The Hartree-Fock Method

Above, we explained how it is the wave function(s) that describe our system of interest. As such, the focus of QM calculations for many years was on inventing efficient numerical routines for calculating wave functions. One of the first was the Hartree-Fock (HF) method [100–104], which we will now be describe.

If we consider a system of nuclei (comprised of protons and neutrons) and electrons, then we can apply the first important assumption, namely the Born-Oppenheimer [105] approximation. This simply states that the motion of the nuclei and electrons can be treated as mathematically independently of each other. This assumption is reasonable since the mass of a proton or neutron is more than 1,800 times that of an electron; this means that the nuclei, from the point of view of the electrons, will react very slowly in response to changes in the environment, to the point of being practically static. We can thus treat the nuclei as occupying fixed points in space, hence it is not required to calculate their wave functions, only those of the electrons.

The second assumption is less rigorous and is mainly introduced for the sake of computational efficiency - it is assumed that the electrons do not interact with each other. Introducing this assumption means that we can rewrite the Hamiltonian for our system of N electrons as follows:

$$\hat{H} = \sum_{i=1}^{N} H_i \tag{2.89}$$

Splitting up the Hamiltonian in this manner means that the eigenfunctions of each individual TISE are not the wave function of the system but a collection of spin orbitals, i.e. the overall wave function of the whole system is approximated by a combination of spin orbital sets for each electron. Since our electrons are non-interacting and our nuclei are fixed, i.e. are enacting a static field, we can introduce the following for each electron Hamiltonian

$$H_i = \frac{-\hbar^2}{2\mu}\nabla^2 + V(\mathbf{r}) + V_H(\mathbf{r}) \tag{2.90}$$

It is clear that this is very similar to **Equation 2.87** but with the addition of an extra

term, namely the Hartree potential $V_H(\mathbf{r})$. This describes the average effect of the field generated by all electrons on a single electron, i.e. it is a mean field approximation and does not take into account instantaneous interactions between electrons. A corollary of this is that electron repulsion is overestimated, since the electrons do not know to avoid each other. This also means that the electron that the mean field is acting upon also contributes to the mean field, i.e. is effectively counted twice and hence must be corrected for. Another assumption made is that a Slater determinant can describe the relationship between the individual spin orbitals and the overall wave function. The Slater determinant has the following form for a two electron system [106]

$$\Psi(\mathbf{x_1}, \mathbf{x_2}) = \frac{1}{\sqrt{2}} \det \begin{Bmatrix} \chi_j(\mathbf{x_1}) & \chi_j(\mathbf{x_2}) \\ \chi_k(\mathbf{x_1}) & \chi_k(\mathbf{x_2}) \end{Bmatrix} \tag{2.91}$$

This form is preferred to the simpler Hartree product

$$\Psi(\mathbf{x_1}, \mathbf{x_2}) = \chi_{j1}(\mathbf{x_1})\chi_{j2}(\mathbf{x_2}) \tag{2.92}$$

since it satisfies the antisymmetry principle (that switching two electrons reverses the sign of the wave function) and the Pauli exclusion principle (the value of the wave function, i.e. the probability, is zero if two electrons have the same spin and location). The $\frac{1}{\sqrt{2}}$ is a normalisation factor that generalises to $\frac{1}{\sqrt{N}}$ for an N-electron system.

The final information that has to be provided for the calculation is a set of functions that together define each spin orbital, i.e. are of the form

$$\chi_j(\mathbf{x}) = \sum_i^K \alpha_{j,i}\phi_i(\mathbf{x}) \tag{2.93}$$

where χ is a spin orbital, α is a fitting parameter and ψ is a function. This collection of functions and fitting parameters is referred to as a basis set. The larger the basis set, the better the chance that it can describe the spin orbital accurately. However, this also significantly increases the computational workload since it also increases the number of parameters that have to be fitted. This basis set, along with the TISE for each electron, gives us two independent equations that allow us to solve iteratively for

our two unknowns (the mean field and the spin orbitals) as follows:

- 1) Guess the spin orbitals by setting the fitting parameters
- 2) Combine the proposed spin orbitals in order to produce an electron density
- 3) Using the electron density, compute the mean field and then solve for the spin orbitals using the TISE
- 4) If the spin orbitals match those from the first step, then the system has converged and the answer has been found. Otherwise, a new guess for the spin orbitals has to be employed and the process repeats until convergence is found
- 5) The individual Hamiltonians and spin orbitals can be combined to produce the overall wave function and Hamiltonian, and hence the total energy of the system can be calculated from the TISE

To conclude, the Hartree-Fock method manages to capture electron-exchange effects via the Slater determinant (which is an advantage compared to density functional theory (DFT), which will be explained later) but its biggest weakness lies in its neglect of electron-correlation effects. As such, higher level QM methods try to address this by building such interactions into the HF treatment; these will also be discussed later. However, in the next section we will focus on some of the general numerical aspects of solving QM calculations.

2.3.2 Numerical Details of Quantum-Mechanical Calculations

The above sets out how it is theoretically possible to solve the TISE in a numerical fashion, but how is this done in practice? Two details have to be decided upon in order to make this possible:

- A basis set that approximates the one-electron wave functions
- A relation that combines the spin orbitals together to produce the overall wave function of the system of interest

It was stated above that basis sets are of the form given in **Equation 2.93**. Here we will describe which types of functions are commonly used for $\phi_i(\mathbf{x})$. Fundamentally, basis sets are of two types: plane waves and atomic orbitals. These approaches are strongly divergent; plane waves are periodic functions and are hence very good at approximating condensed materials with repeating structures, while atomic orbitals are localised functions and therefore particularly effective for individual molecules [107].

Basis Sets - Plane Waves

The use of plane waves is based on Bloch's theorem [107, 108], which states that for a perfect crystal there exists a set of wave functions that are energy eigenstates and that they take the form

$$\phi(\mathbf{r}) = u(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}} \tag{2.94}$$

These are hence called Bloch waves. $\mathbf{k}$ represents the crystal momentum wave vector of the particle ($p_{crystal} = \hbar\mathbf{k}$). Here $e^{i\mathbf{k}\cdot\mathbf{r}}$ is known as a plane wave and represents the wave function of the particle, and $u(\mathbf{r})$ is the modulation enacted upon the particle by the periodic environment [108]. The presence of plane waves here lends itself to the name of this class of basis sets. Since this is a periodic approach, it necessitates the definition of a supercell, which is given by the lattice vectors $\mathbf{a_x}, \mathbf{a_y}, \mathbf{a_z}$. This means that $u(\mathbf{r})$ has the property

$$u(\mathbf{r}) = u(\mathbf{r} + \mathbf{a_x}n_x + \mathbf{a_y}n_y + \mathbf{a_z}n_z) \tag{2.95}$$

and that it is periodic for any integer (n_i). The best choice for the supercell is the so-called primitive cell; this is the supercell with the least amount of atoms necessary to represent the infinite material by periodicity. The simplest example would be the simple cubic cell; this is given by the lattice vectors

$$u(\mathbf{r}) = u(\mathbf{r} + an_x + an_y + an_z) \tag{2.96}$$

where is a the lattice constant, which in this case is the same in all directions. This describes a cube of edge length a, where an atom lies at every corner. In this way, only one atom is necessary to define the whole cell, since the coordinates of every other atom can be calculated relative to the one atom whose position is defined. Correctly identifying the primitive cell is extremely advantageous since QM calculations often scale unfavourably with the number of atoms.

The reason for the use of plane waves is that it is easier to solve QM problems in terms of $\mathbf{k}$ (reciprocal space) than $\mathbf{r}$ (real space). Due to the presence of $\mathbf{k}$, reciprocal space is also known as k-space. The equivalent supercell vectors ($\mathbf{b_i}$) are defined such that $\mathbf{a_i} \cdot \mathbf{b_i} = 2\pi$ if $i = j$ but equal to zero otherwise. The factor 2π comes from the periodic nature of plane waves. A consequence of this definition is that, as the vectors become longer in real space, their reciprocal counterparts, unsurprisingly, become shorter. The k-space equivalent of the primitive cell is called the Brillouin zone (BZ) and has several important properties. The origin of the BZ is referred to as the Γ-point. For example, it is important for computing integrals of periodic systems that take the following form in reciprocal space

$$\bar{g} = \frac{V_{cell}}{2\pi^3} \int_{BZ} g(\mathbf{k}) d\mathbf{k} = \sum_j w_i g(\mathbf{k_j}) \tag{2.97}$$

where $g(\mathbf{k})$ is the Fourier transform of $\bar{g}$ and w_i is a weighting factor for the numerical version of the integration. The values of $\mathbf{k_j}$ are called k-points and are chosen so as to produce a mesh that appropriately samples the BZ. The most common method for choosing these points in reciprocal space is that of Monkhorst and Pack [109], where the points are uniformly distributed in k-space. A further simplification is the use of irreducible Brillouin zone (BZ); this is a subset of the BZ, being the smallest amount of points needed to define the BZ if symmetry is applied. Therefore, for materials with a high degree of symmetry, the computational load is greatly reduced. Increasing the number of k-points naturally leads to better convergence since a finer mesh is being used for the integration. Another way that the convergence of plane-wave calculations can be trivially improved is by using energy cut-offs. The Fourier constants of the plane

wave functions decrease between phases meaning that they can be easily truncated by removing all terms above a certain kinetic energy according to

$$\frac{\hbar}{2m}|\mathbf{k}+\mathbf{G}|^2 < E_{cut} \tag{2.98}$$

E_{cut} introduces error into the calculations however this can be negated by increasing its value, although this requires extra computational expense. However, this intuitive way of economising precision and computational tractability is a useful characteristic of plane-wave basis sets.

In many cases, the core electrons of an atom have only a very minor effect on the property that we wish to analyse. This is because they are bound much more strongly to the nucleus and are therefore less free to participate in chemical processes. For this reason, it is often acceptable to neglect a full calculation of the core electrons, the so-called "frozen-core approximation" [107, 110], and only compute this for the valence electrons. This drastically increases the speed of calculations for a minor decrease in chemical resolution. The core electrons are hence represented in these calculations by a pseudopotential, whereas calculations that do not employ these are termed all-electron calculations.

Basis Sets - Atomic Orbitals

For atomic orbitals there is no periodicity, meaning that there is no need to define a supercell, and hence they decay to zero far away from the atoms. Although they can be technically centred anywhere, they are usually placed at the centre of an atom. One of the first types of molecular orbital basis functions were so-called Slater-type orbitals (STOs) [111]. They have the form

$$\phi_{STO}(x,y,z) = Nx^a y^b z^c e^{-\zeta r} \tag{2.99}$$

where N is a normalisation constant and a, b and c relate to the angular momentum such that their sum is equal to it. r is the distance from the atomic nucleus and ζ controls the width of the function. Large values of ζ produce sharply decaying functions, whereas

low values result in diffuse functions. The main motivations for this type of basis function are twofold: at long range, it exponentially decays and at short range it has a cusp, meaning that it satisfies the Kato theorem [112]. In this way, orbital behaviour at both extremes is treated correctly. However, despite their physical accuracy, their computational tractability is poor due to it being difficult to integrate them. As a compromise, Gaussian-type orbitals (GTOs, also known as Gaussian primitives) are commonly used. These have the form

$$\phi_{GTO}(x, y, z) = Nx^a y^b z^c e^{-\zeta r^2} \tag{2.100}$$

The disadvantage of this formulation is that they lose the cusp at short range (i.e., near the nucleus). The principal advantage arises from the Gaussian product theorem, which states that Gaussians situated on different atoms can be reformulated as a sum of several Gaussian functions at some point on the axis located between them. In this way, many-body interactions can be reduced to a sum of Gaussians, dramatically increasing the speed of integration. In order to combine the advantages of both types of orbitals, Boys came up with the idea of using a linear combination of GTOs to approximate STOs [113], such that

$$\phi_{STO-nG}(x, y, z) = N \sum_{i=1}^{n} k_n x^a y^b z^c e^{-\zeta r^2} \tag{2.101}$$

This, of course, significantly increases the number of fitting parameters in the basis set, however it is more than justified by the increase in speed compared to STOs and in accuracy with respect to GTOs. This is called a contracted Gaussian-type orbital (CGTO), often referred to as an STO-nG orbital (where n refers to the number of Gaussians used to approximate the STO). Combinations of the above types of orbitals are then assembled to represent each atomic orbital (AO).

An AO that is described by only one of the above functions (STO, GTO or CGTO) is referred to as a minimal basis set, i.e. it is described by the minimum number of functions, namely one. Larger sets are referred to as double-ζ, triple-ζ (alternatively as DZ, TZ), etc. for AO basis sets containing two and three functions, respectively.

A larger basis set naturally produces a better fit at the cost of more computational expense, since there are more parameters that have to be iterated over.

A further subdivision is for so-called split-valence basis sets. In this case, core AOs use only one basis function, whereas outer AOs use larger sets. They generally use the notation of Pople's basis sets, e.g. n-ijG or n-ijkG [114]. n states the number of Gaussians used to approximate the core shell, whereas i and j represent the number of Gaussians used to describe the first and second functions of the basis set for the outer electrons. In this way, 6-31G uses six Gaussians for the core AOs and a double-ζ basis set for each outer AO, which contains three Gaussians in the first and one in the second basis function. Similarly, 6-311G would be a triple-ζ basis set.

To represent distortions of orbitals by polarisation, additional functions have to be added. In Pople's terminology, this is noted with either one or two asterisks; one asterisk declares that only the heavy (i.e., non-hydrogen or -helium) atoms include polarisation functions, whereas two means that they are present for all atoms. For example, 6-31G* or 6-31G**. The polarisation functions are d-type orbitals for heavy atoms and p-type otherwise. Therefore, 6-31G** is also sometimes written as 6-31G(d,p). In general, an AO of angular momentum number L is polarised with an AO function of L+1. Polarisation corrections are important wherever electron correlation plays a role, since they add the necessary asymmetry to the orbitals. However, they are costly [115]. Another commonly used additional feature are diffuse, i.e. very slowly decaying, functions. These are useful for describing weak long-range (e.g., hydrogen bonding) interactions [115]. An example of the notation for Pople basis sets is 6-31+G and 6-31++G. + states that s and p functions have been added to heavy atoms, whereas ++ shows that an s function has also been added to the light atoms.

The main deficiency with the Pople sets is that they were not parametrised against calculations with correlation effects [115]. As such, the state-of-the-art for basis sets are so-called "correlation-consistent" basis sets, as proposed by Dunning [116]. The exponents of these Gaussian basis sets are parametrised against such calculations, hence the name. They do this by including successively larger polarisation functions that add extra asymmetry and can therefore capture correlation effects. Another advantage is

that they use extrapolation techniques to approach the basis-set limit (BSL) [117]. The BSL is the result that would be obtained if an infinitely large basis set was used. At this point, a calculation is not computationally limited, since an infinitely large basis set would be able to perfectly describe the wave function, but dependent instead on the level of theory. For this reason, correlation-consistent sets are designed to quickly and monotonically approach the BSL.

For first- and second-row atoms, the notation for correlation-consistent basis sets is cc-pVnZ, where n represents the number of basis functions in the same way as in Pople's sets. Traditionally, n is written as S (short for "single"), D ("double"), T ("triple") and Q ("quadruple") for n = 1,2,3 and 4, respectively, whereafter they are only referred to by integers. They can also include diffuse functions and are then defined as aug-cc-pVnZ. The V in the nomenclature stands for valence since they were intended to only be for valence electrons, however core-correlated versions have also been produced and they are listed as cc-pCVnZ. Finally, for post second-row atoms additional functions are necessary, using the form cc-pV(n+d)Z. For DFT calculations, which will be discussed later, an alternative choice of extrapolated basis sets is available, namely the polarisation-consistent basis sets of Jensen [118]. The advantage of these basis sets is that they converge much more quickly to the BSL, i.e. less basis functions are necessary, meaning that the same accuracy can be achieved at less computational expense. Although they were designed for HF and DFT methods, they have also been shown to retain their advantages at explicitly correlated levels of theory [119], which will be discussed in the next section.

Finally, one important basis-set phenomenon is basis-set superposition error (BSSE). BSSE is important where the system we are interested in is a function of its geometry, e.g the adsorption interaction energy as a function of distance. When molecules, or even parts of the same molecule, come close together, an electron can use the basis set functions of other electrons, leading to a decrease in energy. Consider a complex of molecules A and B, where the interaction energy is represented as a function of geometry by [120]

$$\Delta E_{complex}(\mathbf{r}) = E_{AB}(\mathbf{r}, \hat{AB}) - E_A(\hat{A}) - E_B(\hat{B}) \tag{2.102}$$

where E_A and E_B represent the energies of molecules A and B in isolation, and $E_{AB}(\mathbf{r})$ is the energy of the system where both A and B are present and separated by $\mathbf{r}$. $\hat{AB}$, $\hat{A}$ and $\hat{B}$ refer to the combined basis-set functions of A and B together, as well as the basis-set functions of A and B in isolation, respectively. When $\mathbf{r}$ nears infinity, the result will tend to zero since there is no complex and the energy of the system will be equal to the sum of E_A and E_B. As the two molecules are drawn together, the energy will decrease until a minimum is achieved at the equilibrium distance between the fragments in the complex. However, at short range there is also mixing of the basis functions, which is not present in the isolated fragments. Since we are comparing against these, which do not have this effect, our result is therefore not purely a function of geometry, as we are trying to obtain. As such, BSSE is not necessarily unphysical, but rather muddies the process of trying to find the effect of one variable. As such, it is often necessary to subtract the BSSE. It is particularly important for calculations of binding energies and complexes [121].

There are two main methods for correcting BSSE, namely the chemical Hamiltonian approach (CHA) and the counterpoise method (CP). The counterpoise approach rewrites **Equation 2.102** as

$$\Delta E_{CP,complex}(\mathbf{r}) = E_{AB}(\mathbf{r}, \hat{AB}) - E_A(\hat{AB}) - E_B(\hat{AB}) \tag{2.103}$$

This means that the "isolated" part of the calculations is computed where both sets of basis sets are present, but where only one set of nuclei and electrons are. The basis functions that are present in the absence of its nuclei and electrons are called "ghost orbitals". Since in all configurations the superposition of basis sets is possible, the problem is effectively reduced to the calculation of one target variable, i.e. its spatial geometry. Some authors have argued that this approach overcorrects since the absence of the ghost orbitals' electrons means that more basis functions are free to stabilise the non-ghost monomer, whereas other authors attribute this to other errors, e.g. basis set incompleteness errors (BSIE) [120].

2.3.3 Post-Hartree-Fock Methods

Although HF is exact within the assumptions that it makes, these assumptions hinder it from being able to make predictions compatible with experiment. The assumption within the HF framework that has the greatest effect on its chemical accuracy is the lack of electron correlation. As stated above, HF is a mean field approach; the electrons only feel the average effect of their counterparts and, bizarrely, since they also contribute to the mean field, act upon themselves. As such, in keeping with the assumption that relativistic effects are negligible and the Born-Oppenheimer approximation is valid, we can write the following expression for the correlation energy

$$E_{corr} = E_0 - E_{HF} \tag{2.104}$$

where E_0 is the ground state energy and E_{HF} is the energy from HF. Since E_{HF} must be greater than E_0 (by definition, a molecule cannot have a energy lower than its true ground state), E_{corr} is of negative sign. As such, most approaches try to calculate E_{corr} to complement E_{HF}, which can be calculated reliably. For this reason, approaches of this type carry the moniker *post-Hartree-Fock methods*. In this section, we will briefly describe the three most important of these methods, namely configurational interaction (CI) [122], coupled cluster (CC) [123] and Møller-Plesset perturbation theory (MP) [124].

The CI method is the simplest of the above methods and the crux of the technique is the reformulation of the ground state wave function as

$$\Psi = \sum_{I=0} c_I \Psi_I \tag{2.105}$$

where Ψ_I is known as a configurational state function and describes a particular configuration, i.e. defines the quantum numbers of electrons. c_I are coefficients that are solved for iteratively in order to minimise the ground state energy. Ψ_0 corresponds to the HF Slater determinant, i.e. the HF wave function. The integer I describes the number of electrons that have moved from the HF determinant to the so-called virtual orbitals. $I = 1$ corresponds to one electron having made the transition, $I = 2$ to two,

and so forth. For a system of N electrons, when $I = N$ the CI calculation should be exact. However, this would be computationally infeasible due to numerical problems arising from the enormous basis sets required [125].

Further problems are size consistency and size extensivity. For a full configuration interaction (FCI) method (where $I = N$), the solution would be exact. However, since this is in practice computationally infeasible, the expansion is contracted, normally to doubles ($I = 2$). This leads to the situation where, for two molecular fragments, A and B

$$E(A|B) \neq E(A) + E(B) \tag{2.106}$$

where $E(A|B)$ is the energy for an interacting pair A and B when A and B are separated at an effectively infinite distance within in the same system. $E(A)$ and $E(B)$ are the energies of A and B in separate systems, respectively. In other words, the two systems are not size consistent for truncated CI methods. Similarly, size extensivity deals with how the energy scales of a non-interacting system should scale linearly with the number of molecules [126]. CI is neither size consistent nor size extensive. Next, we will discuss coupled cluster (CC) techniques, which do demonstrate size extensivity.

The CC formulation also rewrites the wave function, but this time as [126, 127]

$$\Psi = e^{\hat{T}}\Psi_{sl} \tag{2.107}$$

where Ψ_{sl} is a Slater determinant and $\hat{T}$ is the excitation operator, which describes excitations of the electrons caused by electron correlation. $\hat{T}$ is comprised of different types of many-body excitations. For example, up to triples, it would be

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 \tag{2.108}$$

where each term describes the excitations arising from that interaction. The advantage of the exponential ansatz is that it maintains size extensivity; the multiplication of two wave functions simply leads to the operators adding together within the exponential term [127]. Size consistency, however, depends on the reference wave function. Due to

this advantage, it is currently the gold standard of quantum-chemistry calculations [127]. However, a disadvantage of CC is that, unlike CI, it does not obey the variational principle, i.e. energies lower than the ground state energy are technically possible [127].

Finally, we have Møller-Plesset perturbation theory (MP). This uses perturbation theory, whereby, starting from the exact solution of a simpler problem (Hartree-Fock, in our case), the system is perturbed slightly in order to search for the true solution. This works because HF already recovers most of the energy of the system and the electron-correlation component is relatively small in terms of the whole system, despite being large on the scale of chemical processes. MP theory can be formulated as

$$\hat{H} = \hat{H}_0 + \lambda \hat{V} \tag{2.109}$$

where $\hat{H}$ is the real Hamiltonian, $\hat{H}_0$ is the HF Hamiltonian, $\hat{V}$ is the perturbation and λ is a parameter controlling the strength of the perturbation. The energy is then given as a power series by

$$E = \sum_{n}^{i=0} \lambda^i E_i \tag{2.110}$$

The level of theory is given by n; for $n = 1$, the energy will be the same as in HF. Due to the expense of the procedure, $n = 2$ is the most common choice, although higher-order variants are also possible. Like CC, MP calculations are size extensive but non-variational. Due to these similarities with and being more tractable than CC, it is often used as a cheaper variant thereof.

2.3.4 Density Functional Theory

Although the post-HF methods mentioned above are highly accurate techniques, they are slow and often scale poorly. For this reason, the currently dominant method in computational chemistry is density functional theory (DFT). DFT was developed in the 1960s by Kohn, Hohenberg and Sham [128, 129] and the key point is that it uses the electron density as a proxy for the wave function. The benefit of this is that it replaces a function of $3N$ dimensions by only three spatial dimensions, thereby mak-

ing the computation quicker. Formally, this is stated as "*the ground state energy from Schrödinger's equation is a unique functional of the electron density*" [107] and is referred to as the first Hohenberg-Kohn theorem. Since there cannot be an energy for a molecule lower than its ground state energy, minimising an appropriate functional will result in the ground state energy. This is known as the second Hohenberg-Kohn theorem, i.e. the electron density minimising the overall energy is the true electron density. While Hohenberg and Kohn proved that such a functional exists, they did not provide any clues as to its form. Accordingly, a plethora of functionals has been suggested. The form of the TISE for DFT is similar to that of HF, but without the summations over all electrons since these have been replaced by a density. Formally,

$$H\psi = [\frac{-\hbar^2}{2\mu}\nabla^2 + V(\mathbf{r}) + V_H(\mathbf{r}) + V_{XC}(\mathbf{r})]\psi = E\psi \tag{2.111}$$

where V_{XC} is the so-called exchange-correlation energy functional and is a catch-all expression for all the unknown QM contributions. These are principally electron exchange and correlation, hence the name. It is this that we mean when we speak of a DFT functional. Like HF, DFT is an iterative procedure, but in this case the guesses for spin orbitals are replaced with estimates for the electron density, i.e. HF is a mean field approach, whereas DFT is, of course, density functional theory. It is important to stress that DFT is a general technique, rather than a specifically quantum-chemical approach. In this way, we have a form of the TISE and a procedure by which to solve it; the only missing piece is now V_{XC}.

As such, the art to an accurate DFT calculation is the application or development of an appropriate functional, where the development thereof is a far from trivial task. In this section, we will touch upon four subcategories of DFT functionals. The simplest is the local-density approximation (LDA) [107,128] series of functionals, which take the form

$$E_{XC}(n) = \int V_{XC}(n)\, n(\mathbf{r})\, d^3r \tag{2.112}$$

In this formulation, a common and successful approach is to use the one exchange-

correlation functional whose form is known exactly, namely that of the uniform electron gas (EG). As would be expected for this case, the electron density ($n(\mathbf{r})$) is constant at all points in space (i.e., uniform). Due to the simplicity of this formulation, it is possible to calculate V_{XC} exactly for a given electron density. The formula then becomes

$$E_{XC}(n) = \int V_{XC,EG}(n)\, n(\mathbf{r})\, d^3r \tag{2.113}$$

i.e., the value of the exchange-correlation functional is a function of only the local electron density, hence the term *local-density approximation*. The advantage of this approach is naturally its simplicity, and it is most appropriate for systems where the electron density varies gradually throughout the system, so that there are no large differences between neighbouring locations. As such, the LDA is most appropriate for solids, where the structure is repetitive, i.e. doesn't decay to zero at any point and the electron density changes smoothly in space. For the same reason, it is less successful when applied to molecules or clusters. Consequently, there also exists another class of functionals under the banner *generalised gradient approximation* (GGA). GGA functionals are an extension of the LDA in that they also incorporate the gradient of the electron density at a point [107, 130]. This leads to the formalism

$$E_{XC}(n, \nabla n) = \int V_{XC}(n, \nabla n)\, n(\mathbf{r})\, d^3r \tag{2.114}$$

Two common variants of this type are the Perdew-Burke-Ernzerhof (PBE) [131] and Perdew-Wang (PW91) [132] functionals. Additionally, there also exist meta-GGA functionals [107], which go one step further and include the second derivative of n.

$$E_{XC}(n, \nabla n, \nabla^2 n) = \int V_{XC}(n, \nabla n, \nabla^2 n)\, n(\mathbf{r})\, d^3r \tag{2.115}$$

The final type of functionals is the hyper-GGA (also known as hybrid-GGA) class [107]. Functionals of this breed combine a GGA functional with a calculation of a part of the exact exchange energy. The most famous example of this class is the Becke, three-parameter, Lee-Yang-Parr (B3LYP) functional [133, 134]. The Becke part represents the exchange part (V_X), while the Lee-Yang-Parr component describes correlation (V_C).

The three parameters referenced in the name are three tunable parameters (a_0, a_x, a_c) that are optimised so that the functional produces accurate results for a large test set of molecules. The functional form is

$$E_{XC}(n, \nabla n) = E_{x,LDA} + a_0(E_{x,HF} - E_{x,LDA}) + a_x(E_{x,GGA} - E_{x,LDA}) + E_{c,LDA} + a_c(E_{c,GGA} - E_{c,LDA}) \quad (2.116)$$

2.3.5 *Ab Initio* Molecular Dynamics

As mentioned previously in our discussion of phase space, the Hamiltonian for classical systems is comprised of potential ($3N$ degrees of freedom) and kinetic energy (also $3N$) components

$$H = K + U = \sum_{i=1}^{3N} \frac{1}{2} m_i v_i^2 + \sum_{i \neq j}^{3N} U_{ij} \quad (2.117)$$

We have also stated how K is easy to compute since it is a single-body quadratic function but that the same is not true for the many-body potential term. Classical potential functions tend to only include dispersion, repulsion and electrostatic terms and are consequently rather simplified. One way to circumvent this is to use so-called *ab initio* molecular dynamics (AIMD) to describe the potential term. There are two main strands of AIMD: Car-Parrinello molecular dynamics (CPMD) [135] and Born-Oppenheimer molecular dynamics (BOMD) [107].

The latter, BOMD, is theoretically simpler. At every frame in the trajectory, the potential energy is calculated using a standard DFT calculation. Upon obtaining the potential energy of the system, the rest of the procedure is identical to standard MD simulations; from the energy, forces are calculated and then the positions of the nuclei are updated.

CPMD is an approximation of BOMD and was initially favoured because it was more computationally efficient than BOMD. However, due to new computational methods making BOMD quicker, BOMD is now more widely used [107]. Car and Parrinello used

a Lagrangian formulation for CPMD, i.e.

$$L = K - U \tag{2.118}$$

which leads to [107]

$$L = \frac{1}{2}[\sum_{i=1}^{3N} m_i v_i^2 + \sum_{i=1}^{j} 2\mu_E \int d\mathbf{r}|\dot{\psi}_j(\mathbf{r})|^2] - E[\psi_j] + L_{ortho} \tag{2.119}$$

where μ_E is a tunable parameter representing the electron mass and the L_{ortho} term keeps the one-electron wave functions orthogonal. The key detail here is that Car and Parrinello gave electrons a fictitious mass and treated them as classical particles, thereby integrating the electronic degrees of freedom into the nuclear ones. For this purpose, the electrons are given a much heavier mass (μ_E); the default value in the CPMD software is 400 a.u., i.e. 400 times that of the actual mass of an electron. The choice of the value of μ_E must be made judiciously; smaller values make the system more adiabatic since we are closer to the Born-Oppenheimer approximation. However, this will also mean that smaller MD step sizes have to be used (the maximum time step is proportional to $\sqrt{\mu_E}$ [136]), slowing down an already computationally intense technique.

2.3.6 Summary of Quantum-Mechanical Methods

DFT, due to favourable scaling derived from its density-based formulation, is the most widely used framework for QM calculations. Its tractability for systems numbering thousands of atoms has made it the workhorse of modern computational chemistry, since this makes it suitable for a much more diverse range of systems. CC and CI methods are of a similar level of theory and represent the highest level of theoretical complexity generally available for QM calculations. Nevertheless, CC is considered the gold standard due to it possessing size-extensivity; this is regarded as more important than conforming to the variational principle, as is the case for the CI formulation. MP methods are similar to CC and can be thought of as a compromise between DFT and CC; they are less computationally intensive than CC and generally more accurate than DFT.

2.3.7 Quantum-Mechanical Calculations for Force-Field Development

In this work, we will not be studying our chosen system, the synthesis of mesoporous silica, which we shall cover in **Chapter 3**, via QM calculations due to the fact that this would be computationally burdensome to the point of impossibility. Rather, we will be using them to obtain data against which we can parametrise classical potential functions, i.e. a force field, for molecules involved in this system. This section will therefore provide the reader with a brief overview of how QM data can be used to derive force field parameters.

Bonded Parameters

For the simpler bonded parameters, i.e. bond-stretching and angle-bending, QM calculations are useful because they can provide us with geometric data about the molecule. As mentioned previously, the geometry of the lowest energy configuration corresponds to the ground state, therefore optimised molecular structures from QM can provide us with equilibrium values of the bond and angle terms. Additionally, QM can be used to fit the spring constants for bond stretching and angle bending. This can be done in one of three ways: 1) fitting to vibrational spectra; 2) calculating the Hessian matrix (second derivatives of the potential energy surface) of the equilibrium structure of the molecule; 3) scanning the potential surface directly over a range of small deviations away from the equilibrium bond length or angle [137]. In the second protocol, the Hessian allows the potential energy surface to be approximated and therefore the energy penalties for small perturbations away from the equilibrium values to be computed [137, 138]. The energy of the molecule can then be plotted against deviations in the bond length and angle, from which the constants for the chosen potential functions can be fitted.

The torsional parameters are usually fitted to data obtained from the third strategy listed above. Since the dihedral angle can move far from its equilibrium value, perturbation of the Hessian is not appropriate. Usually, a complete 360° scan is carried out [137, 139]. Due to the large changes in the relative displacements of different parts of the molecule caused by twisting through a whole revolution, the non-bonded interactions between different parts of the molecule will also vary strongly. These inter-

molecular forces must be subtracted from the QM energy before fitting, so as to prevent them being counted twice in the simulation [139]. Technically, this should also be done for bonds and angles, however the variations involved in the bond lengths and angles are insignificant compared to those stemming from the dihedral angles. A corollary of this is that the non-bonded parameters must be parametrised before fitting those for dihedral torsion. QM calculations are normally the only available data for conformational energies and are therefore an important part of force field parametrisation, especially for larger molecules (e.g., biomolecules) [139]. When fitting bonded parameters from the QM data, it is important to weight the fit more strongly around the minimum because these are the values of the potential that are most likely to occur in an MD simulation and are therefore the most important to describe correctly.

Non-Bonded Parameters

As discussed previously, the non-bonded parameters are usually implemented in the form of the Coulombic, i.e. electrostatic interactions, and the LJ, i.e. dispersion and repulsion, potentials. The latter of these two potentials is problematic, were we to want to parametrise a force field entirely via DFT calculations, the most common QM method, because dispersion is described very poorly by standard density functionals [139, 140]. This means that standard DFT data cannot be used for fitting the LJ potential. Dispersion corrections have, however, been included in DFT calculations in recent years [140] to account for this flaw and in some cases have been used to fit the LJ parameters. For example Grimme [141] used dispersion-corrected DFT to calculate whole force fields using only QM data. However, these dispersion corrections may not be totally accurate, may be impractically expensive for larger molecules and decomposing the dispersion potential into pairwise interactions can be far from trivial [138]. Visscher and Geerke also parametrised a complete force field via QM data [138], but they fitted their dispersion coefficients against the electron density partitioned among the constituent atoms of a molecule. A different approach that has also looked to parametrise the LJ potential from QM calculations is that of Cole et al. [142]; they scaled up the LJ parameters of free atoms calculated from time-dependent DFT according to their relative size in a

given molecule. The latter two approaches are also attractive because they allow the coefficients to change in response to their environment. However, such approaches require much more expertise and technical knowledge than simply fitting the LJ parameters to experimental data. For all the above reasons, fitting the LJ parameters to experimental data remains the standard approach [138, 142].

Finally, point charges can be determined for the Coulomb potential from QM data in several ways. Two common approaches are to use an appropriate piece of theory to fit point charges to the electron density or the electrostatic potential, both of which are available from QM calculations. Such fitting procedures will be one of the topics dealt with in this work and therefore these approaches will instead be discussed in detail in **Chapter 5**. Using charges from QM calculations is often necessary due to the fact that point charges are less transferable than LJ parameters. Since, as was mentioned above, it is normally necessary to fit the LJ parameters empirically to experimental data, fitting the charges in an empirical manner as well would require many experimental data points. This is problematic in the case of less well-studied molecules, for which the amount of experimental data is very limited.

Condensed-Phase Parameters

When we talk of a QM calculation on a single molecule, it is often understood that this is carried out in vacuum, i.e. in the absence of interactions with other chemical species. However, in many cases, it is necessary to parametrise a force field for condensed-phase molecules. In condensed phases, interactions with other molecules may significantly change the properties of individual ones. For example, the geometry of the molecule will change. This is especially important for conformational energies in very large molecules, where interactions with the solvent may make the equilibrium conformation, i.e. dihedral angles, in the liquid phase different to that in vacuum, i.e. in a single-molecule QM calculation. The Coulomb potential is another part of the parametrisation that will be strongly affected, at least in non-polarisable force fields. Since liquid-phase polarisation is accounted for approximately in such force fields by inflating the values of the point charges, liquid-phase charges for polar molecules fitted to vacuum QM data

will be far too low and will fail to accurately represent the electrostatic potential of the molecule.

There are two strategies that can help resolve this issue: 1) optimising the whole liquid structure using QM; 2) employing a continuum approximation of the condensed-phase environment. The first approach is rarely used due to the enormous computational expense that would be required to optimise a quantity of molecules large enough to accurately reproduce the liquid environment. The second approach is more commonly employed due to its superior computational tractability. However, due to being an implicit continuum, it cannot reproduce explicit local interactions that may be crucial for describing some properties. In this work, we will be particularly interested in liquid-phase charges for polar molecules, therefore continuum solvent models will also be an important part of this work. For this reason, they will also be discussed in depth in **Chapter 5**.

2.4 Conclusions

In this section we have introduced many theories, simulation techniques and calculation methods that are necessary to understand the results contained in this work. We have acquainted the reader with both classical and quantum theories of matter, since both are employed in this work. On the classical side, we focussed principally on molecular dynamics simulations, which will be the workhorse for our molecular-level simulations, and statistical mechanics. In our discussion of QM, we invited the reader to pay special attention to density functional theory. In the following section, we will demonstrate how the aforementioned techniques have been applied in the literature to model and further our understanding of the synthesis of mesoporous silica materials.

Chapter 3

Literature Review

3.1 Periodic Mesoporous Silica

Periodic mesoporous silica (PMS) materials consist of a three-dimensional amorphous silica network; they exhibit short- but not long-range order. As the name suggests, they have large pores; these are typically 2—50 nm in size. Their amorphous wall structure is in contrast to the regular, crystalline structure of the chemically similar zeolites, although both have the overall chemical formula SiO_2. The discovery of the first PMS material is generally credited to employees of Mobil [5]. The company therefore lent its name to the most famous example of these materials, MCM-41 (MCM = Mobil Composition of Matter), a material characterised by a hexagonal arrangement of cylindrical pores. Interest in the scientific community in this new class of materials was generated by the work of Beck et al. [5,6], which then led to a variety of discoveries that will be touched upon in this section, namely:

- The extension of syntheses of PMS materials to a wide range of pH conditions [143]
- The discovery of block-copolymer-templated materials, epitomised by SBA-15 (SBA = Santa Barbara Amorphous) [8, 15, 144, 145] in much the same way as MCM-41 typifies surfactant-templated PMS materials
- The incorporation of functionalisation into the synthesis process [9–11]

The widespread interest in this class of materials derives itself from the variety of

mesostructures that will form under the influence of different synthesis-solution conditions. The ability to functionalise these materials has led to many current and proposed applications, such as in microelectronics, biomedicine, separations and catalysis [8, 35, 146, 147]. In order to realise many of these applications, it will be necessary to tune these materials to very specific tasks; this will in turn necessitate an excellent understanding of all the simultaneously competing chemical and physical phenomena occurring during the synthesis. Solvation and hydrophobic effects, acid-base and phase equilibria, and chemical reactions [36, 148, 149] can all significantly influence the morphology of the synthesised material. To this end, it is important to understand how the ordered structural, physical and chemical properties of such materials arise from the initially disordered synthesis conditions.

Due to the wide range of length and time scales on which this cocktail of chemical and physical phenomena is taking place, it is extremely challenging to probe the whole process *in situ* using any one or even a combination of experimental techniques. As a consequence, the exact synthesis mechanism is still unknown and debated. However, due to increases in computing power, it is becoming increasingly possible to simulate such systems *in silico*. The advantage of this approach is that it allows the systems to be studied with atomic, or near-atomic, resolution. However, accurately modelling such complex systems brings other challenges: on the one hand, it is necessary to represent the key chemical behaviour of the system accurately; on the other, models are desired that provide transferability between similar systems and are computationally tractable. Providing this balance is hence one of the main challenges of modelling work. Furthermore, the discovery and design of new nanomaterials in the laboratory is a largely heuristic process, meaning it is generally expensive with regards to time and financial resources. Developing accurate computational models would allow potential candidate materials to be designed and tested for specific applications without stepping foot in a laboratory, i.e. to be identified and characterised in a more cost-effective manner.

Generally speaking, there are four components in the self-assembly of mesoporous silica materials. These are a structure-directing agent (SDA), a silica source, a solvent

and a catalyst (normally H^+ or OH^- ions, i.e. acidic or basic conditions) [150]. The templating agent promotes the formation of the desired morphology: in the case of MCM-41, the SDA is usually an ionic surfactant; for the synthesis of SBA-15, a block copolymer is used. The silica source polymerises around this template, creating a negative image thereof in solid form, thereby giving rise to the material. In the following sections, we will start with a more detailed discussion of the roles of each of the synthesis components. First of all, we will discuss the theory behind the self-assembly of SDAs. Subsequently, we will look at the the silica source and the materials arising from its polymerisation. Finally, we will carry out a detailed review of theoretical attempts to probe the synthesis mechanism of PMS materials.

3.2 Synthesis of PMS Materials

3.2.1 Self-Assembly and Structure-Directing Agents

The science behind the synthesis of PMS materials is an extremely complex process, whose mechanism is still poorly understood in many cases [15, 35, 151], whereas the experimental synthesis is in itself relatively trivial. This is because these materials are synthesised in part via a *self-assembly* mechanism. This means that, under appropriate conditions, these ordered materials will spontaneously arise from a disordered precursor solution. These processes are carried out in solution so that the molecules are mobile, thus allowing assembly to occur. The concentration of the solution must be sufficiently high so that the structure will be formed relatively quickly. The driving force for self-assembly syntheses are in the vast majority of cases weak non-covalent interactions, e.g. Van der Waals, electrostatics, hydrophobic interactions, hydrogen bonds [152–154]. Another characteristic of self-assembly is *reversibility*. The necessity of this attribute is so that components may rearrange themselves, i.e. molecules will not become trapped in disordered configurations. As a result, crystals are synthesised by reversible aggregation, i.e. self-assembly, whereas for glasses it is an irreversible process [152, 154]. This reversibility also tends to result in structures with a low degree of defects. One key condition for this is that the ordering and disordering forces must be of a similar

magnitude. The disordering force is usually the thermal motion of the molecules and therefore weak interactions are usually the ordering force [152].

Since self-assembly is by definition a spontaneous process, it is accompanied by a drop in the appropriate free energy of the system, which, since this is usually carried in solution and does not rely on an external force, is the Gibbs free energy in this case. This can be written generally as [155]

$$\Delta G_{SA} = N_{SDA}\mu_{agg} - kT(N \ln \phi_{SDA} - \ln \phi_{solv}) \tag{3.1}$$

where ΔG_{SA} is the free energy of self-assembly, N_{SDA} is the number of structure-directing agents (SDA), i.e. surfactants, in the micelle, μ_{agg} is the interaction energy for forming a micelle per SDA, and ϕ_{SDA} and ϕ_{solv} are the volume fractions of SDA and solvent, respectively. The first term on the right-hand side represents the attractive interaction of the aggregate, whereas the second term describes the entropic contribution to the free energy. Since a system undergoing self-assembly is becoming by definition more ordered, the entropic term will be negative. Therefore the interaction term must outweigh this for self-assembly to take place.

The SDAs are usually amphiphilic molecules, i.e. one part of the molecule is hydrophobic and the other hydrophilic. In water, the hydrophobic tails of the molecules will group together in the core of an aggregate, while the hydrophilic head groups will arrange themselves at the surface. There are four principal interactions that are applicable to this type of system: head-solvent electrostatic interactions, repulsion between head groups, solvent-tail repulsion and the conformational energy of the tail [154, 155]. The first two apply primarily to the head groups, whereas the latter relate to the tails; this relative separation of interactions is useful for tuning the structures formed from these types of molecules. As such, not all interactions are as influential for all SDAs. For example, the electrostatic interactions are patently more important for ionic surfactants, whereas the tail conformation energy will have a much greater weight in copolymer SDAs, whose tails are much longer [154].

The morphology of the self-assembled aggregate depends on several factors. At low surfactant concentrations above the critical micelle concentration (i.e., the minimum

surfactant concentration at which micellisation will take place in solution), the morphology of the aggregates tends to that of curved surfaces, such as spheres, [12, 156], since this will maximise the surface-to-volume ratio of the aggregate. Since the hydrophilic head groups are at the surface and the hydrophobic tails are in the core, this will maximise attractive interactions, while minimising repulsive ones. However, at higher surfactant concentrations, interaggregate interactions are no longer negligible. For cationic surfactants, for example, this then results in rod-like or worm-like micelles, i.e. a lengthening of the micelles [156]. In the case of MCM-41, its hexagonal structure is formed by the stacking of these rod-like micelles [35]. Other components in the precursor solution can also determine the morphological outcome of the aggregates. For example, counterions at the surface of a charged micelle can shield the repulsion between head groups, thereby reducing the necessary curvature of the micelle and promoting a sphere-to-rod transition [156]. As will be discussed in detail later in this chapter, it has been a matter of debate in the literature as to whether silica also plays such a role in the synthesis of PMS materials.

One rule of thumb for predicting the morphology of aggregates is the *packing parameter* (p) [150, 154, 157]

$$p = \frac{V_T}{A_H L_T} \tag{3.2}$$

where V_T is the volume of the hydrohobic tail, L_T is its length and A_H is the area of the hydrophilic head group. Generally, $p < 0.33$ corresponds to spheres, 0.33—0.50 to cylinders, 0.50—0.99 to flexible lamellae or vesicles and $p = 1$ to planar lamellae [150, 154, 157].

Surfactant morphologies are often understood using phase diagrams that describe how different system variables affect the geometry of a self-assembled structure. For example, **Figure 3.1** [153] provides an example of a temperature-concentration phase diagram for surfactants. Once again, the importance of surfactant concentration is marked; simply by increasing the concentration of the amphiphile, it is possible to access four different phases. The influence of temperature can also be seen, however **Figure 3.1** shows us that less phases are accessible by this route alone. Generally,

increasing temperature leads to an increase in entropy and therefore a tendency for ordered aggregates to break down. This is especially true for ionic surfactants, however the behaviour of polymer SDAs is often more complicated [153]. For example, it is possible for them to instead display a higher degree of ordering upon heating. These micelles increase in size as the temperature increases (due to the breaking of polymer-water hydrogen bonds) until the micelle radius is equal to the length of the hydrophobic tail. After this point, the once spherical micelle stops swelling and starts lengthening in size, i.e. undergoes a sphere-to-rod transition [157]. An example of such behaviour can be seen in the phase diagram of P123 (see **Figure 3.2**); it can be seen that an ordered hexagonal phase can be converted into a more disordered phase by either increasing or decreasing the temperature, behaviour that is absent in **Figure 3.1**.

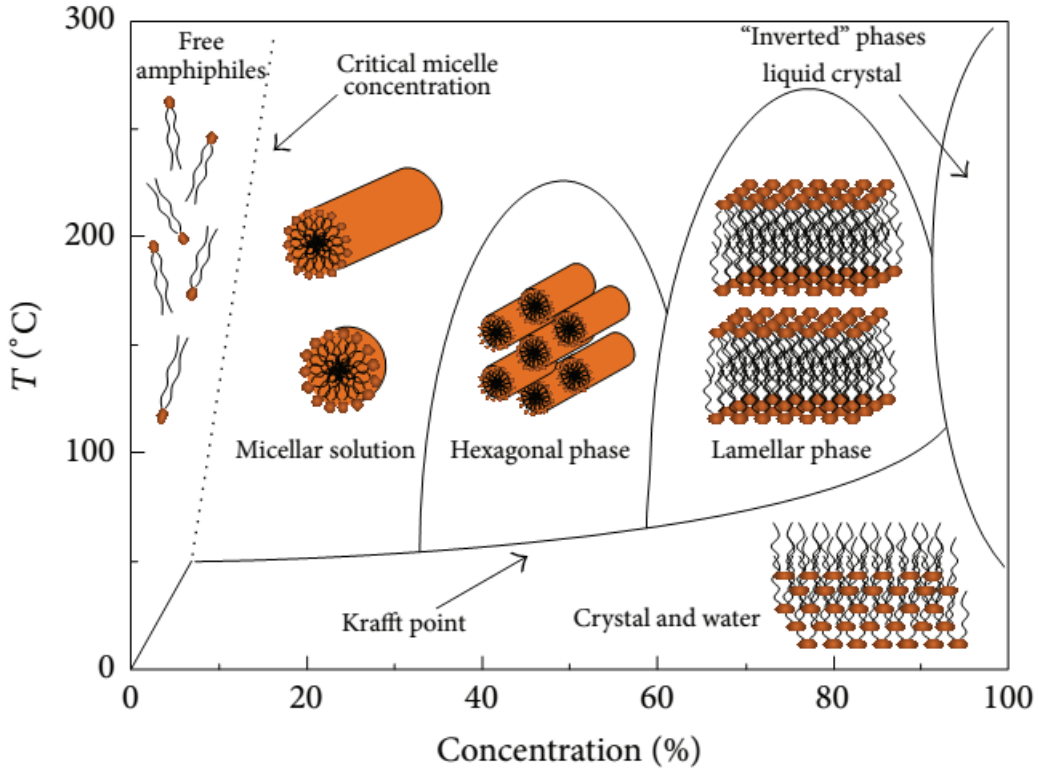


Figure 3.1: Generic example of a typical phase diagram for a surfactant-solvent phase that generates a liquid crystal. Adapted from Lombardo et al. [153]

As can be seen, the morphology of aggregates in self-assembly syntheses is a complex and delicate balance of various interactions and it is replicating this that is so challenging

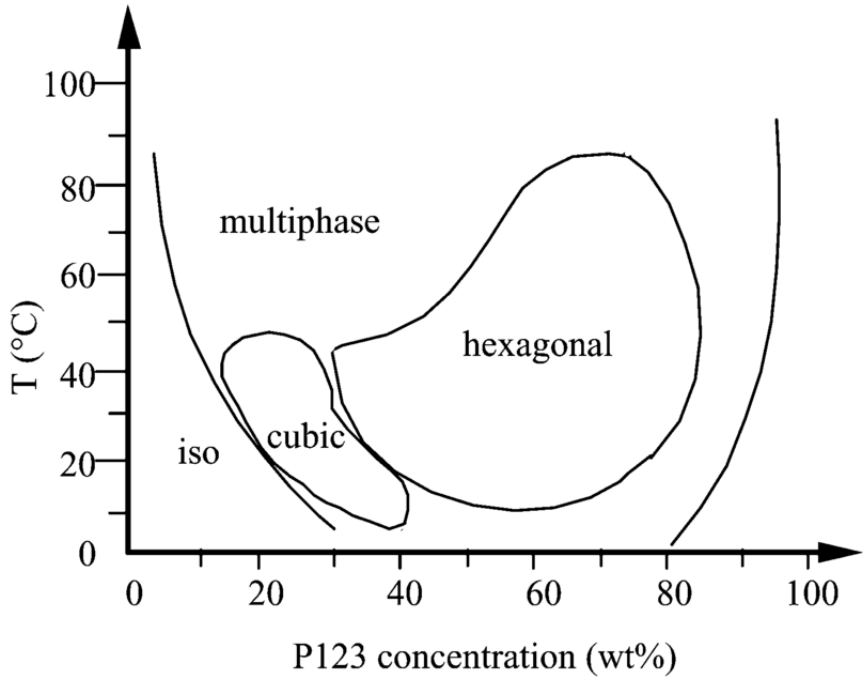


Figure 3.2: Phase diagram of P123, a block-copolymer surfactant, in water. Adapted from Johansson [157]

for the development of computational models.

As mentioned above, the SDA can either be an ionic surfactant or a block copolymer. The most common ionic surfactants are quaternary alkylammonium salts, of which the most popular is n-cetyltrimethylammonium bromide (known as CTAB) [158]. Other common variants are CTAC (the chloride-salt analogue of CTAB), octadecyltrimethylammonium bromide/chloride (OTAB/OTAC) and hexadecylpyridinium bromide/chloride (CPB/CPC) [12]. As can be seen in **Figure 3.3**, CTAB has a polar, i.e. hydrophilic, cationic head group with a bromide counterion and an apolar, i.e. hydrophobic, tail. Some anionic surfactants have also been used [143] but this is uncommon. This leads to the CTAB (and other ionic surfactant) molecules arranging themselves in water so that the polar head groups are positioned at the surface of the micelles, so as to interact with the polar water molecules as much as possible, whereas the hydrophobic tails remain in the micelle core.

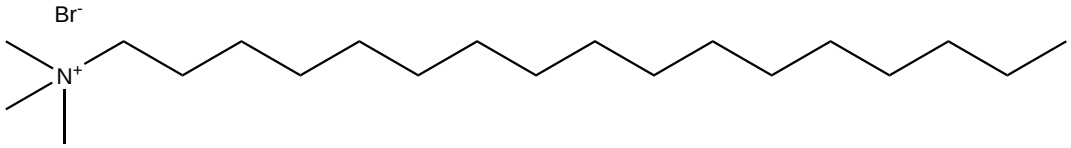


Figure 3.3: Chemical structure of CTAB, the most commonly used surfactant in the synthesis of MCM-41. Altering the length of the hydrophobic chain can be used to tune the pore size of MCM-41

Counterions can play not only an important part in directing the morphology of aggregates, but can also play a key role in facilitating the synthesis mechanisms employing ionic surfactants in the first place. For example, they are not necessary under basic conditions with CTAB-like surfactants; this is because the inorganic silica will be negatively charged, whereas the surfactant head groups will be positively charged. This ensures that there is the strong interaction needed for self-assembly to occur [12]. However, under acidic conditions, the silica will become positively charged; the corollary of this would be repulsion between the surfactant head groups and the silica source, thereby precluding self-assembly. In systems like this, counterions (also called "mediator ions" [12]) of opposite charge are needed to mediate between the two components. If the surfactant was to be anionic instead (as is the case for long-chain alkyl phosphates), then the cases in which counterions were necessary would be reversed.

The above demonstrates how strongly dependent this synthesis method is on electrostatics. Since the necessarily attractive silica-surfactant interaction arises in this way in ionic-surfactant systems, it is important to describe it accurately [8]. Something that complicates this is the medium for the system, namely water. Since water is a highly polar molecule, the charges for the surfactant and silica in aqueous solution will be different as compared to in their pure forms. This means that the charges on the surfactant and silica could be different dependent on the synthesis mixture; some syntheses employ highly concentrated systems, for which this effect would be minimal, whereas in dilute solution it would be appreciable. Furthermore, self-assembly in such systems is determined by phase equilibria. In polar solvents, this will depend strongly on polarisation. For this reason, if an overarching understanding of silica synthesis in

general is to be obtained, it will be necessary to develop a more thorough understanding of polarisation effects in solution. This theme will be the subject of **Chapter 4**.

The second type of commonly used SDAs are block copolymers. Block-copolymer SDAs are different to charged SDAs inasmuch as their interactions are primarily driven by hydrophobic effects and hydrogen bonds, as opposed to ionic (electrostatic) effects. However, the overall principles of their formation of aggregates are similar to that of their charged brethren [154]; one part of the SDA will interact with the solvent molecules in the opposite manner to the other, leading to a rearrangement of the SDA molecules that gives rise to an ordered structure. Block copolymers have the advantage that their aggregates demonstrate higher stability and durability by virtue of their mechanical and physical properties [154].

The experimental aggregation of block copolymers is usually carried out by dissolving them in a common solvent, i.e. one in which both blocks are miscible, and then gradually adding a selective solvent (normally water), i.e. one where only one block is miscible. The addition of water is continued until well above the critical water concentration (i.e., that at which micelles will begin to form), which is normally 25—50 wt.%. Finally, the mixture is quenched via the addition of a large amount of water in order to halt any kinetic processes [154]. This allows aggregates with many of the same morphologies as those from ionic molecular surfactants, e.g. spheres, rods, hexagonally packed rods, bicontinuous phases, lamellae and vesicles, to be formed. Since they are not ionic surfactants, electrostatic interactions are of relatively minor importance for this class of surfactants. Thermodynamically stable morphologies are therefore principally determined by the influence of the remaining three interactions on the free energy. These are: the interfacial tension between the aggregate core and the solvent, the repulsive interactions between the surface groups and the conformational energy of the blocks in the core [154]. The principal routes for influencing these parameters are modifying the type and number of copolymer blocks, the concentrations of copolymer and water, the type of common solvent and the use of additives (usually ions or other polymers). For example, at low water concentrations, i.e. at the beginning of the process, the size of the aggregates is small and spherical due to the number of copolymer-water interactions

being too low to promote phase separation. As the water concentration rises, the size of the aggregates increases; this is to reduce the core-water interfacial energy. Finally, at still higher water concentrations, the rod-to-sphere transition may be observed, driven by the benefit of reducing the stretching penalty of the blocks in the aggregate core. For block-copolymer-templated silica materials, the most common template is a chain with a poly(propylene oxide) (PPO) core that is capped on either end by poly(ethylene oxide) (PEO) (see **Figure 3.4**). The PPO core has an extra methyl group attached to the monomer in order to increase its hydrophobicity, thereby making it amphiphilic.

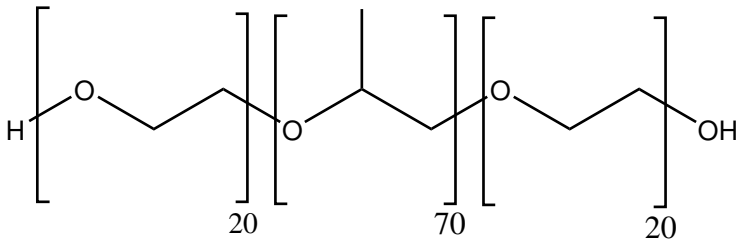


Figure 3.4: Chemical structure of Pluronic P123, which is used as the template for the synthesis of SBA-15. Different templates can be formed by altering the number of monomers in each block

3.2.2 Silica Polymerisation and Materials

Silica-containing (SiO_2) materials are naturally occurring, as well as being produced synthetically. The backbone of these materials are four oxygen atoms arranged tetrahedrally around a silicon atom [159]. The numerous ways in which tetrahedra can be packed and the high level of flexibility in the $Si-O-Si$ bond allow a large variety of complex morphologies to be formed [8]. Such materials can either be crystalline, meaning that their structure possesses short- and long-range order, or amorphous, i.e. demonstrating short- but not long-range order. The most common form of silica is quartz, which is abundant in the earth's crust and which is familiar to the layperson as sand. Quartz is a natural, crystalline form of silica, however synthetic amorphous forms have also been long known, with glass being the most patent example. In this work, we are interested in the chemistry of the self-assembly of synthetic amorphous

silica materials.

Self-assembled silica materials are synthesised via the polymerisation of silicic acids [8,12]. The most simple, and therefore the most well-studied, is orthosilicic acid, which is also the simplest soluble form of silica ($Si(OH)_4$) [160]. Due to its ubiquity, it is commonly referred to as simply *silicic acid*, as will be the case in this work. Its characteristic functional group is the $Si-O-H$ sequence and for this reason it belongs to the silanol class of molecules, i.e. the silicon analogues of alcohols. Although it is soluble in water, it is only stable up to approximately 1 mM, i.e. at very low concentrations, after which silicic acid molecules will start to undergo condensation reactions with one another, polymerising into an amorphous phase [160]. For this reason, silicic acid is usually stored as a precursor molecule that undergoes hydrolysis, i.e. that liberates silicic acid upon entering aqueous solution. The precursor molecule is usually tetraethylorthosilicate (commonly abbreviated as TEOS - see **Figure 3.5**) or tetramethylorthosilicate (TMOS) [158]. These molecules are generally referred to as alkoxysilanes and their characteristic functional group is $Si-O-R$, where R is an alkyl chain. The alkyl chains capping the arms of the alkoxysilane molecule make it more stable in aqueous solution than silicic acid. This is due to the alkoxy groups not reacting with each other, as the silanol groups rapidly would. The hydrolysis and polymerisation reactions can be seen in **Equations 3.3** and **3.4**, assuming full hydrolysis [8,161].

$$Si(OR)_4 + 4(H_2O) \longrightarrow Si(OH)_4 + 4(ROH) \tag{3.3}$$

$$2Si(OH)_4 \longrightarrow (OH)_3Si-O-Si(OH)_3 + H_2O \tag{3.4}$$

Since there are still three silanol ($Si-OH$) groups left after this reaction, the dimers can then polymerise with further silicic acid molecules, leading to oligomers and, if the right conditions are present, a silica mesophase [8]. It is these multiple silanol groups per silicon atom that allow 3D structures to be formed. Both the condensation and polymerisation reactions can be either acid- or base-catalysed; the respective mechanisms

are [8]

$$H_2O + Si(OR)_4 + H^+ \longrightarrow H_2O - [Si(OR)_3] - ROH^+ \longrightarrow HO - Si(OR)_3 + ROH + H^+ \tag{3.5}$$

$$Si(OH)_4 + H^+ \longrightarrow (OH)_3Si(OH_2)^+ \tag{3.6}$$

$$Si(OH)_4 + (OH)_3Si(OH_2)^+ \longrightarrow (OH)_3Si - O - Si(OH)_3 + H_3O^+ \tag{3.7}$$

under acidic and

$$Si(OR)_4 + OH^- \longrightarrow [(OR)_4Si] - OH^- \longrightarrow (OR)_3Si - OH + RO^- \tag{3.8}$$

$$Si(OH)_4 + OH^- \longrightarrow (OH)_3Si(OH)^- + H_2O \tag{3.9}$$

$$Si(OH)_4 + (OH)_3Si(OH)^- + \longrightarrow (OH)_3Si - O - Si(OH)_3 + OH^- \tag{3.10}$$

basic conditions, respectively. The choice of catalyst has an effect on the end product of the process; acid-catalysed polymerisation tends to produce linear chains with occasional cross-linking, whereas alkaline media promote the formation of highly branched clusters, leading to 3D amorphous silica structures [8].

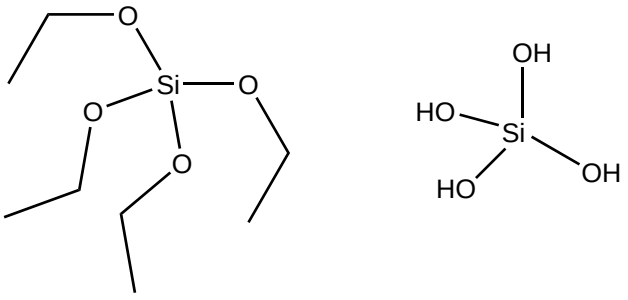


Figure 3.5: Left: chemical structure of TEOS, the most common silica source in the synthesis of PMS materials; Right: chemical structure of silicic acid, the monomer for the polymerisation in the synthesis of PMS materials

In the absence of an SDA, the polymerisation reaction proceeds as a sol-gel process [8, 159]. Sols are dispersions of solid particles in a liquid where the rate of settling or aggregation of the particles is very low [159], i.e. it is a discontinuous system of silica

particles. The first and most famous example of this is the Stöber process, whereby it is possible to synthesise monodisperse silica nanoparticles up to 3 μm [159]. Gels, on the other hand, are formed when the silica particles that have arisen from polymerisation very gradually aggregate and eventually produce a continuous solid phase. The liquid phase can then be removed by evaporation, leading to a reduction in the volume of the material by a factor of 5—10 due to capillary forces, in which case the product is called a *xerogel*. Alternatively, this shrinkage can be avoided by drying the gel in an autoclave above the critical point of the liquid [159]. In this case, the product is classed as an *aerogel*. As the name implies, this process produces materials that are extremely light, with solid volume fractions as low as 0.001 by virtue of their high air content.

As with the synthesis of many silica materials, the growth mechanism in sol-gel processes is still a matter of debate. Matsoukas and Gulari [162] argued that growth of silica particles is achieved only by monomer addition and that no significant nucleation takes place during this stage. They supported this by stating that the growth rate constant was nearly identical to the rate at which monomers are produced by hydrolysis from its precursor. The other camp is that of Bogush and Zukoski [163], who put forward the thesis that nucleation is continually ocurring throughout the synthesis and that these nuclei then aggregate further, producing silica particles.

While sol-gel processes are an appropriate starting point for our discussion of silica materials, in this work we are principally concerned with their formation via self-assembly with SDAs. These surfactant-templated silica materials can be separated into several classes based on pore size and whether they are crystalline or amorphous. Porous materials are classified according to their pore size by IUPAC: microporous materials have pores below 2 nm; mesoporous ones inhabit the range 2—50 nm; above 50 nm, we enter the domain of macroporous materials. The attractive characteristics shared by all materials in this class and that generate such interest are large specific surface areas, ordered pore systems and well-defined pore-size distributions [12]. Of the crystalline type, zeolites are by far the most common and well-studied. However they are restricted to a pore size of 1.5 nm, i.e. are microporous, and have therefore been largely superseded by amorphous mesoporous materials since the discovery and publication of

MCM-41 by Mobil due to their larger range of pore sizes [12].

In this work, we are concerned with the mesoporous silica class and an overview of the most common members in this grouping can be found in **Table 3.1**. We can further subdivide this domain depending on whether the SDA is ionic or a block-copolymer. The classic examples of each route are MCM-41 for the ionic synthesis and SBA-15 for that employing a polymeric surfactant. Since these two materials are by far the most well-studied in their class, we will focus our discussion on these two examples. Both these materials possess a hexagonal pore arrangement (see **Figure 3.9**), although other materials in the same class can have different morphologies (see **Figure 3.6**). SBA-15 tends to have larger pores (4—12 nm [164]) than MCM-41 (usually 2.5—4.5 nm, but can extend up to 10 nm [12, 150, 165]). This is a general trend; ionic-surfactant-templated materials tend to have smaller pore sizes than their block-copolymer counterparts. SBA-type materials tend to have micropores that link their macropores [166], as well as thicker internal walls than their ionic-surfactant brethren [167].

Table 3.1: Enumeration of a selection of the most common periodic mesoporous silica materials. Data taken from Narayan et al. [168]; supplementary and complementary data are noted by additional references

Family	Material	Morphology	Pore Radius (nm)	Pore Volume (cm^3/g)
M41S	MCM-41	2D Hexagonal	1.5—10 [169, 170]	>1.0 [169]
	MCM-48	3D Cubic	1.5—5 [170]	
	MCM-50	Lamellar	2—5	
SBA	SBA-1	3D Cubic [171, 172]	1.4—3.0 [170, 172]	0.5 [171]
	SBA-2	3D-Cage Hexagonal [173]	1.0—4.4 [173–175]	0.14—0.56 [175]
	SBA-3	1D Hexagonal [169]	1.5—3.5 [170]	0.564 [169]
	SBA-11	3D Cubic	2.1—3.6	0.68
	SBA-12	3D Hexagonal	2.5—5.0 [170] [175]	0.83
	SBA-15	2D Hexagonal	4—15 [170]	1.17
	SBA-16	3D-Cage Cubic	4.7—15 [170]	0.91
KIT	KIT-5	3D-Cage Cubic	6.0—12.5 [170]	0.45
	KIT-6	3D Cubic [176]	3.8—18.5 [170, 177]	0.57—1.44 [176, 177]

The thicker pore walls mean that the thermal stability of materials synthesised via polymer-surfactants is higher than those created with the aid of ionic SDAs. The thermal stability of different PMS materials varies depending on the precise synthesis and functionalisation, but is often in the range 400—600°C [12, 178] and can even be as high as 800—1,000°C [179, 180]. While the lower end of this range would be

appropriate for the catalysis of organic materials, it would likely be too low for inorganic catalysis [12], at least for the current generation of materials. Hydrothermal stability may also be an issue due do hydrolysis of the $Si-O-Si$ backbone of these materials. This is most acute in materials where the wall thickness is small; after 120 hours of mild hydrothermal treatment, a 34% decrease in the pore size of the thin-walled MCM-41 is observed, whereas the decrease in SBA-15 (a material with the same hexagonal morphology) is only 3% [167]. This is not the only defining factor though; MCM-48 was shown to experience a fall in pore size of less than 3% [167]. This was put down to the cubic, i.e. different, geometry of its structure. Mechanical compression of 400 MPa is generally enough to destroy most PMS materials [167]. PMS materials typically have large specific surface areas and volumes, meaning that they are suitable for adsorption applications. These are usually in the ranges 600—1,000 m^2/g [12,181,182] and 0.6—1.0 cm^2/g [182], respectively.

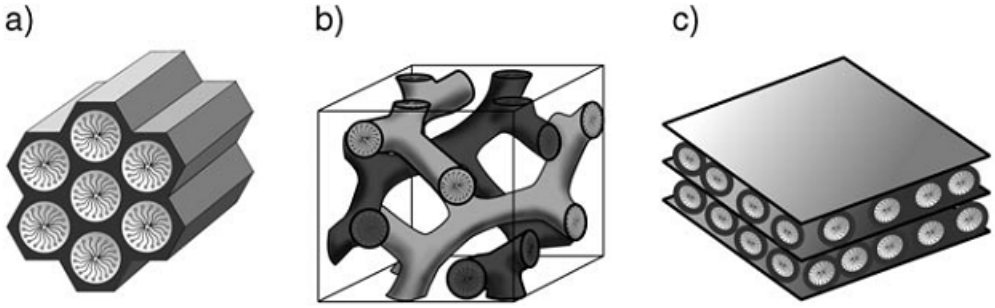


Figure 3.6: Morphologies of various MCM silica materials: (a) MCM-41: 2D hexagonal; (b) MCM-48: cubic; (c) MCM-50: lamellar. Adapted from Hoffmann et al. [12]

The main stages of the synthesis of periodic mesoporous silicas are as follows [15,150]: mixing of the four components (alkoxysilane precursor, SDA, synthesis medium and catalyst); hydrothermal reaction and solid-product phase separation; template removal to free the pores of the silica structure. In the first stage, several processes are at work. First of all, the four synthesis components are stirred in solution. The reaction medium is usually a mixture of water and ethanol; the latter aids the dissolution of the alkoxysilane precursor and the water hydrolyses it. As discussed above, the alkoxysilane condensation reaction is both acid- and base-catalysed and will therefore proceed very

quickly under those conditions, meaning that full hydrolysis is likely, leading to the formation of silicic acid (see **Figure 3.5**). However, the reaction will be much slower at pH values around 7, leading to several TEOS "arms" remaining in solution, which may then be incorporated into the walls of the final material. In some cases, this may be desirable so as to allow functionalisation of the material with these unreacted chemical groups. It is also interesting to note that the hydrolysis reaction consumes water and produces alcohols (see **Equation 3.3**), i.e. the solvent properties of the synthesis medium will be in flux. However, this would likely only be significant at very high silica concentrations. In this initial stage, the template self-assembly proceeds rapidly; for MCM-41, long-range order can be observed within 5—8 minutes [15].

In the second stage, the mixture is heated (to around 100°C for MCM-41 [15,150]) and the mixture is left to age for a period of hours to days. At this point, the silica monomers polymerise at the surface of the SDA micelles [8,35] according to **Equation 3.4**. For ionic-surfactant-templated materials, basic conditions promote smaller silica structures (oligomers and nanoparticles), whereas an acidic medium will favour the formation of cross-linked gels [183]. Block-copolymer-templated materials are normally synthesised under acidic conditions [8,12]. Again, the medium will also be changed by this reaction, since water is produced rather than consumed. After the silicate structure has been formed, the template is then removed by calcination or acid reflux [15,150,158], leaving behind empty pores in its wake.

While the introduction of SDAs complicates the chemistry of the synthesis further compared to non-templated silica materials, it also introduces the ability to tune these materials by utilising these added interactions. A good example of this is the use of surfactants with varying lengths of hydrophobic tails, so as to create a range of pore sizes. Beck et al. showed that increasing the length of the alkyl chain in an ionic surfactant above six, in which case a zeolite, i.e. a microporous material, was formed, they could synthesise MCM-41, a mesoporous one [184]. This occurs because a longer tail section will not be able to pack into the same space and the core of the micelles will therefore swell. Since the volume taken up by the micelles is where the pores will later be located, larger micelles will therefore lead to larger pore radii in the silica

backbone. The interactions between the silica and surfactant can also cause even more fundamental changes to the product morphology. For example, the addition of ionic silicates can in some cases induce a sphere-to-rod transition in micelles, i.e. can change the morphology of the template qualitatively, as well as quantitatively [15]. It has been shown that this is likely due to the high electrostatic contribution of multiply charged silicates interacting strongly with the polar head groups of the SDA [35,36]. The larger silicate charges screen the repulsion between the head groups more strongly, thereby allowing a less curved surfactant geometry to be stable. In other words, the presence of silica allows a material morphology to exist that would not otherwise be available at that surfactant concentration. The addition of further silica may then change the morphology once more, demonstrating how the surfactant/silica ratio can be used to control the synthesis product.

Another factor that can have a strong influence on the synthesised morphology is temperature. Since the self-assembly is driven by phase equilibria, changes in the temperature of the synthesis will change the solubility of the SDA. For example, at low temperature the methyl groups of the hydrophobic PPO chain in P123 (the SDA for SBA-15) become hydrated, thereby forming hydrogen bonds with the water and reducing the hydrophobic effect. When the temperature rises, these hydrogen bonds break, promoting phase separation [157, 185, 186] and causing the micelle to swell, leading to a greater pore size. Ionic surfactants also experience an increase in size with increasing temperature, giving rise to materials with larger pores [187, 188]. Additionally, the rate of polymerisation of the silica will also change with temperature. Finally, swelling agents can also be used to increase pore size [189]. A purely hydrophobic, i.e. non-selective, component may be added; due to its hydrophobic nature, it will locate itself preferentially in the micelle core, causing the micelles to swell. Aromatic hydrocarbons and trialkylamines are common swelling agents [189].

There are three principal ways by which these materials can be functionalised, i.e. functional groups with desired properties can be integrated into them. They are *grafting*, *co-condensation* and *periodic mesoporous organosilicas* (PMOs) [12]. During the silica polymerisation, not all silanol functional groups will react, i.e. some of them will be left

in the structure of the final product. Grafting makes use of this by reacting the exposed silanol groups with another molecule (most commonly: organosilanes, chlorosilanes and silazanes) with a desired functional group so that it becomes incorporated into the surface of the material. The advantage of this method is that it leaves the silica backbone intact. The disadvantage is that the space occupied by the grafted functional groups results in a decrease in the porosity of the material. In the worst-case scenario, these grafted molecules may react preferentially at the pore openings, thereby sealing them and rendering the material useless. Co-condensation employs both tetra- (e.g., TEOS) and trialkoxylsilanes ($R - Si(OR)_3$). The alkane group of the trialkoxysilanes will not be hydrolysed, unlike the alkoxy branches, and will therefore be present in the post-synthesis structure. Since the functional group in this method is incorporated directly into the main backbone of the material, there will not be as large a reduction in porosity as in grafting. The disadvantage is that as the concentration of trialkoxysilanes increases, the degree of disorder will increase due to the lower degree of cross-linking. Additionally, the functional groups will be incorporated into the silica skeleton, but they will not all be accessible. Lastly, calcination cannot be used for the removal of the surfactant; this would likely remove the organic functional groups. PMOs are synthesised using silica precursors of the type $(R'O)_3Si - R - Si(OR')_3$, where R and R' may represent two different organic groups. Using a synthesis of this type solves the problems of the two methods discussed prior. The functional group in this approach is integrated into the silica backbone (i.e., does not block the pores as in grafting) in a regular manner, unlike in co-condensation, where it depends on how the two different precursors react with each other relative to the other. This produces an amorphous material with a narrow pore-radius distribution.

The functionalisation of these materials opens the door to a multitude of possible applications. Due to the high specific surface areas and volumes of these materials as well as the ability to functionalise them to interact preferentially with certain target molecules, gas adsorption and separations are obvious applications. Prominent examples include the separation of CO_2 from air [190, 191] and water treatment [165, 192]. The large pores of PMS materials also mean that catalysis and host-guest chemistry (see

reviews [193–195]) are potential applications due to the larger, i.e. possibly more complex, molecules that can be accommodated compared to, say, microporous zeolites [189]. Controlled transport of molecules, especially for pharmaceutical APIs [182], is also an area of active interest [196]. It has been shown that, by tailoring these materials with certain functional groups, it is possible to control the opening and closing of pores in response to various stimuli. Materials designed to exhibit such behaviour are called *gated materials*. Photochemical [197], thermal [198], pH [199], redox-reaction [200] processes, among others, have all been experimentally demonstrated to be suitable stimuli for controlled molecular transport using this class of materials. For a more detailed discussion of this topic, the reader is directed to the reviews of Alberti et al. [196] and Vallet-Regi et al. [182]. Gated materials could also be used as sensors by designing them so that their pores open and release an indicator molecule in the presence of a chosen chemical. A suitable review of this topic is that of Sancenon et al. [201]. The above are but a selection of the most common applications for PMS materials and demonstrate their enormous potential. However, this vast potential will only be realised if a fundamental understanding of their synthesis can be developed, allowing us to design and tailor new materials systematically rather than heuristically.

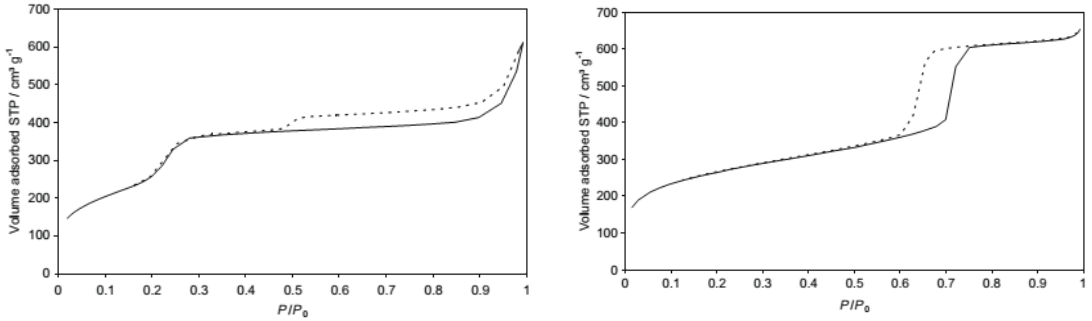
3.3 Characterisation Methods for Porous Silica Materials

The aforementioned key properties of a PMS material, such as pore radius, volume and specific surface area, can be measured in various ways. One common technique for such characterisation is gas adsorption. In this method, the material is first degassed in a glass cell by using an inert gas and heating at 200°C under vacuum overnight [202]. Next, gas molecules are gradually introduced in steps, allowing equilibrium to be established, into the chamber while measuring the amount of gas adsorbed and the pressure. The plotting of the relative pressure and the amount of adsorbed gas is called an isotherm due to the process being carried out at constant temperature. Examples of gas-adsorption data for MCM-41 and SBA-15 can be found in **Figure 3.7**. Thereafter, an isotherm model is chosen, allowing properties such as the surface area to be derived.

By far the most famous of these is Brunauer-Emmett-Teller (BET) theory [203], which is an extension of the Langmuir equation to multilayers. The BET equation is

$$\frac{p^*}{n(1-p^*)} = \frac{1}{n_m C} + \frac{C-1}{n_m C} p^* \tag{3.11}$$

where p^* is the relative pressure, n is the amount of adsorbed material, n_m is the monolayer loading and C is the BET constant. C is related exponentially to the energy of monolayer formation [204]. The BET isotherm data used for analysis are usually those in the relative pressure range 0.05—0.30 [204]. However, if microporosity is present in this range, it may lead to erroneous values [202]. Standard BET analysis employs nitrogen and is carried out at its boiling temperature (77 K) [176,205–207]. Argon is also commonly used (boiling temperature: 87 K) [176]. If the monolayer loading is known, then, in conjunction with the effective molecular area, then the specific surface area can be computed [206]. For nitrogen, it is assumed that the monolayer is close-packed and liquid-like, resulting in an effective area of 0.162 nm^2 [206]. Despite its ubiquity, it has long been known that BET theory is an over-simplified model of gas adsorption. For example, it does not accurately describe adsorbate-adsorbate interactions well at low coverage, as well as assuming that the surface is energetically homogeneous, which is untrue for most materials [206].



(a) MCM-41 (b) SBA-15

Figure 3.7: Adsorption isotherms for nitrogen in: a) MCM-41; b) SBA-15 at 77 K. The solid line indicates adsorption, while the dashed line shows desorption. Adapted from Meynen et al. [202]

For calculating the pore size of porous materials from gas adsorption data, the

Barrett-Joyner-Halenda (BJH) method is most commonly employed [208]. This method makes two assumptions: firstly, that the pore geometry is cylindrical; secondly, that the adsorbed material results from both physisorption and capillary condensation in mesopores (therefore it is valid mainly for mesoporous materials) [209]. The BJH method, therefore, equates the radius of the pore to be the sum of the thickness of the multilayers and the meniscus of the adsorbent. The meniscus is calculated from the Kelvin equation [206, 208, 209]

$$\ln p^* = \frac{2\gamma V_M}{r_m RT} \tag{3.12}$$

where γ is surface tension of the adsorbate, V_M is its molar volume and r_m is the radius of the meniscus. As the pores are evacuated, i.e. the relative pressure is decreased, the film thickness decreases. The decrease in the film thickness can then be correlated with the pore volume, which can then be solved for numerically [208, 209]. Like BET theory, the BJH method relies on several assumptions, such as the validity of the Kelvin equation, that the pores are rigid, that filling and emptying of pores is independent of their location in the network, etc. [206], which may not be fulfilled in practice. Care should therefore be taken in interpreting the accuracy of such data.

Another popular method, especially for macroporous materials, but also for mesoporous ones, is mercury porosimetry. It is routinely used in the pore-size range of 40 nm to 400 μm [176]. Mercury does not readily wet the walls of most materials, therefore a hydrostatic pressure must be applied so that it enters the pores. This applied pressure (P_H) can be related to the pore radius (r_p) according to the Washburn equation, namely [176, 206]

$$P_H r_p = -2\gamma \cos\theta \tag{3.13}$$

where γ is the surface tension of mercury and θ is the contact angle between mercury and the solid pore. Hysteresis is seen in porosimetry for mercury intrusion and extrusion; this is analogous to the hysteresis seen in gas adsorption (see **Figure 3.7** and the difference between the adsorption and desorption lines) due to capillary evaporation

and condensation [176, 206]. By applying percolation theory to this hysteresis loop, information about the pore network can be gleaned [176]. Assumptions made are that the contact angle is constant and that there are no structural changes at high pressures [206]. Gas adsorption and mercury porosimetry are often employed together so that independent measurements may be obtained for the same property [176, 206].

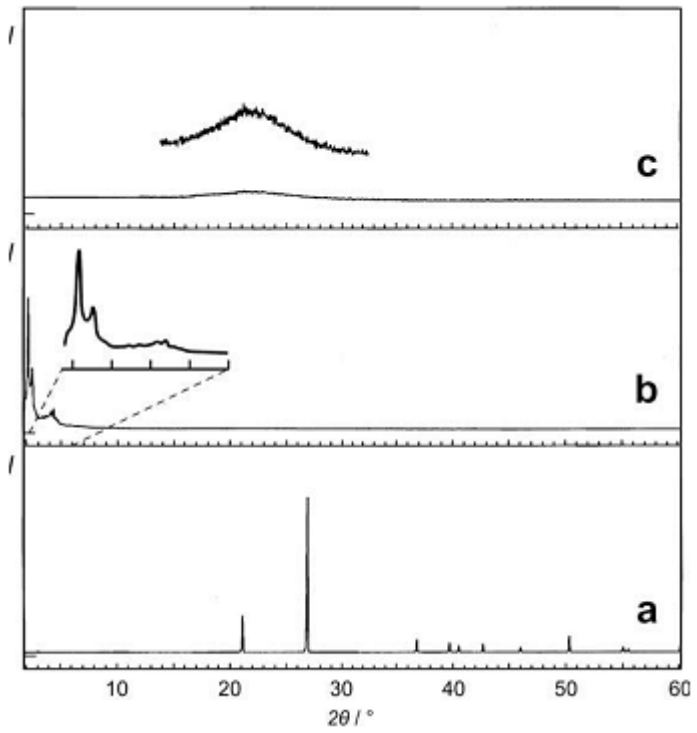
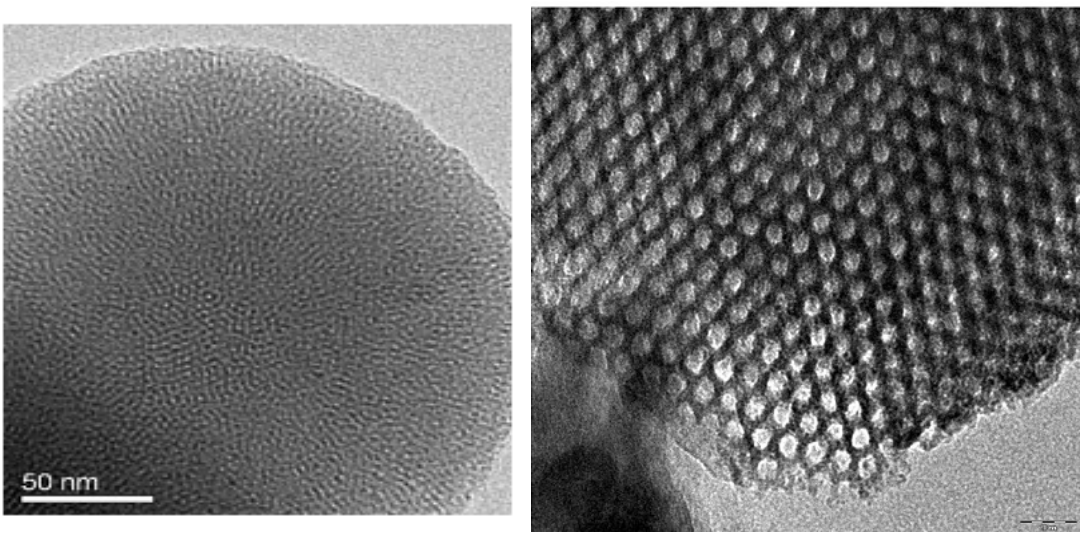


Figure 3.8: Example X-ray diffraction data for different silica materials: a) crystalline α-quartz, resulting in narrow, well-defined peaks; b) mesoporous MCM-48, cubic arrangement of amorphous silica units. Order is only seen at low values of 2θ (see enlarged region) due to having a regular pore structure; c) amorphous silica gel, resulting in one very broad peak due to total lack of long-range order. Adapted from Meynen et al. [202]

A common method for characterising mesoporous silica materials is X-ray diffraction (XRD). In XRD, X-rays of known wavelength are directed at a material. X-rays are used because they have wavelengths in the nanometre range (0.01—10 nm), i.e. on the same scale as the lattice constants of many materials. These waves enter the material's lattice structure and are then diffracted by the atoms comprising it. Constructive interference occurs according to Bragg's Law, namely [202]

$$n\lambda = 2d\sin\theta \tag{3.14}$$

where n is the periodicity, λ is the wavelength of the X-rays, d is the lattice separation distance and θ is the angle of diffraction. Due to constructive interference, more intense peaks will be observed at angles corresponding to the lattice separation, allowing the material to be characterised. The material is therefore acting as a diffraction grating. For crystalline materials, the peaks will be sharply defined, i.e. very narrow and tall, across the range of 2θ, whereas these peaks will be much broader for amorphous materials. However, at small values of 2θ (typically in the region $2\theta = 0.8$—$5°$ [202]), some sharply defined peaks will be observed for mesoporous materials. This arises not from a periodic arrangement of atoms, but rather from the regular pore structure of these materials, whose size happens to be within the wavelength range of X-rays. See **Figure 3.8** for an example of XRD data for different silica materials. This technique can therefore be used for such purposes as characterising a pore arrangement by comparing the diffraction peaks to that of a known material, or relating the intensity of the diffracted beam to pore size [210]. XRD is usually carried out on mesoporous silica in powder form.



(a) MCM-41 (b) SBA-15

Figure 3.9: Transmission electron microscopy images of MCM-41 [202] and SBA-15 [164]

Finally, for imaging samples of materials, scanning and transmission electron microscopy (SEM and TEM, respectively) are two common techniques. In SEM, a beam of electrons are focussed and directed at a material sample. The electrons will interact with the sample and the intensity of the beam of scattered electrons is measured. By scanning across the surface, the intensity at different points is collected, from which an image can be created. These differences in intensity originate from variations in the surface's topography, which are thereby captured in the image. SEM is therefore primarily used for assessing surface topology and morphology. A resolution of around 10 nm is typical [202]. Only conducting materials produce sharp SEM images, therefore silica materials have to be coated with a thin layer of conducting material prior to SEM imaging [202]. Whereas in SEM the detector is in front of the sample, in TEM it is located on the other side, i.e. it measures the electrons transmitted through a thin sample (usually less than 100 nm [202]) of the material. Structural variations throughout the slice cause different amounts of electrons to be diffracted rather than transmitted, meaning that the detector will register different intensities. Atomic resolution can be obtained using TEM, making it the best method for obtaining topographical and morphological information of a sample [202]. An example of TEM images can be seen in **Figure 3.9**. Three-dimensional TEM imaging, i.e. TEM tomography, is also available but is much more burdensome than the two-dimensional variant [202].

3.4 Synthesis Mechanisms for Mesoporous Silica Materials: Liquid-Crystal or Cooperative Templating?

While much is already known about the syntheses of PMS materials, many key aspects of their mechanisms are still unknown, precluding our ability to precisely design and tune materials for specific applications. For example, the precise role that the templating agent plays in the synthesis of surfactant-templated PMS has long been controversial. Beck et al. noted that the mesostructures in solution closely resembled the final material [5] (hexagonal for MCM-41, cubic for MCM-48, lamellar for MCM-50), and on this basis suggested a liquid-crystal templating (LCT) mechanism, whereby the polymerisation

of silica occurs at the surface of a preformed template structure. However, further analysis also showed that these mesostructures formed in the precursor solution (i.e., in the presence of silica) at low concentrations, whilst they only formed dispersed micelles in an aqueous solution at the same concentration. This implied that the silica was no mere spectator in the template formation, but rather an active participant. This led to a second proposed mechanism, the so-called cooperative templating mechanism (CT).

These two proposed routes precipitated a plethora of experimental studies to elucidate the correct mechanism. Techniques applied included: XRD, TEM, infrared and Raman spectroscopy, solid-state and solution NMR, temperature-programmed desorption, gas adsorption, polarised optical microscopy and fluorescence quenching [211–224]. From this multitude of approaches arose significant evidence to support the CT mechanism [211–216], whilst it is still unknown if the LCT mechanism is possible under any set of conditions [217]. Since these studies have seemingly cleared up the principal controversy surrounding the global mechanism, attention has now been directed to elucidating the sub-mechanisms involved under various conditions. Some open questions are:

- Do silicates promote a rearrangement of the surfactant template from spheres to rods/wormlike micelles or not?
- Do silicates prefer to interact with micelles or monomers in the early stages of the synthesis?
- Does the liquid-crystal mesophase arise from the aggregation of spherical or cylindrical micelles?

As can be seen, although significant progress has been made in illuminating the complex formation of PMS materials, many challenges remain before we can begin to be able to claim any profound understanding of their synthesis mechanisms. Against this background and the rapid development of computational power, the appeal of computational modelling is obvious; the ability to observe these processes with atomic-level resolution would clarify many of the uncertainties still associated with the synthesis of these materials.

In the following sections, the results of modelling studies will be presented and discussed with regard to their insights into the mechanisms involved in the formation of PMS materials. From a material perspective, these studies can be divided into three broad categorisations:

1. MCM-41 and ionic-surfactant-templated PMS materials
2. SBA-15 and block-copolymer-templated PMS materials
3. *in situ* functionalised PMS materials

From a modelling perspective, studies can broadly be placed under one of three headings:

1. Lattice models - primarily used for exploring the phase diagrams of precursor solutions
2. Off-lattice models - primarily used for probing silica self-assembly
3. Reactive potentials - primarily used for simulating silica condensation

It should also be clarified here that we are only interested in phenomenological models, i.e. those which produce ordered material structures from a disordered precursor solution along a realistic kinetic pathway. Many studies have focussed instead on creating realistic PMS structures using either stochastic methods [225–227] or by modifying those of known silica structures [228]. While these are useful for practical applications, they do not illuminate the mechanism by which a given material forms experimentally. Therefore, they do not help us with our goal to further our ability to design these materials *in silico*.

3.5 Lattice Models - Exploring Template Phase Diagrams

3.5.1 MCM-Type Materials

The very first computational work on PMS synthesis was carried out in 2001 by Siperstein and Gubbins [229,230]. They focussed on exploring the phase diagrams of ternary

surfactant/solvent/silica systems, i.e. a surfactant-templated silica system. They did this by expanding an earlier model of Larson et al. [231] for surfactant self-assembly. They modelled the surfactants using hydrophilic head and hydrophobic tail beads (H_iT_j); three variations thereof were investigated - H_4T_4, H_4T_6 and H_2T_6. They also employed solvent (S) and silica (I) beads. The interactions between beads can be summarised as follows: silica beads interacted strongly with heads in order to describe their ionic interactions; tails interacted strongly with each other in order to describe hydrophobic clustering; finally, silica-solvent interactions were varied in order to describe various association regimes. They then carried out lattice Monte Carlo (MC) simulations in the canonical (NVT) ensemble, where each bead could occupy precisely one site on the cubic lattice. Three different MC moves were allowed: reptation, chain regrowth and cluster moves.

Siperstein and Gubbins' first point of call was to validate their model for binary systems, i.e. surfactant/solvent and surfactant/silica systems. For the former system, they managed to replicate Larson et al.'s result, demonstrating a surfactant phase transition from a hexagonal to a lamellar structure (via a small gyroid phase) by increasing the surfactant volume fraction. For the second system, they also achieved similar results; with increasing surfactant volume fraction, the system progressed from micelles to a hexagonal phase, before ending up as perforated lamellae. They then compared the size of the aggregates in both phases; the surfactant/solvent aggregates contained three times as many surfactant beads than the surfactant/silica system. This was due to silica-head interactions being weaker than solvent-head ones, leading to less phase separation, i.e. smaller aggregates.

They then turned their attention to the more complex ternary systems (surfactant/solvent/silica) and considered two different scenarios: in the first, silica was completely soluble in the solvent (I-S interaction energy equals zero); in the second, it was immiscible (positive I-S interaction). They proposed that these two scenarios corresponded to two different mechanisms, namely associative and segregative separation, respectively. In the former, phase separation is driven by strong interactions between the two species, while in the latter a strong repulsion is the main driver. They showed

that, independent of the type of mechanism, hexagonal phases were formed at high silica/surfactant ratios, while a low ratio produced lamellar phases, as shown in **Figure 3.10**. This corresponds to what is seen during the experimental synthesis of MCM-type materials [5,217]. While this work was extremely informative, it was only able to capture qualitative trends due to the inherent simplicity of the lattice model that restricts the number of surfactants that can interact with a silica species at any given time.

Finally, they investigated the effect of temperature and surfactant composition on the phase diagrams. They noted that increasing temperature led to lower surfactant but higher silica concentrations in the surfactant-rich phase, which promoted the formation of hexagonal as opposed to lamellar phases. This was in accordance with experimental results [213]. Finally, they analysed the effect of surfactant composition by altering the length of the hydrophobic surfactant tails. Lowering the number (i.e., reducing the surfactant hydrophobicity) promoted hexagonal phases (i.e., better miscibility between phases), whereas increasing the length saw the formation of lamellar phases (i.e., those which result from a high degree of phase separation).

3.5.2 SBA-Type Materials

One of the complexities of modelling triblock-copolymer-templated materials like SBA-15 is that the template molecules are much larger than for traditional surfactants. On the one hand, this gives them their characteristic larger pores, however it also makes modelling such materials at an atomistic level extremely challenging. For comparison, Pluronic P123 is comprised of around 110 monomers, whereas CTAB has a chain length of around 20 heavy atoms. For this reason, it was only natural for the coarse-grain method to be extended to simulations of SBA-like materials. The first work of this type was carried out in 2005 by Bhattacharya and Gubbins [166,232], expanding the model from surfactant-templated PMS to block-copolymer-templated materials. They did this by creating a new template structure, changing the usual template model H_iT_j, representative of the hydrophilic head and hydrophobic tail of ionic surfactants, to $H_iT_jH_i$, which imitates the arrangement of Pluronics, i.e. a hydrophobic core section bookended by two hydrophilic caps. They then recalibrated the bead interaction parameters for

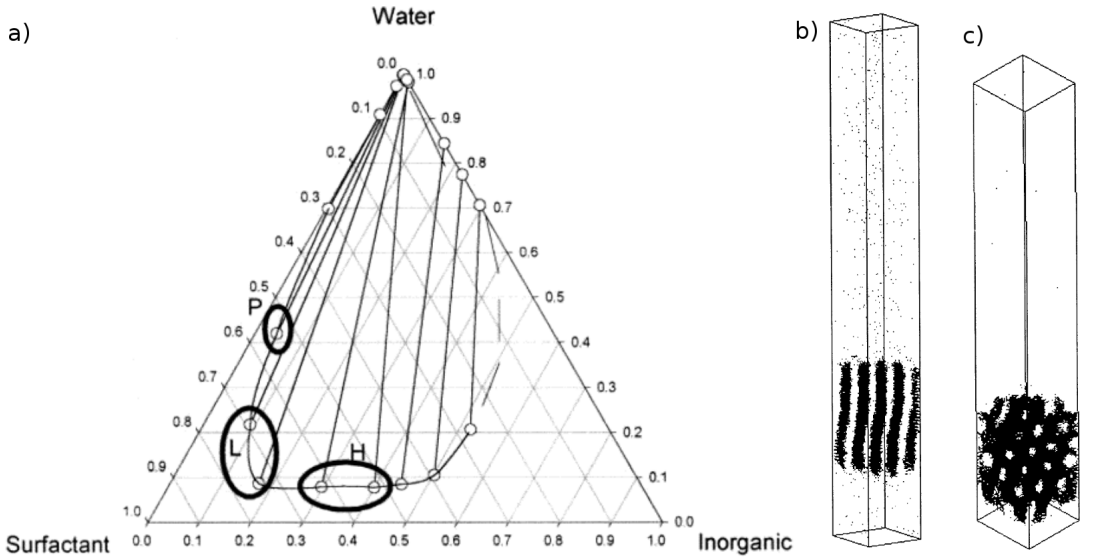


Figure 3.10: (a) Ternary phase diagram of surfactant/water/silica obtained from Monte Carlo simulations, showing the phase separation region. Circles highlight different mesostructures formed in the concentrated phase: H - hexagonal; L - lamellar; P - perforated lamellae. Simulation snapshots showing examples of the lamellar phase (b) and of the hexagonal phase (c). Only silica beads are shown for clarity of visualisation. Adapted from Siperstein and Gubbins [230]

the new species.

In the first stage of their simulations, they achieved qualitative agreement of the binary water/surfactant phase diagram (using a template of $H_3T_5H_3$) for Pluronic P123 with their model. At low concentrations, the simulations produced micelles before transitioning to a hexagonal phase at intermediate concentrations and then finally forming a lamellar phase at the highest ones. They then probed the effect of the head-tail bead ratio; to this end, they studied different template architectures - $H_3T_5H_3$ and $H_2T_7H_2$. Interestingly, this second variant has the same ratio as Pluronic P123. They found that the second template structure organised as lamellae, apart from at very low concentrations, where spherical micelles are formed, seemingly avoiding curved structures entirely. On the other hand, $H_3T_5H_3$ templates gave rise to both cylindrical and lamellar structures. This shows good qualitative agreement, showing that increasing the degree of hydrophobicity of the template leads to structures that minimise the surface area of the aggregate (i.e., lamellae). However, the quantitative result is imperfect since Pluronic

P123 can form cylindrical structures and has the same head-tail ratio as $H_2T_7H_2$.

Next, the authors looked at the ternary system of template/solvent/oil. Oil here is generally used to refer to any hydrophobic solvent. The authors were also interested in materials apart from those of the SBA-15 type, such as mesostructured cellular foams (MCFs), which have very large spherical pores on the range of 10 to 50 nm. A key component of the synthesis is a hydrophobic solvent, normally trimethylbenzene (TMB), whose presence above a critical value causes a transition from a cylindrical template (á la SBA-15) to a spherical one, all the while significantly increasing the pore size. In these ternary simulations, they observed spherical micelles at low surfactant and oil concentrations, then elongated and hexagonal structures at intermediate surfactant concentrations, before becoming lamellae when the oil concentration was increased further.

Finally, the model was extended to quaternary phases by adding silica to the ternary system. In this system, since the silica interacted more strongly with the head groups than with the water beads, the phase separation of a surfactant- and silica-rich phase from a water-rich phase was observed. The authors then explored the effect of increasing the oil volume fraction on the self-assembled template structures. They observed that at low oil volume fractions (2%) cylindrical micelles were present (i.e., SBA-15-type templating). After increasing the oil volume fraction to 9%, a lamellar phase was observed. At high oil volume fractions (23%), the authors observed a mesocellular structure (i.e., MCF-type templating). Interestingly, this means that the model was capturing the experimental behaviour of the system at high and low oil volume fractions. However, the intermediate phase did not agree with the experimental work of Lettow et al. [233], who had posited undulating cylinders as the transitional phase.

Up until now, the results have mainly focussed on phase equilibria and have neglected dynamics. One method that has been successfully applied to illuminate this other important aspect of SBA-like material synthesis are mesoscopic-dynamics (MesoDyn) models [234], i.e. dynamic models with a high degree of coarse graining for large systems. MesoDyn simulations have been widely used to successfully model the self-assembly of other polymer systems, striking a good balance of detail and computational tractability

for such systems. This method is based upon dynamic mean-field density functional theory and the system develops over time based upon chemical-potential gradients and Langevin noise. The beads interact according to Flory-Huggins theory and thereby the free energy is minimised. The noise term is important for hindering the mesostructure becoming stuck in local energy minima, i.e. provides a way to more efficiently sample the whole (and in this case very large) energy surface.

The first group to use this technique to explore the synthesis of PMS materials was that of Chen et al. [185] in 2009, who used it to study the effect of temperature on the mesostructure pore size of SBA-15. Building on the work of Vlimmeren et al. [235], who had previously proposed a mapping for the coarse graining of P123, they initially investigated a binary system of P123/water (10% P123) using a cubic lattice of edge length 32 nm (see **Figure 3.11a**). The simulation time was 1.0 ms (20,000 time steps) and the P123 topology after coarse graining was $(PPO)_5(PEO)_{21}(PPO)_5$. Based on their order parameter (P), which described how heterogeneous the system was compared to the initial homogeneous configuration, they showed that the formation of spherical micelles followed three stages: (i) P increased slowly during the initial stages; (ii) after 750 time steps, P increased rapidly; (iii) at 1,500 time steps, P once again increased very slowly and gradually reaches equilibrium, i.e. phase separation has completed. They then proceeded to simulate the binary system at two temperatures, namely 293 and 333 K. They found that the size of the micelle cores increased with temperature and posited that this was due to the hydrophilic nature of polyethylene oxide (PEO) being reduced at higher temperatures, causing them to migrate from the micellar surface to the core. This expansion is likely to be the reason for the increase of SBA-15 pore size under higher-temperature synthesis conditions. While this study was useful in illuminating this aspect of the synthesis, they neglected to include silica, which could change the micellar structure once more.

In 2012, the same group further developed their model to include silicates and electrostatics [237]. They stated that they estimated interaction parameters for silica using "the cohesive energy density of different molecules and the miscibility behaviour of binary mixtures", but other details about this process were scant. They also introduced

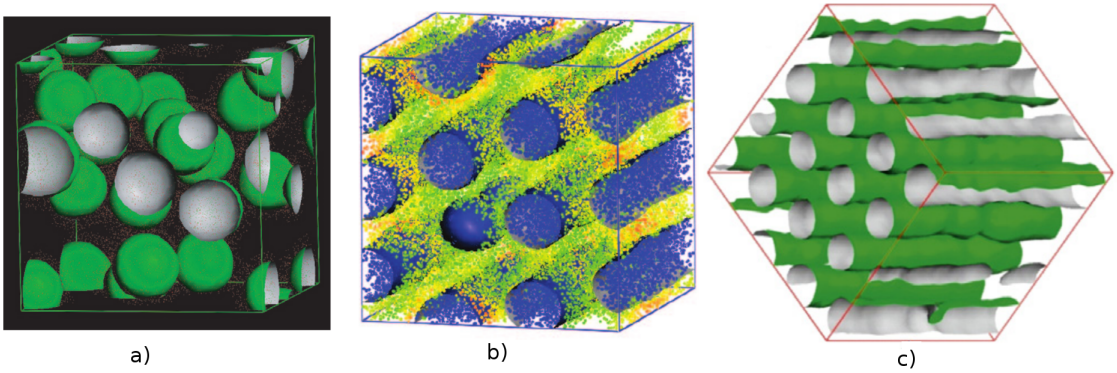


Figure 3.11: Examples of structures obtained from MesoDyn simulations of Pluronic 123 surfactants: (a) spherical micelles obtained in quiescent copolymer/water solutions, adapted from Chen et al. [185]; (b) hexagonal mesophase obtained in copolymer/water solutions with constant imposed shear and charged PEO beads, adapted from Yuan et al. [236]; (c) hexagonal mesophase obtained in silica/copolymer/water solutions in the presence of shear with charged PEO beads, adapted from Chen et al. [237]

shear in order to imitate the stirring which occurs in the experimental synthesis. They fitted Mulliken charges to *ab initio* QM calculations to arrive at a PEO bead charge of -0.3, while PPO was left uncharged. Silica was assigned a charge of +1 since it was argued that the anionic silica would be surrounded by H^+ ions under the acidic conditions of the synthesis. The same lattice size, simulation time and time step were used as in their previous work. They found that the mesophase formation underwent four stages: (i) the solution formed a continuous phase with some pre-micelles between 0 and 0.1 ms; (ii) spherical micelles were formed between 0.1 and 0.2 ns; (iii) at this point, a shear was applied and the structure was rapidly broken up before cylindrical micelles formed by 0.4 ms; (iv) from 0.4 ms until 1.0 ms, the cylindrical micelles aggregated into a hexagonal arrangement. In this way, they achieved the hexagonal mesophase that is typical of SBA-15 (as shown in **Figure 3.11c**). Interestingly, they also varied the charges of the silica and PEO beads. They found that changing the silica charges to +0.5, +1.5 or +2.0 destroyed the hexagonal structure of the mesophase, leaving only disordered cylindrical micelles. For the PEO beads, they tried charges of -0.1, -0.2 ... -0.5; they only saw the appearance of the hexagonal mesophase at -0.2 and -0.3.

A similar work to Chen et al.'s was that of Yuan et al. [236]. They also used

MesoDyn simulations to model the self-assembly of P123 but used a much wider range of concentrations and, like Chen et al.'s later work, they employed shear and applied charges to the PEO beads. However, they carried out their simulations in the absence of silica. They observed mesophases ranging from spherical micelles to a bicontinuous phase and finally a lamellar arrangement. When they made the PEO beads positively charged (to mimic H^+ ions) in the presence of shear, they noted the successful formation of a hexagonal mesophase via a progressive elongation of the micelles (see **Figure 3.11b**). Most interestingly, they found that both shear and charge were necessary to achieve the hexagonal geometry; the presence of only one of the two factors was not sufficient to induce it. These two studies are an excellent example of the fact that the success of such simulations depends strongly on an accurate rendering of the competing chemical and physical interactions.

3.5.3 *In Situ* Functionalised Materials

As we have seen, the surfactant model of Larson et al. [231] was successfully used to describe the first two (namely MCM-type and SBA-15 type) of our three material streams. In 2007, Patti et al. [238] further expanded the use of this method to describe our third stream, *in situ* functionalised materials. They did this by modelling hybrid organic-inorganic precursors in the synthesis solution. As in the original work, Patti et al. considered linear surfactants, namely H_4T_4. Inorganic silica was mapped as two connected segments on the lattice, with parameters representing either complete miscibility (I_2) or immiscibility (I_2^*) with the solvent (S). The organic groups were considered to be either hydrophobic, i.e. equivalent to tail groups, or hydrophilic, i.e. equivalent to heads. This led to four distinct possibilities for hybrid precursors: IT, I^*T, IH, I^*H. Patti et al. performed MC simulations in the NVT ensemble at different surfactant concentrations and found that the system phase separated into a concentrated surfactant phase and a solvent-rich phase. Interestingly, they found that the nature of the organic group had a profound influence on the formation of ordered mesophases. Hydrophobic groups acted as co-surfactants and hydrophilic groups played the role of co-solvent. In fact, considering a hydrophobic terminal group (a ternary system of

$H_4T_4/IT/S$), no ordered phases were found, regardless of surfactant concentration. The authors proposed that this explained the difficulty in experimentally obtaining ordered materials from terminal organic-inorganic precursors [239]. In all ordered mesophases, Patti et al. observed that the terminal organic group was mainly located in the solvent side of the micellar aggregates, thus suggesting that at least some of the functional groups would not be available on the pore surface upon the completion of the synthesis.

A similar approach was later used [240] to study the phase diagrams of systems containing bridging organic precursors (i.e., IHI, I^*HI^*, ITI and I^*TI^*) and to compare them to the purely inorganic and terminal hybrid precursors. Once more, they observed phase separation in their systems, driven by the presence of inorganic components, in some cases yielding ordered mesophases. The main conclusion of their second paper was that bridging precursors are more effective at promoting phase separation and the emergence of order in the concentrated phase, presumably because the potentially disrupting organic moieties are protected by inorganic groups on both sides.

In a later study [13], the same group extended their approach to consider more complex hybrid precursors, aiming to represent organosilicates functionalised with amine groups, namely: aminopropyl-trimethoxysilane (AP, represented as I_2TH), methyl-3-aminopropyl-trimethoxysilane (MAP - I_2THT) and phenyl-3-aminopropyl-trimethoxysilane (PAP - I_2THT_3). They then examined four-component systems comprised of a pure silica source, solvent, surfactant and one of the aforementioned hybrid aminosilcate species. Importantly, the surfactant was no longer linear, as in previous studies, but composed of a branched head T_5HH_3, which is more in keeping with the actual shape of the molecule used experimentally.

Patti et al. showed that at high concentrations of AP (15 vol.%) and MAP (25 vol.%), the hexagonal geometry of the surfactant mesophase broke down, in qualitative agreement with experimental observations. They posited that this was due to amine groups being relatively solvophilic, leading to the increased solubility of the surfactant in the solvent and promoting the formation of disordered mesophases. Conversely, high concentrations of PAP (above 17 vol.%) resulted in a transition from a hexagonal to a lamellar phase. In this case, the hydrophobic nature of the large tail group of PAP

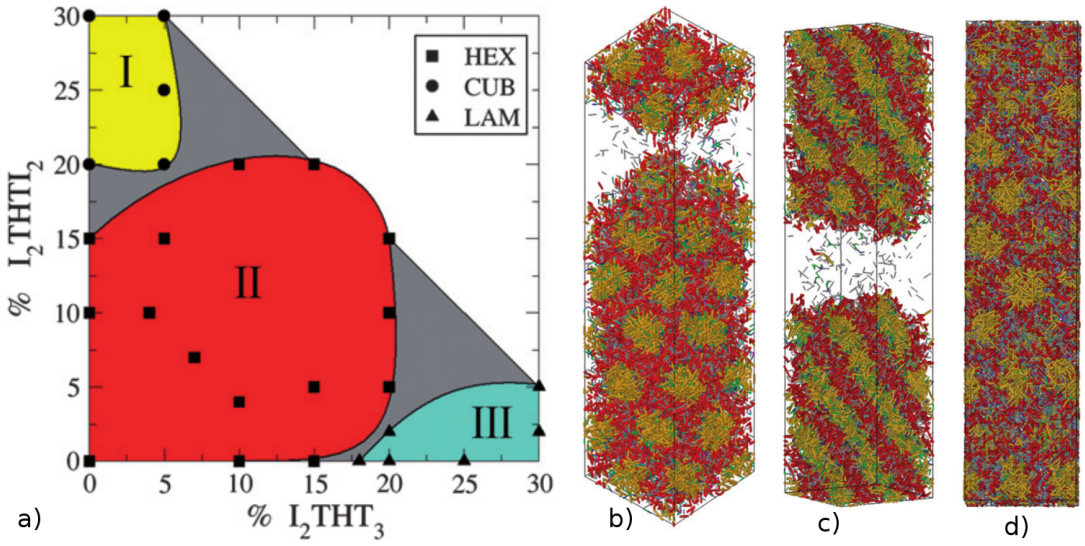


Figure 3.12: Two-dimensional phase diagram (a) obtained from lattice Monte Carlo simulations of solutions with 40% surfactant, 10% silica dimers and variable concentrations of terminal and bridging hybrid organosilicate precursors. The labelled regions indicate stable cubic (I), hexagonal (II) and lamellar (III) mesophases. Snapshots showing examples of hexagonal (b), lamellar (c) and cubic (d) mesophases. Colour scheme is: red - surfactant heads; yellow - surfactant tails; gray - silca; blue/green - organic groups of terminal precursor; magenta/cyan - organic groups of bridging precursor. Adapted from Patti et al. [14]

caused a swelling of the hydrophobic core of the surfactant micelles, which promoted a lower curvature and hence a lamellar phase. In the case of the longer precursors (MAP and PAP), the authors noted that the organic groups penetrated the micelle core, suggesting that they would become available for surface interactions. However, for AP it was shown that the functional amine group was encompassed almost entirely by the inorganic walls, rendering the functional group inaccessible. The authors point out that this contradicts experimental results and ascribe the discrepancy to the inherent simplifications of the lattice model.

The last study in this series by Patti et al. [14] considered longer hybrid precursors in the terminal and bridged forms, namely I_2THT_3 and I_2THTI_2. Similarly to the case with smaller precursors, they observed that the bridged arrangement was more favourable for the formation of ordered hexagonal phases. More interestingly, they considered mixtures of the two precursors, leading to a five-component system, in an

attempt to describe the formation of so-called bifunctional mesoporous silicas. Their results showed that the bridging precursor (BP) acted as a co-solvent due to both ends being hydrophilic, whereas the terminal precursor (TP) acted a co-surfactant. In this way, the surfactant solubility could be controlled by adjusting the concentrations of these two precursor species, and this induced transitions from hexagonal to either cubic or lamellar phases, as demonstrated in **Figure 3.12**. Where the BP and TP concentrations were similar, they tended to cancel each other out and a hexagonal phase was formed. However, an excess of TP relative to the BP precursors produced lamellar phases, and the reverse case resulted in cubic phases.

The final and most realistic set of simulations that were carried out on the basis of Siperstein and Gubbins' model was that of Jin et al. [15]. In their set-up, they introduced three key improvements over their predecessors: (i) they introduced a diamond sub-lattice (i.e., had two separate but interconnected lattices) upon which only the silicates could move on. The reasoning behind this is that it would allow the model to better describe silica's tendency to form tetrahedral networks and hence be more realistic geometrically; (ii) they allowed silica condensation to occur. They modelled this by allowing two oxygen atoms to share the same lattice site with a very favourable interaction energy - an order of magnitude higher than the silica-head interaction. If the polymerisation move was accepted, then the two terminal oxygens were replaced by a $Si - O - Si$ bond. This

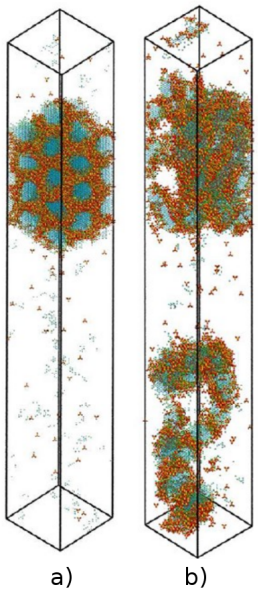


Figure 3.13: Snapshots of mesophases obtained from lattice Monte Carlo simulations with both surfactant self-assembly and silica polymerisation: (a) hexagonal phase obtained from a two-step process; (b) disordered phase obtained from a one-step process. Colour scheme: blue - surfactant tails; green - surfactant heads; red - neutral oxygens; purple - negatively charged oxygens; yellow - silicon. Adapted from Jin et al. [15]

also left an empty site (where one of the oxygen atoms had migrated from), which was equivalent to the water produced during the condensation reaction, since empty lattice sites were treated as an implicit solvent; (iii) deprotonation equilibria were explicitly modelled through a reactive Monte Carlo (RMC) method, thus allowing for simulations at high pH (i.e., with anionic silicates) to be performed. The interactions were calibrated by trying to replicate the high-temperature phase diagram of the surfactant, CTAB, in water.

Jin et al. considered two different scenarios in their attempt to elucidate the synthesis of MCM-41. The first scenario was a two-step process: in the first stage, no silica polymerisation took place and the system was allowed to self-assemble and phase separate; in the second, silica polymerisation was allowed to occur. In the first stage, which was a binary water/CTAB system, spherical micelles arose from the self-assembly process. However, once silica was added to the sub-lattice and polymerisation was allowed to occur, the system formed a hexagonal phase, i.e. an MCM-41 template. This was an important result since it provided strong evidence in support of the CT mechanism, since only in the presence of silica was the desired hexagonal phase able to form. Crucially, they also went to great length to show that this was an equilibrium transition and not an artefact of the simulation. They did this by showing that the system could be made to transition reversibly between lamellar and hexagonal mesophases by heating and cooling. This can be observed in **Figure 3.13**. However, they also noted that this could only occur before extensive polymerisation had occurred, otherwise the structures became locked in place.

The above results were compared to a second scenario in which both mesostructure assembly and silica condensation were allowed to occur simultaneously from the very start of the simulation. Interestingly, they observed that rapid polymerisation precluded the development of template ordering, instead producing a glassy material. This result suggests that the mesoscale structure of the final solid can be controlled only during the initial stages of the synthesis, before silica polymerisation becomes extensive. Even though the results of this study were insightful, the inherent simplicity of the lattice model once again makes it difficult to establish any quantitative comparison

with experiment. For example, the authors were not able to achieve the same degree of polymerisation in their simulations as observed in experiments, a discrepancy that was attributed by the authors to the geometric restrictions of the lattice.

3.6 Off-Lattice Models - Probing Template Self-Assembly Mechanisms

3.6.1 MCM-Type Materials

Up until now, we have addressed models with low degrees of configurational freedom and chemical resolution; we will now move on to models with a continuous configurational phase space and atomistic or near-atomistic resolution. The thought behind this shift is to move away from only qualitative insights towards quantitative ones. In this vein, 2007 saw the first published molecular dynamics (MD) simulations of MCM-41 synthesis using realistic atomistic models [241]. The authors started from a dilute aqueous solution of surfactant, to which anionic silicic acid monomers were added, in an attempt to mimic the initial stages of the synthesis of MCM-41, i.e. at the conditions of the original experimental work, namely high pH and low surfactant concentration. Decyltrimethylammonium bromide (DeTab) was chosen for the surfactant instead of the more common CTAB for the sake of computational tractability; they modelled it using a modification of the OPLS potential [242]. Meanwhile, water was described by the SPC/E model [31]. To describe the silica molecules, Jorge et al. adapted an existing model from Pereira et al. [161] by modifying geometric and electrostatic parameters on the basis of results from QM calculations on a wide range of neutral and anionic silicates [243].

Their initial study [241] yielded two major insights into the early stages of the synthesis: (i) silicates were observed to interact strongly with small surfactant micelles, rather than free monomers; (ii) the adsorption of silica monomers on the micelle surface caused an increase in the size of the micelles. This links well to the study of Jin et al., who showed that if extensive polymerisation occurred before micelles and larger surfactant structures had formed, then they could not produce an MCM-41-like hexagonal

structure, only a glassy material.

This was followed up by a more detailed report, in which solutions of silicate oligomers were also considered [35]. More precisely, the authors ran simulations from several starting configurations, whose composition each corresponded to a different stage of the synthesis. These were: (i) a reference solution (no silicates present, describing the initial surfactant solution); (ii) a monomeric solution (with silica anionic monomers instead of bromide counterions - in this way the charge of the overall solution was maintained); (iii) an oligomeric solution with small silica oligomers corresponding to a distribution from solution NMR experiments at high pH [244]. From their results they concluded that cooperative templating does indeed occur, since the size of the surfactant micelles was larger in the monomeric solution than in the reference solution (see **Figures 3.14a** and **b**), while the surfactant micelles in the oligomeric solution were bigger still. This showed that the silica was no mere spectator, but rather directly influenced the morphology of the surfactant self-assembly. At the pH being modelled (11), almost all of the silica species were deprotonated and it was thus concluded that the increase in micelle size was due to silica species being more effective at screening the positive charge of the surfactants than the bromide ions. They quantified this by the average number of surfactants bound to each counterion: 1.31 for bromide ions but 1.64 for silica monomers. The oligomers were even more effective at producing larger micelles since they can be multiply charged, increasing their screening ability. Interestingly, the multiply charged silica oligomers were also able to bind to multiple micelles simultaneously (1.36 on average, compared to a little more than 1.0 for both bromide and monomeric ions), which was suggested to promote aggregation of micelles.

Jorge et al. [35] also considered an exchange solution, which was produced by taking an equilibrated reference solution and adding silica monomers, while maintaining the initial concentration of bromide ions. This was meant to more realistically mimic the experimental addition of a silica source to a pre-equilibrated surfactant solution. Results of that simulation showed again that silica preferentially binds to micelles, and clearly showed ion exchange at the micelle surface (the degree of silica binding was 76% by the end of the simulation, i.e. they had replaced more than three quarters of the bromide

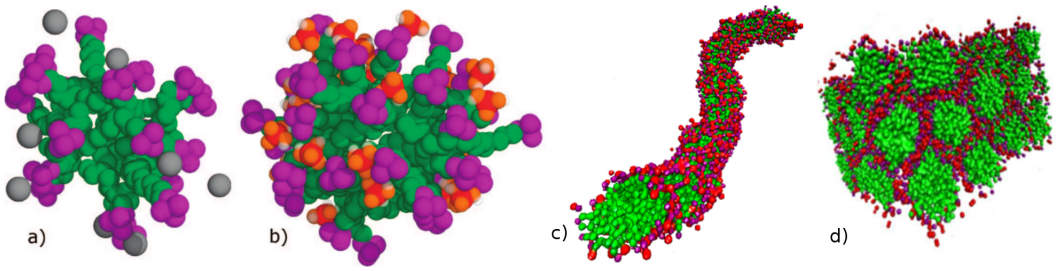


Figure 3.14: Snapshots showing the evolution of silica/surfactant aggregate structures during the early stages of MCM-41 synthesis, as observed from MD simulations using a multiscale modelling approach: (a) small spherical micelle in a dilute surfactant/water solution (atomistic model), adapted from Jorge et al. [35]; (b) larger spherical micelle in a solution with silica monomers at short times (atomistic model), adapted from Jorge et al. [35]; (c) rod-like micelle in a silica monomer solution after long equilibration times (coarse-grained model), adapted from Pérez-Sánchez et al. [36]; (d) hexagonal phase in a solution with silica dimers (coarse-grained model), adapted from Pérez-Sánchez et al. [245]. Surfactant tails are shown in green, heads in purple, silica in red/orange and bromide in grey

counterions). The anionic silica monomers tend to be found inside the layer of the cationic head groups, whereas the bromide ions tend to be located outside, allowing the two types of anions to work together more efficiently to shield the electrostatic repulsion between head groups, thus promoting larger surfactant micelles. Interestingly, when silica monomers were replaced by larger silica oligomers (while keeping the total charge constant), the latter tended to migrate out of the head group in order to interact preferentially with multiple micelles.

The approach of Jorge and co-workers suffered from several limitations due to the computationally intensive nature of the atomistic model. For example, formation of multiple micellar aggregates was only possible for the shorter DeTAB surfactant and not for the more commonly employed CTAB. More importantly, the authors were not able to observe significant shape changes of the micelles (e.g., sphere-to-rod transitions) in their simulations. In an attempt to circumvent these limitations, the same group developed a coarse-grained (CG) model for PMS precursor solutions [36]. The model of Pérez-Sánchez et al. was based on the popular MARTINI CG potential of Marrink and co-workers [246], but included new parameters for neutral and anionic silicic acid

monomers. The CTAB surfactant was described by four hydrophobic tail beads and one positively charged head bead [247], while the models for water and bromide followed the conventional MARTINI prescription. The surfactant model was validated by comparing micellar density profiles at the CG [36] and all-atom (AA) levels [248], and was shown to quantitatively reproduce the experimental average size and aggregation number of small CTAB micelles [36], as well as most features of the experimental CTAB/water phase diagram over a wide range of concentrations and temperatures [249]. The silica monomer parameters were calibrated by matching micellar density profiles against those obtained from AA simulations [241] under the same conditions.

Applying the CG mapping procedure allowed Pérez-Sánchez et al. to significantly extend the length and time scales available to MD models of the MCM-41 synthesis (simulations reached the order of tens of microseconds), while maintaining a good degree of chemical realism [36]. Simulations with silica monomers equilibrated to form long worm-like micelles, as seen in **Figure 3.14c**, while at the same conditions an aqueous solution of CTAB (i.e., sans silica) yielded only a distribution of spherical micelles. They did however observe a sphere-to-rod transition in the silica/surfactant solution that proceeded through the successive fusion of several smaller micelles to form large rods. This transition was driven by a strong adsorption of anionic silicates at the surface of micelles, screening the repulsive interaction between the cationic surfactant heads and promoting the formation of structures with a lower curvature. This was further evidence supporting the CT mechanism for the initial stages of MCM-41 synthesis.

In a later paper, Pérez-Sánchez et al. extended the model to consider silicate oligomers, as well as monomers, with parameters once more calibrated to match atomistic micellar density profiles [245]. They focussed their attention on low surfactant concentration solutions corresponding to the experimental studies of Firouzi et al. [213]. Initially, they tested several compositions, in order to determine the effect of different silica species on the final mesostructure. These were: (i) a binary surfactant/solvent reference solution; (ii) a ternary surfactant/silica monomer/solvent solution; (iii) a ternary surfactant/silica dimer/solvent solution. The silica/surfactant ratio was equal to unity in the latter two cases, i.e. there were half as many dimers in solution (iii) as monomers

in (ii), since monomers have a charge of -1 and dimers -2, hence keeping the total silica charge constant. Their results confirmed the previous assertion [36] that the reference solution produced spherical micelles, whereas the introduction of silica altered their geometry, making them much more elongated. Interestingly, however, the solutions containing monomers and dimers did not produce the same micellar geometry; solution (ii) produced worm-like micelles, whereas solution (iii) gave rise to a phase-separated hexagonal array, strongly reminiscent of the structure of MCM-41 (cf. **Figure 3.14d**). In order to show that these arrangements were equilibrium states, as opposed to kinetic artefacts, exchange simulations were carried out; monomers were replaced by dimers and vice versa, while maintaining the micellar geometry and allowing the system to relax. These simulations did indeed show that the changes were reversible and that the obtained structures therefore corresponded to the equilibrium states of the system.

The hexagonal array reached equilibrium much quicker in the exchange solution than starting from a random initial configuration. This is interesting, since the exchange solution is likely to represent a more realistic scenario; silica monomers initially produce rod-like micelles, then dimers resulting from silica condensation subsequently promote the formation of the hexagonal array. Further simulations with mixtures of monomers showed that formation of the hexagonal mesophase requires the presence of between 15 and 33% of silicates in the form of dimers. Their results thus not only demonstrate that silica is necessary to promote the formation of hexagonal structures, but that it needs to be in the form of (at least) small oligomers. This is mainly due to the ability of multiply charged silica oligomers to bind to more than one micelle at the same time, thus promoting micelle aggregation and eventually phase separation of the mesophase.

Next, Pérez-Sánchez et al. investigated larger oligomers, namely cyclic trimers, cyclic tetramers and cubic octamers. Their simulations showed reversible hexagonal-to-lamellar transitions by changing the silica charge density (either through increasing the degree of condensation or the degree of deprotonation for the same degree of condensation) or by adding a co-solvent (e.g, benzene) to the system. In the former case, a higher charge density reduced the effective area per head group, leading to a decreased curvature and the formation of lamellar phases. In the latter case, the co-solvent was

incorporated into the hydrophobic region of the hexagonal phase (i.e., inside the surfactant micelles), leading to an increase in the effective tail volume, and hence also promoting a transition to a lamellar structure. All of the above results were in qualitative agreement with experimental observations of the synthesis of MCM-41, but a key achievement was the ability to reproduce, for the first time, the formation of a hexagonal silica/surfactant mesophase under conditions that were quantitatively equivalent to the experiments of Firouzi et al. [213].

A very recent paper by the same authors applied the CG model to precursor solutions spanning a wide range of surfactant concentrations in order to probe the whole phase diagram of the system [249]. Once again, they compared the behaviour of a reference CTAB solution to that of solutions containing silica monomers and silica dimers. For the monomer/surfactant/solvent system, they found no surfactant concentration at which a hexagonal phase could be induced - in fact, the presence of silica seemed to inhibit the formation of a hexagonal phase, instead reverting it to worm-like micelles or lamellar phases. Replacing the monomers with dimers produced very different results - a hexagonal phase was formed over a much larger range than in the binary surfactant/solvent solution (mainly at the expense of the micellar-rod region). Interestingly, the results suggest that, even when starting from a pre-equilibrated hexagonal CTAB structure formed in a concentrated surfactant/solvent solution, the original surfactant liquid crystal is destroyed in the initial stages of the synthesis upon addition of silica monomers, only to be reformed later after some degree of silica condensation occurs, producing oligomers. These results are further evidence against the LCT mechanism, since they show that the nature of surfactant/solvent and silica/surfactant/solvent hexagonal mesophases is remarkably different - the former is driven by micellar repulsion (weak-screening limit) and fills the entire space, while the latter is driven by micellar attraction, i.e. strong-screening limit, and constitutes a phase-separated system.

The phase diagrams for the three systems (see **Figure 3.15**) also allowed the authors to hypothesise which of the MCM-like materials might be synthesised by the LCT mechanism. At high surfactant concentrations in the presence of silica, they observed the formation of a bicontinuous phase reminiscent of MCM-48 and of a lamellar phase

similar to MCM-50. Although bicontinuous phases were observed in all three tested solutions, their ranges of stability do not overlap at any surfactant concentration, hence MCM-48 would not be a candidate for the LCT mechanism. The exception, however, is MCM-50; in all three systems there is a region of overlap of lamellar phases at high surfactant concentrations. Taking all these observations together, it seems likely that the CT mechanism is the pathway by which the synthesis of MCM-41 is achieved.

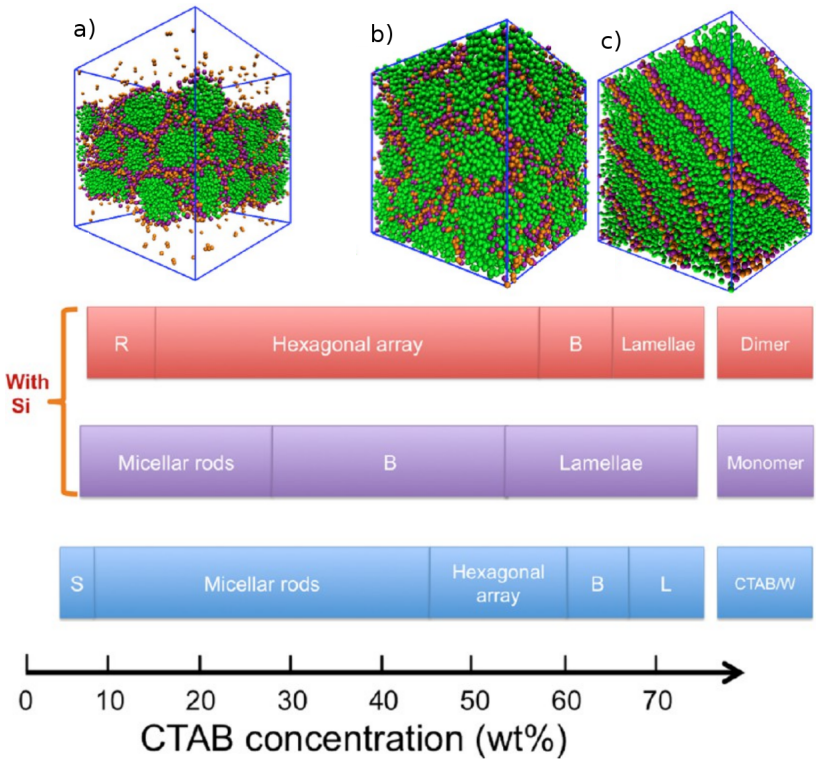


Figure 3.15: Map of the mesophases observed in coarse-grained simulations of different solutions at a range of initial surfactant concentrations, representative of incremental stages of PMS synthesis. From bottom to top: (blue) surfactant/water solution; (purple) silica monomer solution; (red) silica dimer solution. S stands for spherical micelles, R for micellar rods, B for bicontinuous and L for lamellar phases. Adapted from Chien et al. [249]

3.6.2 *In Situ* Functionalised Materials

Similarly to lattice models, atomistically detailed approaches for studying the synthesis of PMO (periodic mesoporous organosilica) materials have been built upon models of

MCM-41. They have been based on information obtained at the QM level [243, 250] in order to develop AA models for the synthesis solution [35, 241, 251] that are able to probe the molecular-level interactions between precursors and surfactants in systems with several thousand of molecules. Futamura et al. [250, 251] developed an AA model for hybrid inorganic-organic precursors $(HO)_3Si-R-Si(HO)_3$ with different organic linkers (R), namely planar organic rings (benzenesilica BENZ-Sil), as well as $C=C$ (ethylenesilica, ETHY-Sil) and $C-C$ bonds (ethanesilica, ETHA-Sil). Their aim was to study the role of the organic linker on PMO formation, elucidating experimental observations where different organosilica precursors induced structural differences in the final material [12]. The model was based on a combination of OPLS parameters for the organic groups [50] with silicate parameters from the group's previous work and new geometric and electrostatic parameters, specific to organosilicates, obtained from DFT calculations in the gas phase [250].

Futamura et al. [251] carried out MD simulations of three different aqueous solutions with DeTAB as the SDA and each of the different organic precursors under high pH conditions, where the majority of the silicate groups were anionic. Their simulations of up to 35 ns were able to capture the early synthesis stages, showing the initial formation of small surfactant aggregates, followed by micellar growth and fusion processes. The systems with BENZ-Sil and ETHY-Sil showed a measurable increase in the average micelle size with respect to the ETHA-Sil system. This was associated with a higher concentration of precursors adsorbed at the micelle interface for the former two cases, which supports the CT mechanism, as previously observed in the case of all-silica systems [241]. From computed micelle density profiles, the authors also showed that not only the inorganic silica but also the organic part of the precursors could be found between the surfactant heads in the micelle surface. More interestingly, they observed that while BENZ-Sil and ETHY-Sil were arranged parallel to the micelle surface, the ETHA-Sil species were perpendicular. This is an important fact because BENZ-Sil and ETHY-Sil are arranged more efficiently in the space between the surfactant heads, making it easier to form ordered structures. Furthermore, visual inspection of the BENZ-Sil system showed evidence of ordered arrangements of organosilicate precursors

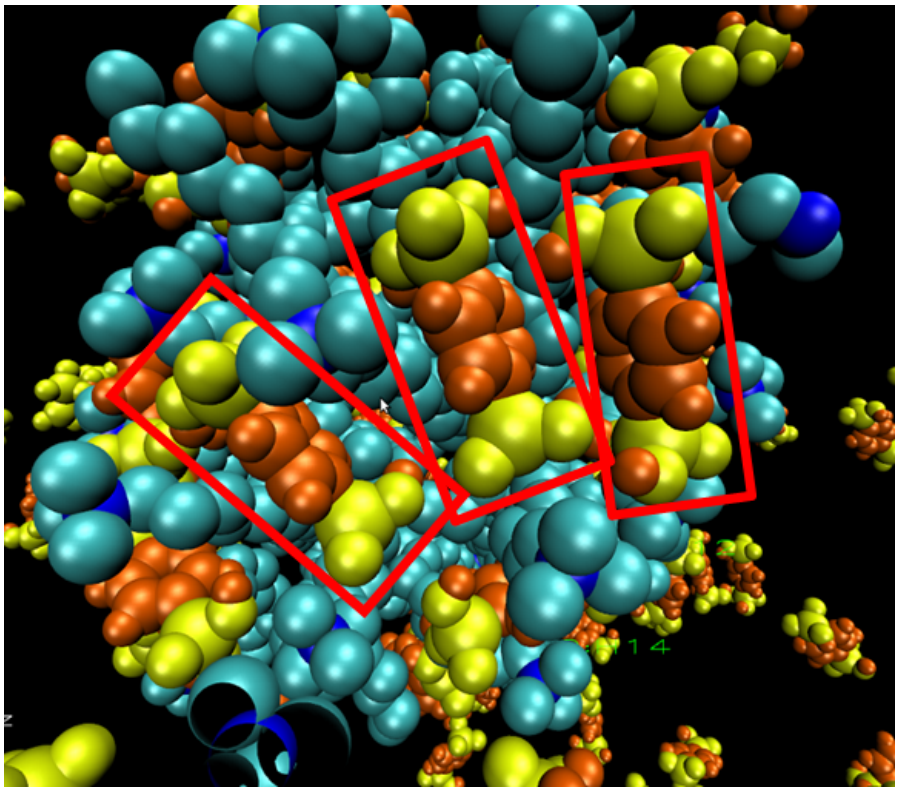


Figure 3.16: Snapshot of a spherical micelle formed in atomistic MD simulations of a precursor solution for a benzene-silica PMO. Colour scheme: light blue - surfactant carbon atoms; dark blue - surfactant nitrogens; orange - benzene groups; yellow - silicate groups. The red boxes highlight hybrid precursors that appear to stack in an ordered fashion at the micellar surface. Adapted from Futamura et al. [251]

at the micelle surface, possibly via π-π stacking between BENZ-Sil species (see **Figure 3.16**). It was suggested by the authors that this could represent the early stages of formation of the molecular-scale periodicity experimentally observed in the pore walls of benzene-silica materials [252]. Unfortunately, the high computational expense of the AA simulations prevented Futamura et al. from probing the later stages PMO synthesis, including the formation of organosilica/surfactant mesostructures.

3.7 Reactive Models - Modelling Silica Condensation

The studies discussed until this point have all focussed on the phase-equilibrium and self-assembly aspects of the PMS synthesis. To investigate these phenomena thoroughly

and efficiently, most studies have foregone investigation of the silica-polymerisation component of the synthesis - in fact, of all the studies discussed in the previous sections, only the work of Jin et al. [15] explicitly described silica condensation reactions, albeit with a rather simplified approach. Most attempts to describe silica condensation have made use of reactive potentials [253–255]. Such studies were able to provide atomic-level detail of the formation of small amorphous silicate clusters in solution, but required simulations to be run at very high temperatures in order to efficiently sample reaction events. The work of Wu and Deem [256] also used a reactive potential, but the configurational space of the system was explored through a rather complicated MC algorithm. Although their work provided important insight into silica nucleation, it was limited to small clusters due to the computationally intensive nature of the method. A very distinct approach is kinetic Monte Carlo (KMC) [257,258], which describes the chemical reactions as independent Poisson processes and thus significantly speeds up the sampling of reactive events (see **Section 2.2.8**). However, this method requires all reactions and rate constants to be explicitly enumerated, which in itself imposes a limitation on the size of oligomers that can be formed (the number of possible reactions increases factorially with cluster size). Models that describe the system energy as a function of bond topology only, such as the continuous random network (CRN) approach [259,260], have also been proposed. Finally, recent work by Malani et al. [95,151] has put forth a reactive Monte Carlo (RMC) model of silica polymerisation, which showed good agreement with experimental NMR measurements of the evolution of silicate distributions under neutral conditions. The RMC approach seems promising, since it is able to sample long time scales while maintaining a reasonable degree of atomistic detail.

None of the methods described above has yet been applied in the context of simulating PMS materials, with one notable exception being the work of Seaton and colleagues [261,262]. The first model which introduced explicit silica polymerisation reactions during simulations of MCM-41 formation was presented by Schumacher et al. [261]. They implemented a KMC reaction scheme, based on the CRN model of the silica network [260], where all silicon atoms are always connected to four oxygen atoms, and tried to describe the entire process of MCM-41 formation. They split their simulations

into four stages: (i) initial adsorption and condensation of a thin silica layer around the surfactant micelle; (ii) aggregation of silica-covered micelles to form a hexagonal mesostructure; (iii) further condensation of silica to form a highly cross-linked network; (iv) high-temperature calcination, causing further silica condensation in the absence of template micelles. To make the scheme computationally tractable, many major assumptions were made, such as describing the surfactant micelles as soft cylinders with no atomistic detail. Furthermore, several non-physical moves were applied to force the reaction along the chosen path (e.g., adding an artificial potential gradient to force the silica to the micelle in the opening stages, which is later removed, instead of correctly describing silica diffusion to the micelle surface). Due to all these special moves and assumptions, one cannot be completely confident that this model produced an accurate description of the real synthesis process. However, it achieved the authors' main objective, which was to generate realistic models of MCM-41 materials (see **Figure 3.17a**) for adsorption predictions (they reported fairly accurate gas adsorption results for nitrogen, ethane and carbon dioxide). More importantly, the work of Schumacher et al. demonstrated that by applying clever approaches, the computational modelling of the entire PMS synthesis is not an impossible task.

A later study by Ferreiro-Rangel et al. [262] proposed some improvements to the method of Schumacher and applied it to the synthesis of two less-studied materials, SBA-2 [263] and STAC-1 [264]. Both are synthesised with gemini quaternary ammonium surfactants and contain connected spherical pores templated by spherical micelles. However, they differ in the way that the micelles pack together - SBA-2 shows a hexagonal close packing of spheres, whereas STAC-1 micelles pack in a face-centred cubic lattice. In order to produce realistic models of these materials, Ferreiro-Rangel et al. [36] had to explicitly include water molecules during the reaction stages (which had been ignored by Schumacher et al.) and make the micelle-micelle interactions more repulsive so as to generate the correct packing. The simulation results were successfully compared against experimental data, including structural characteristics and nitrogen-adsorption isotherms. The authors concluded that the strong adsorption sites observed in adsorption isotherms were actually due to the surface roughness of the pores, and also

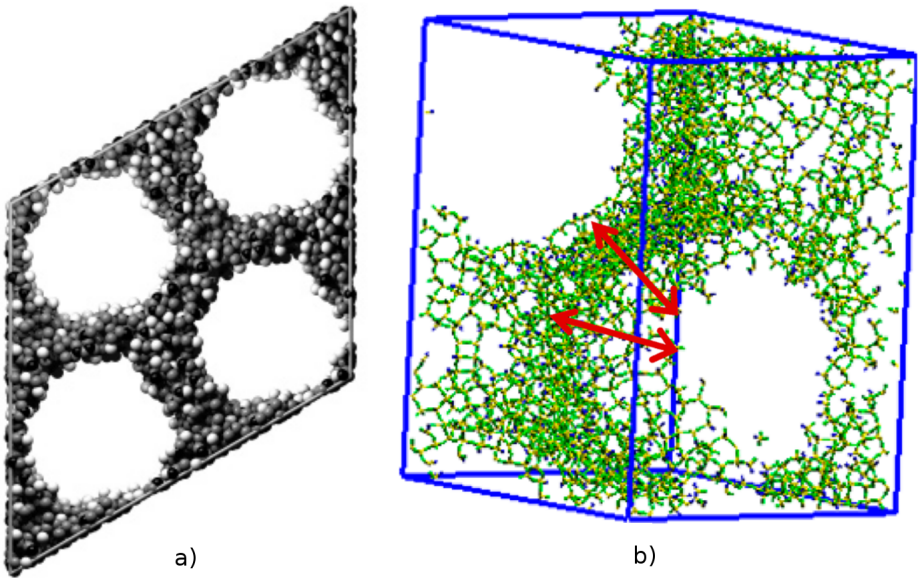


Figure 3.17: Examples of atomically detailed models of MCM-41 (a) and SBA-2 (b) materials obtained from kinetic Monte Carlo simulations of silica condensation. Only the silica framework is shown for the purpose of clarity. In panel (b), the red arrows highlight connection windows between spherical pores. Adapted from (a) Schumacher et al. [261] and Ferreiro-Rangel et al. (b) [262]

suggested a plausible hypothesis for the origin of the connecting windows between large spherical pores during the synthesis (see **Figure 3.17b**). Another interesting observation pertains to the evolution of the degree of silica condensation during different stages of the synthesis. The results of Ferreiro-Rangel et al. show that the transition between the formation of a silica layer at the micelle surface (the first step in the process) and the initial aggregation of micelles to form an ordered mesostructure (the second step) involved a significant presence of dimers and higher silica oligomers, with the concentration of monomers below 10%. This is in qualitative agreement with the interpretaion of Pérez Sánchez et al. [245] in their MD study of the self-assembly of silica/surfactant mesostructures.

3.8 Conclusions

In this chapter we have seen how a myriad of computational techniques have been used by various authors to probe several aspects of the liquid-phase synthesis of periodic mesoporous silica/organosilica materials, namely phase equilibria, self-assembly of SDAs and silica polymerisation. However, we have also observed how most of the studies surveyed here employed either very basic systems, which do not correspond to any real-life system, or very specialised variants that lack transferability. Drawing from these conclusions, we have decided to employ multiscale modelling, starting from a fundamental level, to begin parametrising a transferable force field for organosilica molecules, in the hope of making it easier to systematically test the synthesis of a wider range of materials. Because polarisation, which has been neglected in all of the computational studies described previously, is likely to play an important role in the highly polar and often charged PMS precursor solutions, we have decided to take a broader view of silica synthesis and investigate liquid- and solution-phase behaviour in general. For this reason, we will initially study the effect of polarisation in aqueous phases. With this goal in mind, we proceed to our first results chapter, where we will investigate how well current water models deal with polarisation effects that affect phase equilibria.

Chapter 4

Polarisation Corrections for Free Energy Calculations of Water

4.1 Introduction

In the previous chapter, the literature regarding the modelling of the synthesis of mesoporous silica was surveyed. As could be seen, the role of water, which, in all but the most highly-concentrated systems, will make up the majority of the molecules in the system, was essentially glossed over. Either off-the-shelf models were used or it was treated implicitly. In this section, we will thoroughly investigate how these water models deal with phase changes, since these processes are key for silica synthesis, which is largely defined by phase equilibria. Additionally, we will also demonstrate how their discrepancies can be linked to a key, but under-appreciated factor in fixed-charge models - polarisation.

Water is ubiquitous in chemical, biological, geological and industrial processes, making it the most extensively studied substance. It is also an extremely complex liquid, often demonstrating anomalous behaviour [68, 265–269]. This combination of importance and curiosity has sparked numerous computational studies of liquid water (see, for example, [270–273] and references therein). In particular, the solvation properties of water are of great interest for a vast range of applications, including the synthesis of PMS materials. The key thermodynamic property for describing solvation processes in

water is the Gibbs free energy of hydration (ΔG_{hyd}), and the calculation of ΔG_{hyd} for organic molecules is a huge area of research in computational chemistry [274–276]. In this work, we set out to perform a systematic test of several classical non-polarisable water models for their ability to predict the free energy of hydration of water, which is directly related to, but distinct from, its free energy of vaporisation [277]. In doing so, we uncovered what we believe to be a fundamental inconsistency in the way polarisation effects are currently handled in the calculation of phase-change energies and free energies. This prompted us to reassess the theoretical foundations of analytical polarisation corrections for the case of water hydration, with implications for future model development.

At the highest level of theory, water can be described using *ab initio* QM methods [48, 278–286], as was discussed in **Section 2.3**. However, despite continued increases in computational power, representing liquid water at an appropriate level of detail using a fully QM treatment remains extremely challenging [287]. As such, the vast majority of computational studies that involve water, usually as a solvent, rely on the classical approximation. Within classical models of water, two main approaches may be distinguished: polarisable (see **Section 2.2.4**) and non-polarisable models. The latter describe the electrostatic interactions of water through a fixed set of point charges or, more rarely, multipole moments, which cannot respond to changes in the environment surrounding the molecule (i.e., polarisation is not explicitly accounted for). Polarisable models, on the contrary, are able to respond on-the-fly to changes in the electrostatic environment by inclusion of inducible point dipoles, fluctuating charges or Drude oscillators [33]. This comes at a cost, however, and polarisable models can increase the computational cost by three to 10 times over that of similar fixed-charge pairwise-additive models [34]. In many applications of interest, such as protein folding [288–290], electrolyte solutions [291–293], and, of most relevance to this work, hydration free energy calculations [294–296], the majority of the system is occupied by water molecules; therefore, there is an inherent need to keep water models as computationally simple as possible.

The most popular non-polarisable models of water make use of the point-charge

approximation, differing only in the total number of interaction sites and in the values of the parameters. Early models like SPC [297] and TIP3P [298] were composed of three interaction sites, with point charges on the oxygen and hydrogen atoms, and a single LJ site on the oxygen. An alternative approach, initiated by the publication of the TIP4P model [299], displaced the negative charge along the bisector of the $H-O-H$ angle, creating an additional interaction site (see **Figure 4.1**). As such, a generic three-site model has five independent tunable parameters: the charge of the oxygen (q_O; since the molecule has to be neutral, the charge of the hydrogen atoms, q_H, is not independent), the well depth (ε) and van der Waals radius (σ) for the LJ potential, the $O-H$ bond length (L_{O-H}) and the $H-O-H$ angle (θ). The extra parameter for four-site models is the distance between the oxygen interaction centre and the interaction centre for the oxygen charge, the "M-site", (L_{O-M}). Nevertheless, these models still have many similarities; they have fixed point charges and their bonds and angle are rigid. Other alternatives have been proposed, such as five-site models that place two negative charges near the positions of lone-pair electrons on the oxygen atom [300]. However, the additional level of complexity does not appear to bring a significant improvement in overall performance [301], so only three-site and four-site models will be considered further.

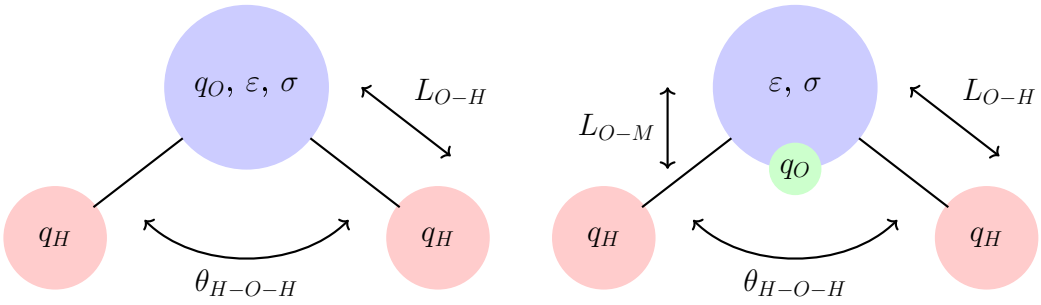


Figure 4.1: Schematic of three- and four-site water models

Regardless of the number and arrangement of the interaction sites, water models can be (rather arbitrarily) classified according to their parametrisation approach into three categories. “Generation-I” models (e.g., SPC, TIP3P and TIP4P) were parametrised to match a small number of structural and thermodynamic properties; typically, the radial

distribution functions (RDF), bulk density and enthalpy of vaporisation. The fact that they are still widely used today, particularly in biomolecular simulations [302], attests to their success in balancing computational speed with a good overall description of liquid water. Nevertheless, several shortcomings of these models have become apparent over the years and have led to several reparametrisation efforts.

A particularly important breakthrough in the context of this work, which will be discussed in more detail later, was proposed by Berendsen et al. [31]. They developed a simple expression to implicitly correct for polarisation effects in enthalpy of vaporisation calculations by approximating the energy cost of distorting the dipole moment of the water molecule from its equilibrium gas-phase value to a larger value more appropriate to the liquid phase. The SPC/E model developed by Berendsen et al. [31] constitutes a reparametrisation of the earlier SPC model by taking into account this polarisation correction. This led to improvements in the description of several properties, such as the self-diffusion coefficient. Another issue with earlier models is that they were parametrised without the use of special techniques like Ewald sums to account for long-range electrostatics. It is well known that neglecting long-range electrostatics can lead to significant artefacts, particularly for polar molecules [303–305], and thus Ewald sums and their derivatives [64] have become standard in modern molecular simulations. Indeed, later tests of Generation-I models revealed systematic deviations from experimental data when Ewald sums were employed in the calculations, which motivated researchers to reparametrise these earlier models by running simulations that made use of Ewald sums; examples thereof are TIP4PEw [306] and TIP4P/2005 [307]. These models also made use of the polarisation expression of Berendsen et al. [31] to correct the enthalpy of vaporisation. We will collectively describe "Generation-II" models as those that were still parametrised to match a restricted set of data, but have taken into account polarisation corrections for the enthalpy of vaporisation during the parametrisation process.

Finally, we will classify as "Generation-III" models those more recent efforts that considered a much larger set of data in the parametrisation process, including the static dielectric constant, and often use advanced parameter-fitting techniques [302,308–310].

These models also make use of the Berendsen expression to implicitly account for polarisation costs in the enthalpy of vaporisation. **Table 4.1** summarises all the models considered in this work, together with their parameters and classification.

Table 4.1: Parameters for all classical non-polarisable water models analysed in this work. Please refer to **Figure 4.1** for variable definitions

Model	Generation	σ (Å)	ε (kJ/mol)	q_H (e)	L_{O-H} (Å)	Θ_{H-O-H} (°)	L_{O-M} (Å)	μ (D)
SPC [297]	I	3.166	0.65	0.41	1.0	109.47	0	2.27
TIP3P [298]	I	3.15061	0.6364	0.417	0.9572	104.52	0	2.35
TIP4P [299]	I	3.15365	0.648	0.52	0.9572	104.52	0.15	2.18
SPC/E [31]	II	3.166	0.65	0.4238	1.0	109.47	0	2.35
TIP4P-Ew [306]	II	3.16435	0.690946	0.52422	0.9572	104.52	0.125	2.32
TIP4P-2005 [307]	II	3.1589	0.7749	0.5564	0.9572	104.52	0.1546	2.305
TIP4P-Ice [311]	II	3.1668	0.8822	0.5897	0.9572	104.52	0.1577	2.426
H2O-DC [308]	III	3.184	0.593	0.45495	0.958	109.47	0	2.42
TIP4P-FB [310]	III	3.1655	0.74928	0.52587	0.9572	104.52	0.10527	2.428
TIP4P-ε [309]	III	3.165	0.7732	0.527	0.9572	104.52	0.105	2.4345
OPC [302]	III	3.1666	0.8903	0.6791	0.8724	103.6	0.1594	2.48

A common feature of all the non-polarisable models described above is that they possess a fixed dipole moment (μ, cf. **Table 4.1**) that is significantly higher than the equilibrium gas-phase value (1.86 D [46]), but also significantly lower than the real dipole moment of water in the liquid phase. Indeed, the best current experimental estimate of the dipole of liquid water is 2.9 ± 0.6 D [47], while a large number of QM calculations [48, 280–283, 286, 312] and classical polarisable models [313–318] estimate values between 2.6 and 3.1 D. The "intermediate" dipole moment of non-polarisable models is a direct consequence of implicitly accounting for polarisation effects when trying to describe the potential energy surface (PES) of the liquid state. Indeed, recent efforts to develop consistent point charges for the liquid state typically yield values that are intermediate between gas-phase and (real) liquid-phase charges [142,319–321]. While this appears to be necessary to provide an adequate representation of the liquid-state PES, the issue of how to account for polarisation effects in properties that involve a change of state (such as the enthalpy of vaporisation and solvation free energies) remains unresolved.

As mentioned above, Berendsen et al. [31] were the first to propose an analytical expression for the energetic polarisation effect when moving a molecule from the gas to the liquid state. More specifically, they accounted for the internal self-polarisation

energy, which represents the energy difference between the equilibrium configuration of the molecule in the gas and liquid phases. Using this polarisation distortion correction, a molecular dynamics (MD) simulation to calculate the hydration free energy can now be represented as a thermodynamic cycle, as seen in **Figure 4.2**.

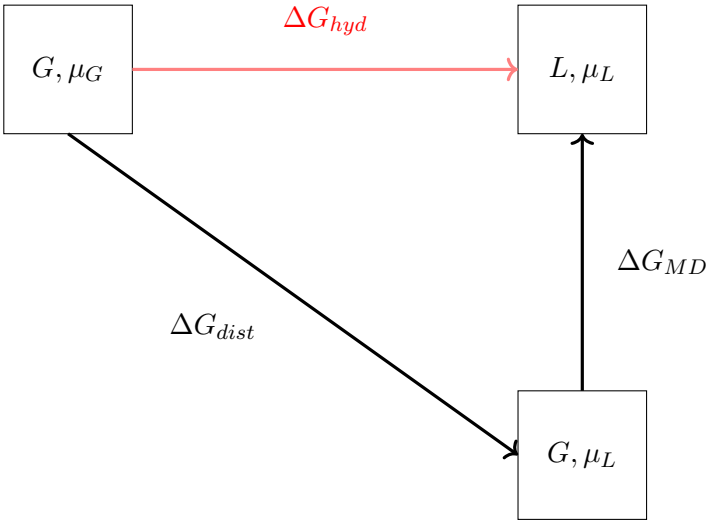


Figure 4.2: Thermodynamic cycle used in water models with polarisation corrections, such as SPC/E, for calculating the free energy of hydration. The red line represents experiment, whereas the black lines represent the various calculations which make up the simulation

This cycle relates results from simulations to those from experiment. In this case, the real process (experimental - ΔG_{hyd}) can be decomposed into two separate steps: i) while in the gas phase (G), the molecule's electronic distribution is polarised from its original state (represented here by the gas-phase dipole moment, μ_G) to a state that is representative of the bulk liquid (μ_L); ii) the polarised molecule is then surrounded by other molecules, leading to a fully liquid state (L). The first step corresponds to the cost of the intramolecular rearrangement of the molecule (ΔG_{dist}), while MD simulations using classical non-polarisable models yield the free energy due to the interactions with surrounding molecules (ΔG_{MD}). Berendsen et al. argued that the energy cost of rearranging the geometry of the molecule is dominated by the change in the dipole moment between the two phases. For a force field with fixed charges, this distortion correction takes the form [31]

$$\Delta E_{dist} = \frac{(\mu_L - \mu_G)^2}{2\alpha} \tag{4.1}$$

where μ_L and μ_G are the respective dipoles of the liquid and gaseous states, and α is one-third of the sum of the trace of the dipole-dipole polarisability tensor. Despite being formulated more than 30 years ago, this is still the most commonly used polarisation correction for non-polarisable water models in molecular simulations. It is in itself an approximation, since it is only the leading term of a summation over all multipoles, and its use implies the assumption that the leading dipole term will dominate. Swope et al. [322,323], however, have challenged this assumption and shown that, specifically for water, the difference between using only the dipole and an expansion up to quadrupoles is approximately 15% (0.4 kJ/mol in their case).

The work of Leontyev and Stuchebrukhov made two important contributions to polarisation corrections [30, 324]. Firstly, they offered a possible explanation as to why the polarisation corrections used in the parameterisation of water models may be incorrect [324]. They posited that, to correctly calculate the distortion correction, the actual experimental liquid-phase dipole moment should be used. Generally, in the then absence of reliable estimates for the dipole of the liquid, μ_L was estimated from the dipole moment of the model itself, which, as we have argued above, is much lower than the real value. Since the distortion correction scales with the square of the change in the dipole, this discrepancy has a large impact. For example, the SPC/E model has a dipole of 2.35 D [31], whereas the experimental value is around 2.9 D [47], and the gas-phase value is 1.86 D [46]. Using the model value, we would obtain a correction of 5.1 kJ/mol. However, application of the experimental value results in a correction of 23 kJ/mol. For comparison, ΔH_{vap} = 44 kJ/mol [270], so this discrepancy is approximately half the total enthalpy change of the process. Even worse, the free energy of self-solvation (hydration) of water is only -26 kJ/mol [325].

Leontyev and Stuchebrukhov also proposed a second polarisation correction [30], which is of opposite sign to the distortion correction, based on the electronic screening experienced by a solvated molecule. As already mentioned, charges on non-polarisable water models are usually lower than indicated by experimental and *ab initio* studies.

This is because classical molecular simulations cannot account for the response of the purely electronic degrees of freedom of the solvent. In practice, classical simulations assume a relative permittivity of unity, i.e. equal to vacuum, for calculating electrostatics using Coulomb's law. Under the Born-Oppenheimer approximation, however, it is more realistic to think of molecules (i.e., nuclei) as immersed in a bath of electrons [324]. This leads to a higher relative permittivity of the surrounding medium, thus effectively reducing the charges. Leontyev and Stuchebrukhov proposed that this effect can be accounted for by instead using the high-frequency dielectric constant of water, which describes the purely electronic response of the liquid phase, and thus managed to approximately reconcile the model dipole magnitudes with those of experimental studies through a simple scaling law [30]

$$q_{model} = \frac{q_{liquid}}{\sqrt{\epsilon_\infty}} \tag{4.2}$$

where q is the charge and ϵ_∞ is the high-frequency dielectric constant, which is commonly expressed as the square of the fluid's refractive index measured at the sodium D-line frequency [326].

Reducing the magnitude of the charges works for a single environment (e.g., a pure liquid) but is not valid when a change of said environment occurs; the model charges will not alter their magnitude to reflect the new level of screening and hence this will not be captured in the difference between two MD simulations (say, in the gas and liquid states). For this reason, Leontyev and Stuchebrukhov suggested a correction to try and quantify this difference analytically. They defined this as the energy of solvating a water molecule (in its liquid-phase configuration) in a purely electronic continuum, i.e. taking the molecule from vacuum (ϵ_0) to a solution of electrons (ϵ_∞). To this end, they applied a simple solvation model to quantify the correction. The approach of Leontyev and Stuchebrukhov can be seen as a thermodynamic cycle in **Figure 4.3**. It it similar to that shown in **Figure 4.2**, but now includes the extra step of solvating the polarised molecule in the purely electronic continuum (ΔG_{el}). Their results for water [30] showed that the two correction terms (distortion and electronic) almost cancelled each other out, leaving an overall correction of around 5 kJ/mol. This correction turns out to be

almost identical to that proposed by Berendsen et al. [31] in the development of SPC/E, as well as close to those of other simple water models. They then theorise that the good overall performance of these water models is due to cancellation of error, rather than to a rigorous treatment of polarisation corrections.

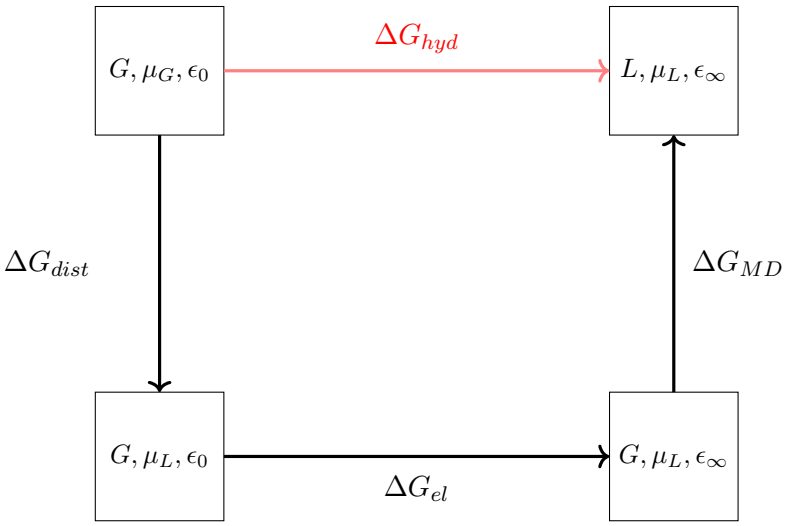


Figure 4.3: Thermodynamic cycle proposed by Leontyev and Stuchebrukhov [324] for calculating the hydration free energy of water. Red line indicates experimental route, while the black lines represent steps in calculation by simulation

In this work, we bring together the various improvements presented in the work of Swope et al. [322, 323] and Leontyev and Stuchebrukhov [324] into one unified correction scheme, and, as a result, attempt to answer the following questions: 1) what is the relative magnitude of the individual components of polarisation and to what extent do they cancel out?; 2) is it necessary to include higher-order multipoles in polarisation corrections or is simply using the leading dipole term sufficient?; 3) how sensitive is the magnitude of the polarisation correction to different values of the input variables (e.g., multipole moments, polarisabilities)? In particular, we have extended the method of Leontyev and Stuchebrukhov to calculate the electronic screening correction up to quadrupoles and combined it with the expressions of Swope et al. [322] for the distortion correction (also up to quadrupoles). After a detailed sensitivity study of these expressions, we have evaluated the significance of polarisation corrections in the calculation of free energies of hydration and heats of vaporisation using Molecular Dynamics (MD) for

the 11 water models shown in **Table 4.1**. We will end this chapter with some general conclusions and recommendations for future development of classical non-polarisable molecular models.

4.2 Theory and Methods

4.2.1 Molecular Dynamics Simulations

Hydration free energies (ΔG_{MD}) for each of the classical non-polarisable water models presented in **Table 4.1** were calculated using MD and thermodynamic integration (TI). This followed closely the procedure described previously [327,328] and the reader is referred to those publications for further technical details. MD simulations were carried out with GROMACS [44,329] using the leap-frog algorithm [330] with a time step of 2 fs. Simulations were performed in the NPT ensemble, with the temperature controlled at 298 K using a Langevin stochastic-dynamics thermostat [331] and the pressure controlled at 1.0 bar using the Parrinello-Rahman barostat [94]. A cut-off of 1.2 nm was used for LJ interactions, together with long-range dispersion corrections for energy and pressure. Long-range electrostatics were handled with the PME method [64], also using a cut-off of 1.2 nm for the direct sum. Water molecules were kept rigid using the SETTLE algorithm [332]. All simulations were carried out with at least 900 water molecules in cubic boxes with periodic boundaries. Further details about MD algorithms were presented earlier in **Chapter 2**.

Each system was first equilibrated for 5 ns, from which pure liquid properties were calculated after discarding the first 1 ns. In particular, the total potential energy of the liquid was sampled and used to calculate the enthalpy of vaporisation (ΔH_{vap}) of water according to

$$\Delta H_{vap} = E_{liq} - E_{vap} + RT - C_{QM} \tag{4.3}$$

In this equation, E_{liq} is the molar potential energy of the pure liquid, E_{vap} is the potential energy of the vapour (it is strictly zero for all the models considered here, as they contain no intramolecular contributions to the energy), R is the ideal gas constant,

T is the temperature and C_{QM} are quantum vibrational corrections. The value for C_{QM} (0.29 kJ/mol) at ambient temperature was taken from the work of Horn et al. [306] and was added to both ΔH_{vap} and ΔG_{MD}.

From the sampling period of the pure liquid simulations, we have extracted 50 evenly spaced configurations, which were used as starting points for TI calculations, as described previously [327, 328]. These 50 configurations hence served as an ensemble; ensemble simulations have been shown to be more precise and offer more robust uncertainty estimates [78, 79]. The usual procedure of decoupling solute-solvent interactions using a coupling constant (λ) was applied, with LJ and electrostatic interactions decoupled in two separate stages. For the LJ part of the free energy, 15 λ values were used - [0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1.0] - with λ = 0.0 corresponding to a fully interacting solute. For the electrostatic component, five λ points were used - [0.0, 0.25, 0.5, 0.75, 1.0]. The soft-core approach [81], with the same parameters as reported previously [327, 328], was employed in the LJ decoupling simulations in order to avoid instabilities in simulations close to $\lambda = 1.0$. Each starting configuration was run for 200 ps, with the gradient of the Hamiltonian being computed over the last 180 ps of each trajectory. This was then averaged over all starting configurations (i.e., over a total of 9 ns for each value of λ) and integrated over all λ points, according to

$$\Delta G_{MD} = \int_{\lambda=0}^{1} \langle \frac{\partial G(\lambda)}{\partial \lambda} \rangle_{\lambda} d\lambda \tag{4.4}$$

It has been shown [333] that integrating the Hamiltonian gradient over λ using the standard trapezoidal rule can lead to systematic errors in the free energy. In this work, we have used cubic splines [334] to obtain a more accurate hydration free energy from the individual simulation points. We calculated the standard error on the Hamiltonian gradient for each λ value by sampling over the the 50 individual configurations. This was then propagated for the solvation free energy and used to calculate the 95% confidence interval, which we report as error bars in our data tables and plots. All the averaging and integration was carried out using GROMACS utilities and in-house scripts.

In order to compare vaporisation enthalpies and hydration free energies with exper-

imental data, we applied the thermodynamic cycle depicted in **Figure 4.3**. The next two sections describe how the two polarisation contributions were estimated.

4.2.2 Multipole Distortion Correction

Swope et al. [322,323] proposed a set of equations that expands the distortion polarisation correction up to quadrupoles. In the process, they tested whether the usual approximation of truncating the series at the leading dipole term was an acceptable simplification. They based their correction on a representation of the molecular electron density in terms of a multipole expansion. The full derivation can be found in their first paper from 2010 [322] and therefore only the key concepts will be elucidated here. In Einstein notation, their correction is

$$E_{dist} = \frac{1}{2}\alpha_{\alpha,\beta}V_\alpha V_\beta + \frac{1}{3}A_{\alpha,\beta\gamma}V_\alpha V_{\beta\gamma} + \frac{1}{6}C_{\alpha\beta,\gamma\delta}V_{\alpha\beta}V_{\gamma\delta} + ... \tag{4.5}$$

where V are the relevant derivatives of the external electric potential, whereas α, A and C are the dipole-dipole, dipole-quadrupole and quadrupole-quadrupole polarisabilities, respectively. The multipoles are linked to the electric-field vectors and polarisabilities by

$$\mu_\alpha - \mu_\alpha^0 = -\alpha_{\alpha\beta}V_\beta - \frac{1}{3}A_{\alpha,\beta\gamma}V_{\beta\gamma} \tag{4.6}$$

$$\Theta_{\alpha\beta} - \Theta_{\alpha\beta}^0 = -A_{\alpha\beta,\gamma}V_\gamma - C_{\alpha\beta,\gamma\delta}V_{\gamma\delta} \tag{4.7}$$

where Θ represents the quadrupole moments as a column vector. These can then be represented in matrix form as

$$\begin{pmatrix} 3(\mu - \mu^0) \\ \Theta - \Theta^0 \end{pmatrix} = -\begin{pmatrix} 3\alpha & A \\ A^t & C \end{pmatrix}\begin{pmatrix} \nabla V \\ \nabla\nabla V \end{pmatrix} \tag{4.8}$$

Hence, knowing the change of the multipoles upon entering the liquid phase and the polarisability tensor, one can solve for the electric-field vectors and then obtain the distortion free energy change of polarisation. However, a final step is necessary for the

application of this set of equations. Although there are an equal number of equations and unknowns (12), several are dependent due to symmetry. Swope et al. [322] applied a spherical coordinate system to make this system of equations solvable, reducing the number of both equations and unknowns to eight.

The above equations provide an estimate of the distortion energy term. In order to apply them for the estimation of *free* energies, Swope et al. [323] made the argument that

$$\Delta G_{dist} \approx E_{dist} \tag{4.9}$$

Hence,

$$\Delta G_{dist} = \frac{1}{2}\alpha_{\alpha\beta}V_{\alpha}V_{\beta} + \frac{1}{3}A_{\alpha,\beta\gamma}V_{\alpha}V_{\beta\gamma} + \frac{1}{6}C_{\alpha\beta,\gamma\delta}V_{\alpha\beta}V_{\gamma\delta} + ... \tag{4.10}$$

This assumption is accompanied by a caveat, namely that the "polarisation cost is approximately constant over the range of thermally accessible conformations on either: 1) the unpolarised quantum-chemical potential surface; 2) the polarised fixed-charge potential surface" [323]. Swope et al. thoroughly analysed this assumption in Appendix B of their paper and found, at least for their chosen set of molecules/side chains, that the assumption held well. In the least successful case, acetamide, this was attributed to the high restructuring free energy cost of that particular molecule. Since the focus of our study, water, does not have a complicated geometry, and **Equation 4.9** was valid for the most similar molecule in their test set (methanol), we do not foresee any difficulties with this assumption.

Solving the above set of equations therefore requires three inputs: the gas-phase multipoles, the liquid-phase multipoles and the polarisability tensor. The former can be taken from experiment or high-level QM calculations. Assuming that the polarisabilities are the same in the gas and liquid phases allows one to also obtain the polarisability tensor from *ab initio* calculations. This leaves the liquid-phase multipoles to be determined, which is a well-known challenge [279]. In their first paper, Swope et al. [322] validated their expressions by comparing distortion energies obtained directly from QM

calculations in a dielectric continuum model with the results of their analytical treatment and found good agreement between the two. We have tested the precision of our implementation of the distortion-correction equations by successfully replicating the calculation for water presented in their paper, starting from the original data kindly provided by one of the authors [322].

In their subsequent paper [323], Swope et al. used their methodology to compare solvation free energies from corrected classical MD simulations with experimental data. In doing so, and similarly to Berendsen et al. [31], they used the effective multipole moments of the classical models as an estimate of the liquid-phase moments. As discussed above, this can make a significant difference to the magnitude of the positive correction. Therefore, we believe that a re-evaluation of this method with more realistic liquid multipole moments is in order.

4.2.3 Electronic Polarisation Correction

The previous two sections described how we accounted for the free energy contributions due to intermolecular (classical, non-polarisable) interactions with the surrounding medium (via MD and TI) and due to intramolecular polarisation effects (via the distortion correction). The missing term, as postulated by Leontyev and Stuchebrukhov [30,324], is the intermolecular nuclear-electron interactions, i.e. those between the purely electronic continuum and the molecule being solvated. Since they defined this as the free energy of solvating a water molecule in a solution of electrons, they applied continuum solvation models to quantify this contribution. More precisely, the simplest solvation model, the Born model [335], was used because it allows for an analytical solution. The Born model provides an expression for the free energy of solvation of an ion in a spherical cavity surrounded by a uniform and infinite dielectric continuum. Kirkwood [336,337] later showed that the Born model was the leading term in an expansion over all multipoles

$$\Delta G_{solv} = -\frac{1}{2}\sum_{l=0}^{\infty}\sum_{m=l}^{l}\left(\frac{\epsilon-1}{\epsilon+\frac{l}{l+1}}\right)\frac{|M_{lm}|^2}{R^{2l+1}} \tag{4.11}$$

where M is any multipole with order l and component m, R is the radius of the cavity and ϵ is the static dielectric constant of the continuum. The static dielectric constant describes the response of the fluid to an electric field in the limit of zero frequency (i.e., a static electric field). As such, both nuclear and electronic responses are included in this quantity. However, in our case we are not interested in the nuclear response of the solvent, but rather in the purely electronic response due to the presence of a solute molecule. Since the nuclei are much more massive than the electrons, they are much slower and therefore cannot respond to the field when the frequency is very high. As such, the relevant quantity for our purposes is the infinite-frequency dielectric constant (ϵ_∞), which captures only the response of the electronic degrees of freedom of the solvent. Thus our equation becomes

$$\Delta G_{el} = -\frac{1}{2}\sum_{l=0}^{\infty}\sum_{m=l}^{l}\left(\frac{\epsilon_\infty - 1}{\epsilon_\infty + \frac{l}{l+1}}\right)\frac{|M_{lm}|^2}{R^{2l+1}} \tag{4.12}$$

Leontyev and Stuchebrukhov [30] only included the dipole term in their work, i.e. l = 1, but here we will extend this treatment to also include quadrupole contributions.

One problem with the Born/Kirkwood expression is that the term R (the cavity radius) is not strictly defined, as well as being assumed to be spherical. The choice of cavity radius is largely arbitrary, with three common choices being: i) the radius of a sphere of volume equal to the molecular volume of the pure liquid; ii) the molecular length plus the Van der Waals radii of the outermost atoms; iii) adding a constant term (0.5 Å) to the value obtained from the molar volume of the pure liquid [338]. Additionally, some authors, such as Luo et al. [338], have attempted to find a unique theoretical determination of the cavity radius. However, this radius is only unique for a given set of multipoles, i.e. a specific QM calculation. In general, consideration of multipoles of higher magnitudes lead to smaller cavity radii.

One way to introduce a degree of self-consistency is to use a modified version of the Onsager reaction field model [339] developed by Barker and Watts [340]. Usually, this is used to make the Born model self-consistent by replacing the liquid-phase dipole with the gas-phase dipole plus a distortion of this dipole induced by an electric field.

However, in our case (similarly to Leontyev and Stuchebrukhov [30]), we want to use the reaction field to make the cavity radius (R) self-consistent. The two necessary equations are

$$\mu = \mu_0 + \alpha F \tag{4.13}$$

$$F = \frac{\epsilon - 1}{\epsilon + \frac{1}{2}} \frac{\mu}{R^3} \tag{4.14}$$

The first merely expresses the dipole in solution (μ) as a function of the permanent (i.e., gas phase) dipole (μ_0) plus a distortion through an electric field (F) multiplied by the isotropic polarisability (α). The second is the reaction-field equation. Equating the two fields gives us

$$F = \frac{\epsilon - 1}{\epsilon + \frac{1}{2}} \frac{\mu}{R^3} = \frac{\mu - \mu_0}{\alpha} \tag{4.15}$$

Then we arrange for R

$$R = \left(\frac{\epsilon - 1}{\epsilon + \frac{1}{2}} \frac{\mu\alpha}{\mu - \mu_0} \right)^{\frac{1}{3}} \tag{4.16}$$

This way, the cavity radius is an analytical function of its environment. For simplicity, this equation only includes the dipole but could potentially be expanded to obtain an estimate of the cavity radius that includes quadrupoles. A final point has to be made here about the choice of dielectric constant in **Equation 4.16**. In the case of the calculation of the cavity radius, the static dielectric constant of the fluid is used in order to provide an appropriate estimate of the electric field caused by the surrounding solvent.

4.2.4 Analysis of AIMD Trajectories

As part of this study, Schienben and Marx generously provided us with the trajectory of their recent AIMD study of water [48] and from this we were able to calculate the multipole moments of liquid water. In this section, we will briefly elucidate how the

trajectory was analysed.

The output files provided by Schienbein und Marx consisted of coordinates of all atoms (given as H and O in the trajectory output files) and maximally localised Wannier functions, where two electrons of an orbital are localised at a point in space, (X), from each frame of their trajectory. These were analysed using an in-house Python script. Although the dipole is origin independent, the traceless quadrupole is not. For this reason the coordinates of all water molecules were altered so that the oxygen atom lay at the origin. Then all water molecules were rotated so that they lay along the same plane. Subsequently the dipoles were calculated as:

$$\mu_i = \sum_l r_{i,l} q_l \tag{4.17}$$

where i refers to the coordinate of the atom along the i-axis and l refers to the atom in the molecule. The overall dipole was then calculated as

$$\mu = \sqrt{\mu_x^2 + \mu_y^2 + \mu_z^2} \tag{4.18}$$

For quadrupoles, this was

$$\Theta_{ij} = \sum_l q_l (3 r_{i,l} r_{j,l} - ||\mathbf{r_l}||^2 \delta_{ij}) \tag{4.19}$$

where i and j are refer to the coordinates of atom l along the i- and j-axes, respectively, $\delta_i j$ is the Kronecker delta and $||\mathbf{r_l}||$ is the magnitude of the vector $\mathbf{r_l}$. The charges (q_l) of the ion centres (O and H) correspond to the number of outer electrons of each but positive, since these are treated as nuclei; q_O is hence +6e and q_H is +1e. Each Wannier function has a charge of -2e since it corresponds to the coordinate of a localised orbital.

Since the calculations of Schienbein and Marx were carried out at the QM level, the angles and bond lengths of the individual water molecules were not constrained. This meant that some were not symmetrical, leading to non-diagonal elements of the 3×3 quadrupole matrix being non-zero. In order to make all molecules comparable the quadrupole matrices were diagonalised so that all off-diagonal elements were equal to

zero. The effect of this was minor (less than a percent difference in the quadrupole magnitudes) and was done purely for the sake of consistency.

The major (Θ_2) and minor (Θ_0) quadrupoles were then computed as

$$\Theta_2 = \Theta_{zz} + \frac{1}{2}\Theta_{xx} \tag{4.20}$$

$$\Theta_0 = \Theta_{xx} \tag{4.21}$$

μ, Θ_0, Θ_2 were calculated for every molecule in every frame of the trajectory, 12,928 in total. These were then summed and divided by the total number of molecules (12,928) to give the average value of each property. These were then binned to provide a distribution of each property. Similarly, for every molecule the necessary polarisation correction was calculated, summed and averaged in order to give the average polarisation correction per molecule.

4.3 Results and Discussion

Our first objective is to assess the performance of classical non-polarisable water models in reproducing the free energy of hydration and the enthalpy of vaporisation. Clearly, a consistent model should be able to predict both of these properties to a good degree of accuracy. The results of our MD calculations for all the water models in **Table 4.1** are reported in **Table 4.2**.

Figure 4.4 shows the deviations between MD results and experimental data for each of those properties, plotted against each other with horizontal and vertical error bars. In this plot, a model showing perfect agreement with experiment would be located at the origin. To obtain the red points in **Figure 4.4**, we have included the distortion correction proposed by Berendsen at al. [31], i.e. **Equation 4.1**, using the dipole moments of each model as a proxy for the liquid-phase dipole, in the calculation of both ΔH_{vap} and ΔG_{MD}. This corresponds to the state-of-the-art in the development of non-polarisable water models [302, 308–310]. It is worth noting that, by convention, ΔH_{vap} is positive, as it is measured as a change from liquid to vapour (hence it implies an increase in energy), while ΔG_{hyd} is negative, going from the vapour to the liquid.

Table 4.2: Calculated heat of vaporisation and free energy of self-solvation data for various water models with two different polarisation corrections. All energies are given in units of kJ/mol

Model	Berendsen Correction	ΔH_{vap}		ΔG_{hyd}	
		None	Berendsen	None	Berendsen
SPC	3.58	43.9	40.3	-25.8	-22.3
TIP3P	5.11	42.3	37.1	-25.3	-20.2
TIP4P	2.18	43.6	41.4	-25.2	-23.0
SPCE	5.11	49.0	43.8	-29.5	-24.4
TIP4P-Ew	4.51	48.6	44.1	-28.9	-24.4
TIP4-P2005	4.22	50.0	45.7	-29.8	-25.5
TIP4P-Ice	6.82	56.2	49.3	-34.6	-27.8
H2ODC	6.68	50.4	43.7	-30.4	-23.7
TIP4P-FB	6.87	51.4	44.6	-31.7	-24.8
TIP4P-ε	7.03	51.9	44.9	-31.7	-24.7
OPC	8.19	53.6	45.4	-33.2	-25.0

Therefore, a model that yields a liquid-phase potential energy that is, say, too negative, will overestimate ΔH_{vap} and likely underestimate ΔG_{hyd}. For the same reason, when the distortion correction is added to ΔH_{vap}, it takes a negative sign, while it is positive when added to ΔG_{hyd}.

As expected when the Berendsen correction is applied (red line), the Generation-I models perform rather poorly, significantly overpredicting ΔG_{hyd} and underpredicting ΔH_{vap}. This is mainly because, although ΔH_{vap} was a target in the parametrisation of those models, the distortion correction was not accounted for in this process. Conversely, models of Generations II and III predict ΔH_{vap} rather well (the exception is TIP4P-Ice, which was not designed to do so). However, all of these models (again with the exception of TIP4P-Ice) systematically overestimate the hydration free energy, albeit not by a very large amount. What is most striking in this analysis, however, is that the performance of all the models falls on a straight line ($R^2 = 0.9905$) that does not pass through the origin. This suggests that, using the current approach for force field development, no classical non-polarisable water model will be able to simultaneously predict the enthalpy of vaporisation and the free energy of water hydration. In fact, the intercepts of the linear fit shown in **Figure 4.4** tell us that the models that give the best performance for ΔH_{vap} overestimate ΔG_{hyd} by about 2 kJ/mol, while a hypothetical model that would be tuned

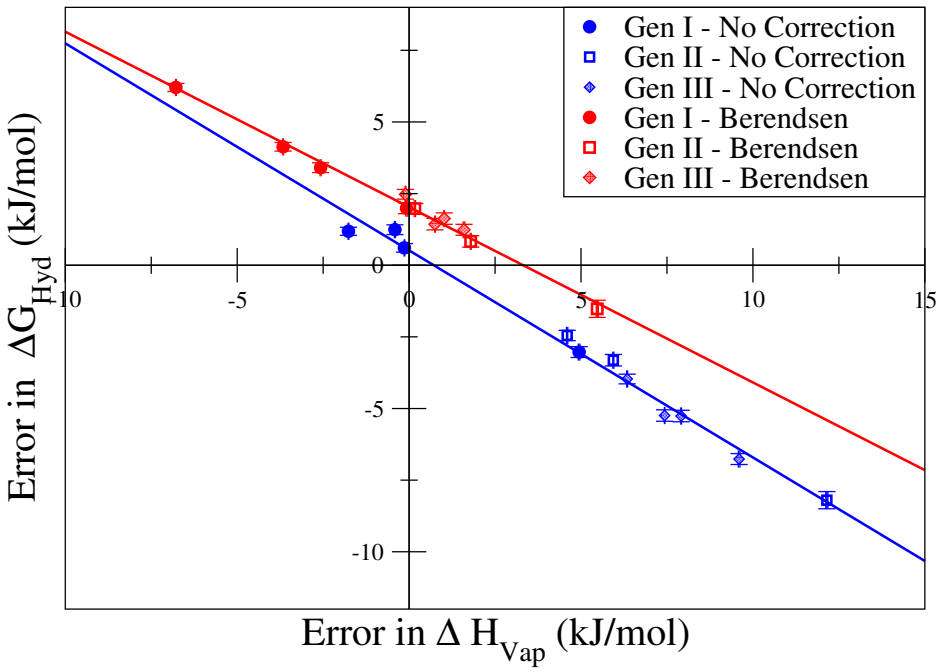


Figure 4.4: Deviations between simulation and experiment for the hydration free energy and enthalpy of vaporisation of 11 different non-polarisable water models (as shown in **Table 4.1**). Generation-I models are represented by solid circles, Generation-II ones by open squares and Generation-III models by shaded diamonds. The different colours represent different approaches to polarisation corrections: no correction (blue points), Berendsen's correction (red points)

to match ΔG_{hyd} would overestimate ΔH_{vap} by about 3 kJ/mol. Although these are not very large values, they are much larger than the uncertainty of the calculations (about $\pm$ 0.2 kJ/mol). Of course, our analysis is not exhaustive, and it is always possible to argue that a model could be found within the current paradigm that could fit both properties simultaneously. However, we argue that the analysis of **Figure 4.4** is compelling enough to warrant a rethink of the way polarisation corrections are currently handled in the parameterisation of classical non-polarisable models.

Further insight is obtained by analysing the blue points in **Figure 4.4**. These were obtained by comparing simulations to experiments without applying any polarisation corrections to either property. This corresponds to the paradigm used in earlier model development (prior to 1987) and unsurprisingly leads to good performance for the Generation-I models in predictions of ΔH_{vap} (and correspondingly poor perfor-

mance for the more recent models). What is perhaps surprising here is that a linear fit leads to an intercept very close to zero (0.51 kJ/mol). In particular, the SPC model is able to predict both energetic properties rather satisfactorily. Of course, it is now well known that these Generation-I models present serious deficiencies in their description of other structural and thermodynamic properties of water (e.g., temperature of maximum density, phase diagrams, diffusion coefficient) [272]. This leaves us with a paradox - newer models are now able to provide a very good description of most water properties [302,307,310], but lead to systematic deviations in the hydration free energy; older models are able to predict the latter quite well, but lead to a generally poor description of liquid water. In this context, it is interesting to note that despite decades of advances in water-model development, the majority of simulations of biomolecular systems and of solvation free energy calculations still make use of rather outdated Generation-I models, such as SPC or TIP3P [302].

To address this apparent paradox, we delve deeper into the issue of polarisation corrections. As discussed in the methodology, application of more rigorous procedures [30, 322–324] requires data on the polarisability tensors of water in the gas phase, as well as estimates of the multipole moments of water in the liquid phase (at least up to quadrupoles). The latter, in particular, are subject to a great degree of uncertainty, due to the well-known difficulty in separating the contributions of each molecule in a QM simulation of a condensed phase [279]. Indeed, there are relatively few QM studies that report both molecular dipoles and quadrupoles for liquid water; most of these have been collected by Niu et al. in their recent paper [341]. Due to the simple geometry of water, all of its constituent atoms can be placed along one plane, with the hydrogens placed symmetrically across the bisector. Since water has one dipole, this means that only one component of the dipole vector will be non-zero. In our representation, the dipole is oriented along the z-axis. For the quadrupoles (we use the traceless version, in keeping with the work of Swope et al.), this means that only the diagonal components will be non-zero, i.e. three components. However, since the trace (the sum of the diagonal) must be equal to zero, this leaves us with only two independent quadrupolar components; thus three independent parameters over both

dipoles and quadrupoles. The QM data collected by Niu et al. are reproduced in **Table 4.3**, together with the experimental values for water in the gas phase. We also include results calculated by applying Wannier functions to individual water molecules in the trajectories of AIMD simulations of Schienbein and Marx [48]. The procedure for extracting these results was described in **Section 4.2.4**. The results shown in **Table 4.3** for the liquid multipoles can be split into two groups: hybrid quantum mechanics/molecular mechanics (QM/MM) studies (entries 2—4) and AIMD (entries 1, 5 and 6).

Table 4.3: Dipole and quadrupole moments of water from various quantum-mechanical calculations

Calculation	μ (D)	Θ_0 (D.Å)	Θ_2 (D.Å)
Gas - Experimental [46]	1.86	0.11	2.57
Liquid - AIMD (2) [342]	2.43	0.10	2.72
Liquid - QM/4MM [341]	2.49	0.13	2.93
Liquid - QM/230TIP5P [343]	2.55	0.20	2.81
Liquid - QM/4TIP5P [341]	2.69	0.26	2.95
Liquid - AIMD (3) [48]	2.88	0.30	3.32
Liquid - AIMD (1) [280]	2.95	0.18	3.27

A significant degree of variability can be seen in the multipole moments for the liquid, with dipole moments between 2.43 and 2.95 D. Although there is still some debate in the literature about the magnitude of the dipole moment of liquid water, evidence from a variety of theoretical approaches is converging upon values between 2.8 and 3.0 D. The early AIMD studies of Parrinello and co-workers [279,280] made use of maximally localised Wannier functions to isolate the contributions of each molecule in the liquid, leading to dipole moments in good agreement with estimates obtained from experimental X-ray scattering [47] and from the experimental refractive index using a mean-field approach [344]. Although it was subsequently shown that the non-local exchange-correlation functionals used in early AIMD studies lead to an overstructured liquid phase with incorrect density [287], more recent simulations using hybrid and dispersion-corrected functionals yield quite similar values for the dipole moment [48, 281,283,286]. Dipole moments around 3.0 D were also obtained from MP2 calculations on large water clusters [312] and from MD simulations using an *ab initio* dipole moment

surface [345]. Given this mounting evidence, we consider that our calculations based on the very recent and highly accurate simulations of Schienbein and Marx [48] represent the best currently available estimates of the multipole moments of liquid water.

Since the full polarisability tensor cannot be obtained directly from experiment, the results of QM studies were used. In particular, we have gathered data from three separate studies where relatively high levels of theory were employed [346–348]. For completeness, we have also considered the data set used by Swope et al. [322], kindly provided by one of the authors, even though it made use of a comparatively low level of theory. All the non-zero values of the polarisability matrices used in this work can be found in **Table 4.4**. Overall, we have tested six different polarisability matrices, calculated using three different computational approaches (DFT, MP2 and Coupled Cluster). Together with the six different sets of multipole moments, this results in 36 individual data sets.

Table 4.4: All non-zero components of the polarisation tensors used in this work **given in atomic units**. α = dipole-dipole polarisability matrix, A = dipole-quadrupole and C = quadrupole-quadrupole. * = d-aug-cc-pVQZ = [7s6p5d4f3g/6s5p4d3], † = cc-pv(t+d)z, ‡ = [5s3p2d/3s2p], ◊ = d-aug-cc-pVQZ = [7s6p5d4f3g/6s5p4d3]

Component	Polarisability Tensor					
	B3LYP†	B3LYP*	MP2*	CCSD*	CCSD(T)‡	CCSD(T)◊
$\alpha_{x,x}$	5.517	9.769	9.488	8.976	9.70	9.25
$\alpha_{y,y}$	8.510	10.204	9.942	9.712	10.06	9.874
$\alpha_{z,z}$	7.285	9.965	9.691	9.310	9.71	9.529
$A_{x,xz}$	-1.42	-1.12	-1.08	-0.88	0.98	-1.04
$A_{y,yz}$	-6.50	-5.47	-4.96	-5.01	-2.08	-5.11
$A_{z,xx}$	3.93	2.96	2.65	2.65	3.64	2.72
$A_{z,yy}$	-3.57	-2.06	-1.79	-2.00	-3.25	-1.93
$A_{z,zz}$	0.36	-0.89	-0.86	-0.65	-0.39	-0.79
$C_{xx,xx}$	5.97	18.23	16.27	15.18	13.73	15.65
$C_{xx,yy}$	-4.63	-9.26	-8.31	-7.87	-8.44	-8.09
$C_{xx,zz}$	-1.34	-8.96	-7.96	-7.30	-5.29	-7.57
$C_{yy,yy}$	8.75	16.55	15.03	14.29	13.58	14.67
$C_{yy,zz}$	-4.12	-7.29	-6.72	-6.42	-5.14	-6.59
$C_{zz,zz}$	5.46	16.26	14.68	13.72	10.43	14.15
$C_{xy,xy}$	4.63	12.84	11.51	10.73	6.90	11.10
$C_{xz,xz}$	3.20	12.97	11.59	10.75	5.99	11.13
$C_{yz,yz}$	7.82	13.96	12.56	12.03	8.02	12.34

First of all, we considered the effect of the method and basis-set size for calculating the polarisability tensor of water. The overall polarisation corrections for all multipole sets and polarisation tensors are shown in **Table 4.7**, while the corresponding individual components for the distortion and electronic corrections are shown in **Tables 4.5** and **4.6**.

Table 4.5: Distortion correction for different sets of multipoles and polarisability matrices at various levels of theory. * = d-aug-cc-pVQZ = [7s6p5d4f3g/6s5p4d3], † = cc-pv(t+d)z, ‡ = [5s3p2d/3s2p], ◊ = d-aug-cc-pVQZ = [7s6p5d4f3g/6s5p4d3]

Calculation	Multipole Distortion Correction (kJ/mol)					
	B3LYP †	B3LYP *	MP2 *	CCSD *	CCSD(T) ‡	CCSD(T) ◊
AIMD (2)	13.4	7.8	8.2	8.3	8.6	8.2
QM/4MM	23.9	12.0	12.6	12.9	13.5	12.8
QM/230TIP5P	21.7	11.8	12.4	12.7	13.1	12.5
QM/4TIP5P	36.7	18.5	19.5	20.3	21.0	19.7
AIMD (3)	80.2	35.9	38.1	40.4	42.5	38.8
AIMD (1)	78.6	37.9	40.1	41.3	43.4	40.7

Table 4.6: Electronic polarisation correction for different sets of multipoles and polarisability matrices at various levels of theory: * = d-aug-cc-pVQZ = [7s6p5d4f3g/6s5p4d3], † = cc-pv(t+d)z, ‡ = [5s3p2d/3s2p], ◊ = d-aug-cc-pVQZ = [7s6p5d4f3g/6s5p4d3]

Calculation	Electronic Solvation Correction (kJ/mol)					
	B3LYP †	B3LYP *	MP2 *	CCSD *	CCSD(T) ‡	CCSD(T) ◊
AIMD(2)	-25.6	-16.5	-17.2	-18.0	-17.0	-17.4
QM/4MM	-31.1	-19.9	-20.8	-21.7	-20.5	-21.1
QM/230TIP5P	-33.5	-21.5	-22.5	-23.5	-22.2	-22.8
QM/4TIP5P	-44.2	-28.3	-29.6	-30.9	-29.1	-29.9
AIMD (3)	-64.6	-40.8	-42.7	-44.7	-42.1	-43.3
AIMD (1)	-68.4	-43.4	-45.4	-48.1	-44.7	-46.0

Studying these results, it is clear that the distortion and electronic corrections largely cancel out, leaving an overall correction of much smaller magnitude than the individual terms. Furthermore, **Table 4.7** shows that the two results calculated with small basis sets (B3LYP† and CCSD(T)‡) are outliers, which confirms the importance of using a large basis set in the calculation of polarisabilities [349]. This leaves four sets, all calculated with the same high-level basis set (d-aug-cc-pVQZ). These remaining sets show a converging trend in the value of the overall correction as the level of theory increases. The difference between the highest and lowest values for the overall correction

varies between 0.7 and 1.5 kJ/mol, although this would be significantly smaller (0.2 to 0.6 kJ/mol) were we to discard CCSD* as an outlier. In any case, it can be seen that the error introduced from the polarisation tensor is fairly small, provided an adequate level of theory is employed. From this point on, we will only consider that of Loboda et al. [348] (CCSD(T)◊), since it is calculated at the highest level of theory with the largest basis set.

Table 4.7: Overall solvation correction for different sets of multipoles and polarisability matrices at various levels of theory. * = d-aug-cc-pVQZ = (7s6p5d4f3g/6s5p4d3) (Reference [346]), † = cc-pv(t+d)z (Reference [322]), ‡ = (5s3p2d/3s2p) (Reference [347]), ◊ = d-aug-cc-pVQZ = (7s6p5d4f3g/6s5p4d3) (Reference [348])

Calculation	Overall Solvation Correction (kJ/mol)					
	B3LYP†	B3LYP*	MP2*	CCSD*	CCSD(T)‡	CCSD(T)◊
AIMD (2)	-12.2	-8.7	-9.0	-9.7	-8.4	-9.2
QM/4MM	-7.3	-8.0	-8.2	-8.9	-7.0	-8.3
QM/230TIP5P	-11.8	-9.7	-10.1	-10.8	-9.1	-10.3
QM/4TIP5P	-7.5	-9.8	-10.1	-10.6	-8.1	-10.3
AIMD (3)	15.6	-5.0	-4.7	-4.3	0.4	-4.5
AIMD (1)	10.2	-5.5	-5.3	-6.8	-1.3	-5.3

Next, we look at the breakdown of the components of the overall correction for all our multipole sets in **Table 4.8**. What is immediately clear is that the multipole set is a much larger source of uncertainty than the polarisation matrix, with the difference between the largest and smallest corrections being 5.8 kJ/mol. Additionally, the first four entries in **Table 4.8** are quite different to the last two, which is unsurprising, since the last two entries have dipoles (2.88 and 2.95 D, respectively) that are much closer to the experimental value (2.9 D) [47]. Furthermore, it can be seen that the overall correction is a near cancellation of the two individual contributions, especially for the the last two entries. Another aspect worth mentioning is that the last two sets of multipoles, those closest to the experimental estimate, yield values for the cavity radius in the Onsager model that are quite close to the value estimated from the experimental liquid density - 1.554 Å (see **Section 4.2.3**). This affords an additional degree of consistency to our approach for estimating the polarisation correction.

Although both effects cancel out to a significant extent, for almost all cases the negative contribution is larger in magnitude, leading to an overall negative correction.

Table 4.8: Breakdown of corrections for various multipole sets using the polarisability tensor of Loboda et al. [348] μ = Dipole, Θ = Quadrupoles

Calculation	Cavity Radius (Å)	ΔG_{el} (kJ/mol)		ΔG_{dist} (kJ/mol)			Total (kJ/mol)
		μ	Θ	$\mu-\mu$	$\mu-\Theta$	$\Theta-\Theta$	
AIMD (2)	1.808	-10.2	-7.3	7.5	-0.6	1.3	-9.2
QM/4MM	1.763	-11.5	-9.6	9.9	-1.5	4.4	-8.3
QM/230TIP5P	1.724	-12.9	-9.9	11.4	-1.3	2.4	-10.3
QM/4TIP5P	1.650	-16.4	-13.6	16.9	-2.3	5.1	-10.3
AIMD (3)	1.575	-21.6	-21.7	27.4	-5.4	16.8	-4.5
AIMD (1)	1.554	-23.6	-22.4	30.3	-5.2	15.7	-5.3

As described above, the magnitude of the correction tends generally to decrease as the multipole moments of the liquid phase increase. However, because there are three variables that can change independently between different rows (one dipole and two quadrupoles), it is not easy to discern clear trends from **Table 4.8**. In order to better elucidate the interplay between the two competing correction factors, we have simplified the calculation by assuming a relationship between the two quadrupole moments and the dipole moment (see **Figures 4.5** and **4.6**). This allows us to obtain an approximate estimate of the polarisation corrections using only the dipole moment as a free variable. The results using the polarisability matrix of Loboda [348] are plotted in **Figures 4.7** and **4.8**. It is important to emphasise that this procedure was carried out merely for mathematical convenience; we are not suggesting that any physical dependence of the quadrupole magnitudes on the value of the dipole moment actually exists.

Figure 4.7 shows how the magnitude of both corrections depends on the multipole moments of the liquid. The positive correction is equal to zero at a dipole magnitude equal to 1.86 (the start of the x-axis in this case), since this is equal to the gas-phase dipole, i.e. no polarisation has yet occurred so the multipoles are not yet distorted. The negative contribution is also zero; there are no other molecules to interact with the gas-phase molecule and produce a reaction field. This corresponds physically to a cavity of infinite size (see **Equation 4.16**), which would cause the continuum correction to tend to zero. One interesting result is the point of intersection between the positive and negative corrections, i.e. where the overall correction is exactly zero. In our case, the crossover is at 3.07 D, which is slightly higher than the experimental value of 2.9 D [47]. Again, in qualitative agreement with Leontyev and Stuchebrukhov [30], this strongly

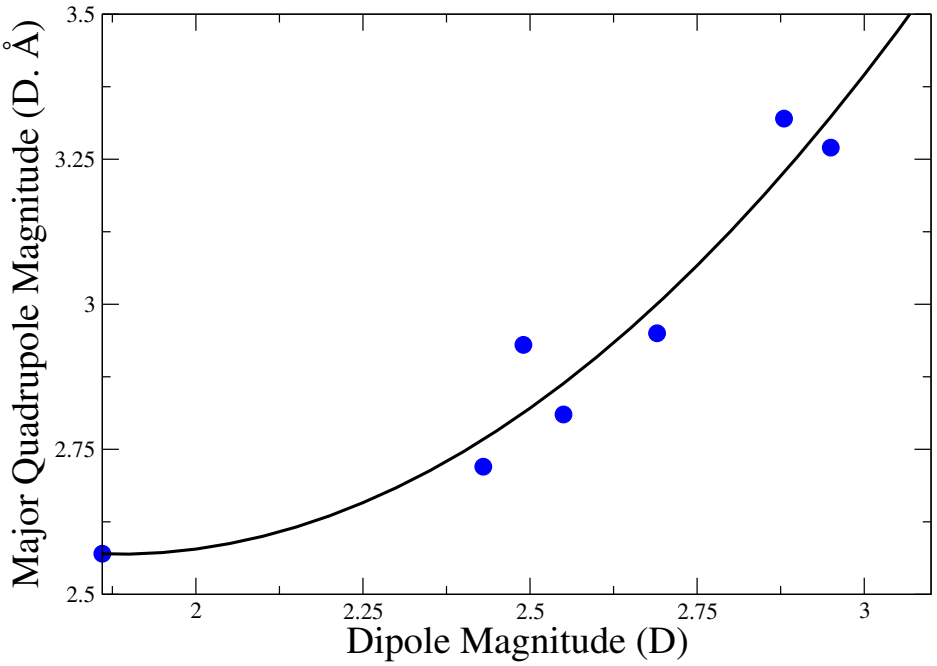


Figure 4.5: Fit of major quadrupole magnitude (Θ_2) to dipole magnitude (μ) for the parametric study. Fit: $\Theta_2 = 0.664\mu^2 - 2.50\mu + 4.92$, $R^2 = 0.93$

suggests that if realistic liquid multipole moments are used, the overall correction will be relatively small in magnitude.

In **Figure 4.8**, the overall correction is broken down into the individual multipole components (dipoles (μ) and quadrupoles (Θ)) using the polarisability matrix of Loboda [348]. The decomposition is as follows:

$$\Delta G_{\mu} = \Delta G_{dist:\mu-\mu} + \frac{1}{2}\Delta G_{dist:\mu-\Theta} + \Delta G_{el:\mu} \tag{4.22}$$

$$\Delta G_{\Theta} = \Delta G_{dist:\Theta-\Theta} + \frac{1}{2}\Delta G_{dist:\mu-\Theta} + \Delta G_{el:\Theta}$$

If only the dipole contributions are used, the correction becomes positive for dipoles larger than 2.73 D. However, the quadrupole contribution is strongly negative up to high values of the dipole moment (intercept at 3.28 D), shifting the intercept of the overall correction to 3.07 D. This implies that, at dipoles close to the experimental value of 2.9 D, the change in magnitude of the dipole hinders solvation, whereas the change in

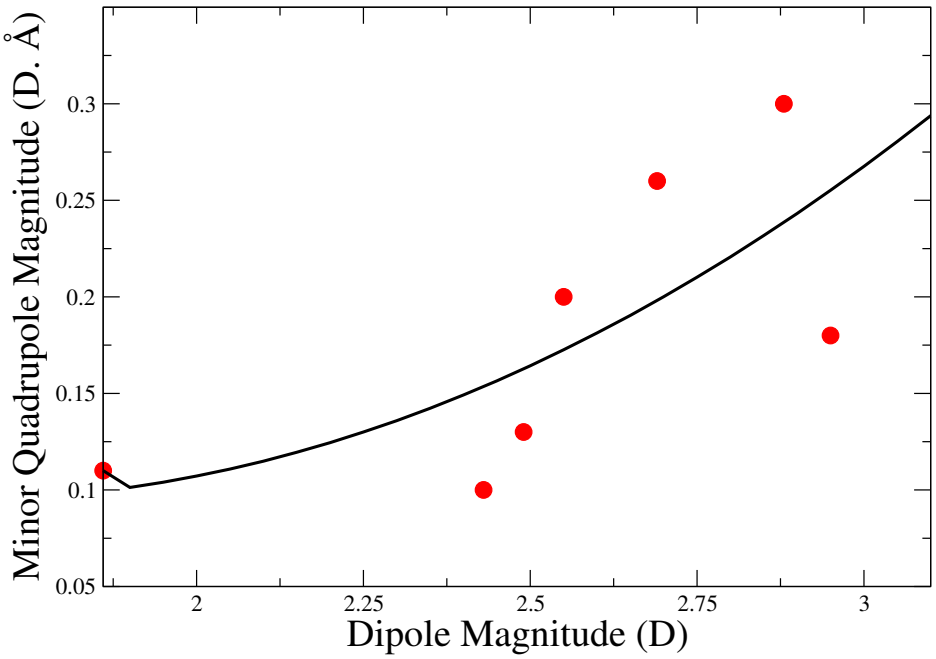


Figure 4.6: Fit of minor quadrupole magnitude (Θ_0) to dipole magnitude (μ) for the parametric study. Fit: $\Theta_2 = 0.0924\mu^2 - 0.302\mu + 0.341$, $R^2 = 0.48$

the quadrupole promotes it. It also explains why our overall polarisation correction is small and negative, while Leontyev and Stuchebrukhov [30], who considered only dipole contributions, obtained a small positive correction in their analysis.

Up until now, we have worked only in terms of the average values of the liquid-phase multipole moments. In reality, a range of fluctuating geometries will be present in the liquid, leading to a distribution of multipole magnitudes. As was explained in **Section 4.2.4**, the AIMD trajectories of Schienbein and Marx [48] allowed us to calculate the distributions of dipoles and both quadrupole moments in liquid water. The results are shown in **Figure 4.9**, where it can be seen that the distribution is nearly symmetric for all multipoles. The distributions for the dipole and major quadrupole seem very similar, both exhibiting a sharp maximum. However, plotting the major quadrupole with respect to the dipole for each molecule showed no correlation between the two properties (see **Figure 4.10**). The same is true for the minor quadrupole (**Figure 4.11**), which in any case shows a much wider distribution than the other two multipoles.

In order to test whether using the average multipoles is a good proxy for the whole

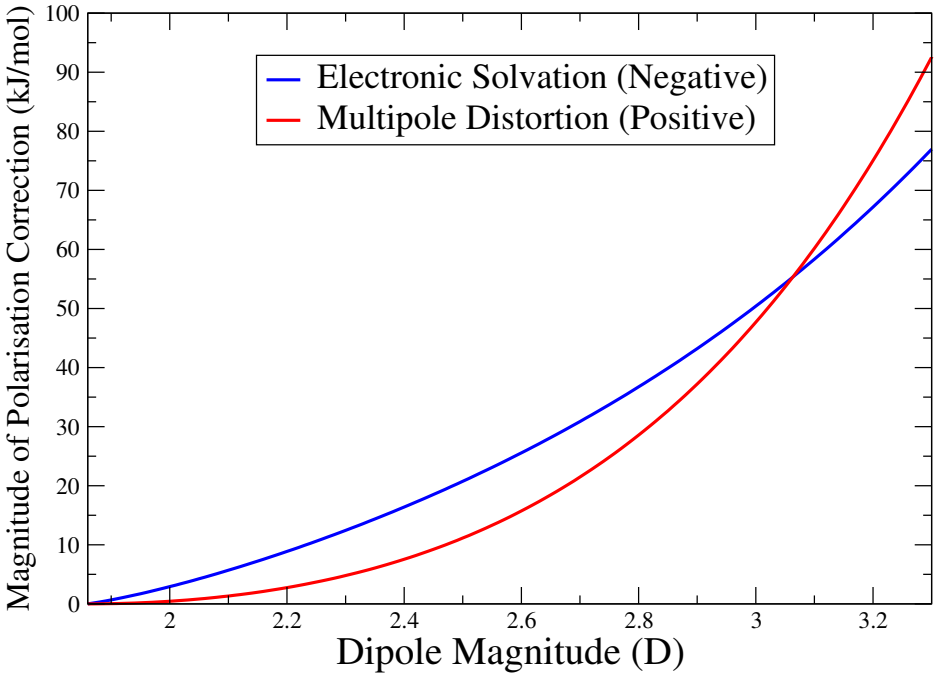


Figure 4.7: Absolute values of the two competing polarisation correction factors as a function of the dipole moment of liquid water using the polarisability tensor of Loboda et al. [348]

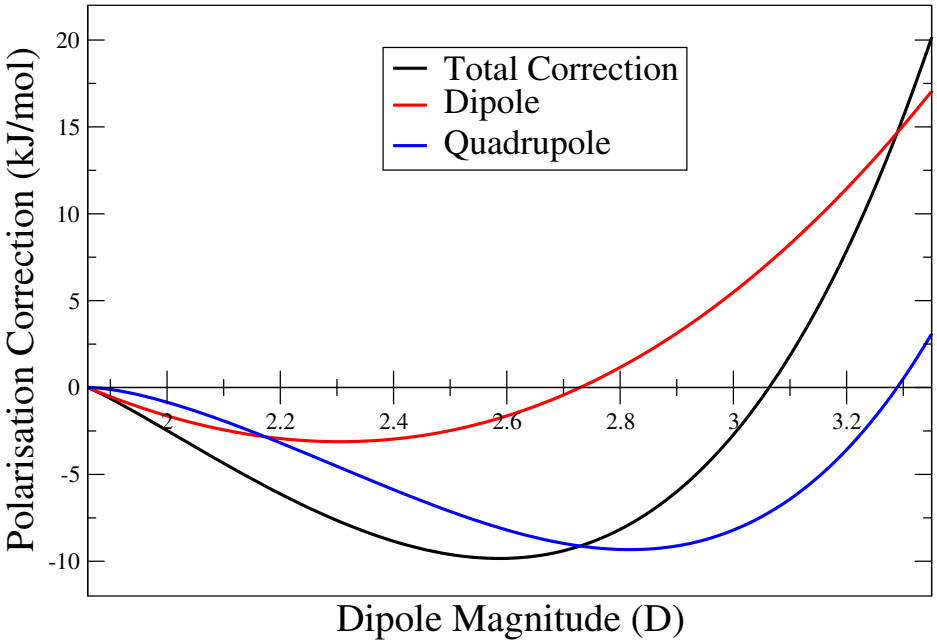


Figure 4.8: Dipole and quadrupole components of the overall polarisation correction (sum of distortion and electronic contributions) as a function of the dipole moment of liquid water using the polarisability tensor of Loboda et al. [348]

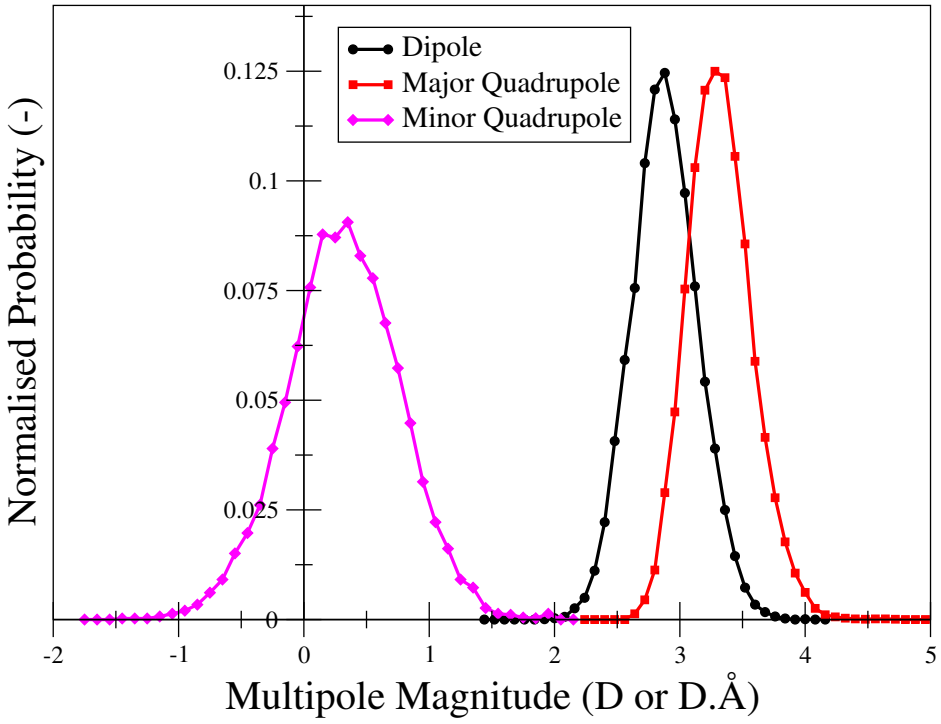


Figure 4.9: Multipole distributions calculated from data provided by Schienbein and Marx from their AIMD study of liquid water [48]

multipole distributions, the polarisation correction was calculated for all 12,928 individual water molecules in the AIMD trajectory and then averaged. The aggregated results of the distribution analysis are shown in **Table 4.9**, along with the equivalent calculation using the average multipoles. For each case, we considered two different ways of calculating the cavity radius: i) applying equation (16) to each individual set of multipoles; ii) using a fixed value obtained from the experimental liquid density (i.e., 1.554 Å).

The results in **Table 4.9** show that using the average multipoles as a proxy for the full distribution produces results which are more negative by about 5 kJ/mol, i.e. a significant discrepancy. Since the full distribution is a better representation of the physical reality of liquid water, we believe that these results are a better estimate of the necessary polarisation correction. Additionally, we can see that using the fixed rather than the individually calculated cavity radius produces results that are more negative

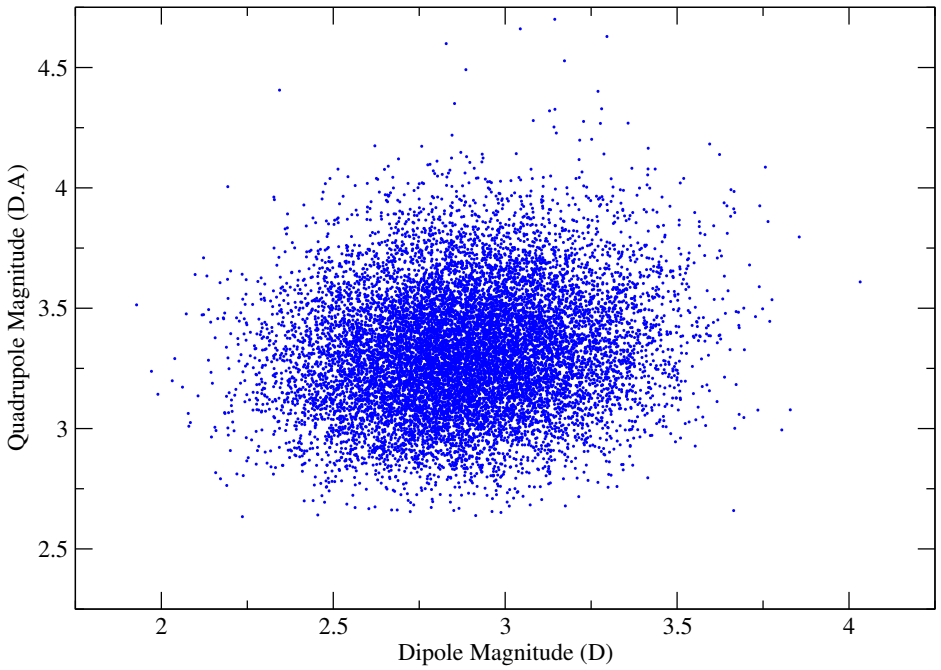


Figure 4.10: Relationship between the dipole (μ) and the major quadrupole (Θ_2). Calculated from data provided by Schienbein and Marx. There are 12,928 data points (101 frames, 128 water molecules per frame)

by about 2 kJ/mol. This is due to the generally smaller cavity radius in the fixed version, increasing the negative correction.

From **Table 4.9**, it can also be seen that although the dipole-dipole interactions for the multipole-distortion component are indeed the largest contribution, they are by no means dominant; they are generally a little more than twice as large as the quadrupole component. Nevertheless, the sum of the dipole-quadrupole and quadrupole-quadrupole interactions (i.e., the missing terms in a dipole-only treatment) is in almost all cases larger than the overall corrections, i.e. extremely important due to the delicate balance between the positive and negative corrections. For the electronic correction, this effect is even more pronounced; the quadrupole contributions are equal or larger than the dipole ones. Additionally, using only dipole terms, the overall correction in every case would be more positive, whereas including quadrupoles makes the overall corrections more negative. This questions the conventional wisdom that accounting only for dipole-dipole interactions is sufficient. Indeed, it also begs the question as to whether excluding

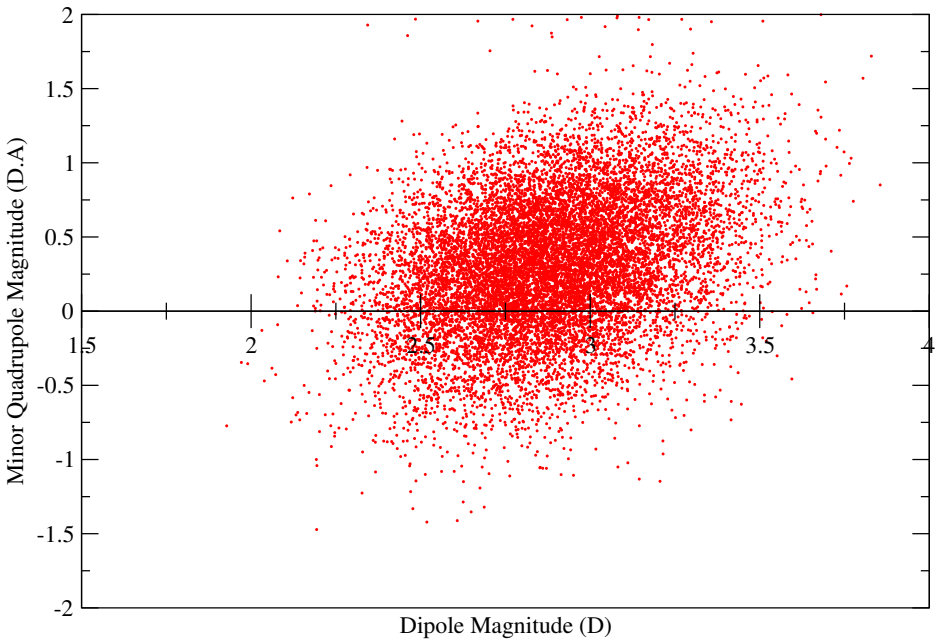


Figure 4.11: Relationship between the dipole (μ) and the minor quadrupole (Θ_0). Calculated from data provided by Schienbein and Marx. There are 12,928 data points (101 frames, 128 water molecules per frame)

multipole moments above quadrupoles (as in our case) is also a reasonable assumption.

Finally, in **Table 4.10**, we compare the results for the overall polarisation correction when different approximations are used: i) only dipoles are considered and the distortion correction is estimated using **Equation 4.1** (Row 1); ii) the electronic solvation correction as per Leontyev and Stuchebrukhov [30,324] is added, but only using dipoles. (Rows 2 and 3); iii) the full treatment, up to quadrupole moments, is employed (Rows 4 and 5, respectively). In all cases, the whole multipole distribution of Schienbein and Marx [48] was used and the final values are the average of the correction over all molecules and frames. We can see that using up to quadrupoles with both terms results in a lower correction, regardless of the method chosen to calculate the cavity radius. In both cases it can also be seen that using the fixed cavity radius, calculated from the liquid density, produces a slightly more negative result; the dipole-only correction is reduced from 1.6 to 1.0 kJ/mol and the combined correction from 0.8 to -1.4.

Having analysed in detail the effect of each variable on the calculation of polarisation

Table 4.9: Breakdown of corrections calculated from the AIMD trajectories of Schienbein and Marx [48] and the polarisability tensor of Loboda et al. [348] using different approaches (see text). μ = dipole, Θ = quadrupole.

Method		Cavity Radius (Å)	ΔG_{el} (kJ/mol)		ΔG_{dist} (kJ/mol)			Total (kJ/mol)
Multipoles	Cavity Radius		μ	Θ	$\mu-\mu$	$\mu-\Theta$	$\Theta-\Theta$	
Average	Calculated	1.575	-21.6	-21.7	27.4	-5.4	16.8	-4.5
Average	Fixed	1.554	-22.5	-23.2	27.4	-5.4	16.8	-7.0
Distribution	Fixed	1.575	-21.8	-22.1	29.0	-5.6	21.6	1.0
Distribution	Calculated	Variable	-22.1	-22.1	29.0	-5.6	21.6	0.8
Distribution	Fixed	1.554	-22.7	-23.7	29.0	-5.6	21.6	-1.4

Table 4.10: Comparison of different methods for polarisation corrections. In all cases, the multipole data of Schienbein and Marx [48] and the polarisability tensor of Loboda et al. [348] were used. All energetic corrections are given in kJ/mol

Method	Multipoles	R	ΔG_{dist}	ΔG_{el}	Total
Berendsen[a] [31]	μ	N/A	5.4	-	5.4
Leontyev [324]	μ	Variable from dipole	23.7	-22.1	1.6
Leontyev [324]	μ	Fixed (liquid density)	23.7	-22.7	1.0
This Work	$\mu+\Theta$	Variable from dipole	44.9	-44.1	0.8
This Work	$\mu+\Theta$	Fixed (liquid density)	44.9	-46.3	-1.4

[a] Berendsen correction is average correction over all water models shown in **Table 4.1**, i.e. using the model dipole in each case as the estimate for the dipole of liquid water

contributions, we are now in a position to identify our best estimate for the overall polarisation correction and revisit the MD results shown in **Figure 4.4**. As discussed above, our best estimate is obtained using the polarisability tensor at the highest level of theory (Loboda's [348]) and the multipole distribution of Schienbein and Marx [48]. As can be seen from **Table 4.10**, our final value could be either 0.8 or -1.4 kJ/mol, depending on which method for defining the cavity radius is chosen. Since the cavity radius is not a strictly defined property, it is hard to judge which method is more "correct" and we shall therefore consider both in our final analysis. We applied each of these corrections to the enthalpy of vaporisation and hydration free energy results calculated from MD simulations for the 11 non-polarisable water models. Fitting a linear trend through each of these data sets, we obtained the dashed and dotted lines shown in **Figure 4.12**. The linear fits yield slightly positive intercepts; 0.13 kJ/mol for the fixed cavity radius and 0.74 kJ/mol for the variable cavity radius. The trend lines obtained by applying the original Berendsen correction (intercept = 2.0 kJ/mol) and applying no polarisation correction (intercept = 0.52 kJ/mol) are also shown in **Figure**

4.12 for comparison.

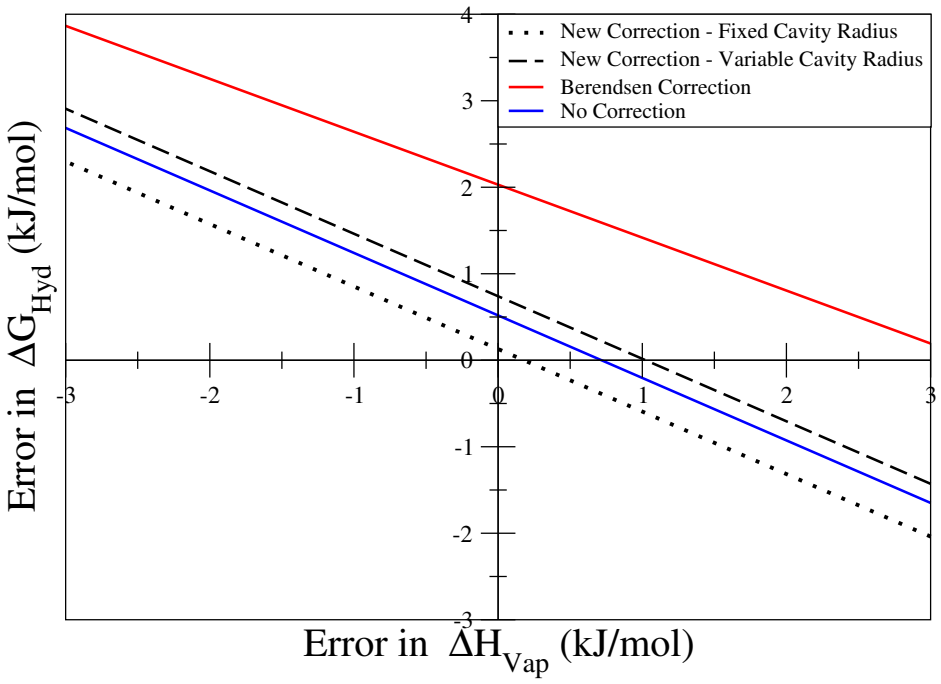


Figure 4.12: Linear fits for the results of the water models in **Table 4.1** using different values of the polarisation correction: Berendsen's original correction using model dipoles and our corrections (using either a fixed or a variable cavity radius) obtained from the multipole distributions calculated from the work of Schienbein and Marx [48]

The difference between the two best estimates of the polarisation correction gives an idea of the uncertainty of our procedure. Taking this into account, it is clear that the polarisation correction for water should be close to zero, corresponding to almost complete cancellation of the distortion and electronic contributions. This leads to an intercept close to the origin in **Figure 4.12**. In contrast, as discussed previously, applying the Berendsen correction with model dipoles, as done in most water-model parametrisation efforts, leads to a systematic deviation between enthalpy of vaporisation and hydration free energy predictions. The implication of our analysis is that it should be possible, at least in principle, to design a new water model that is able to simultaneously predict both these properties, while at the same time yielding an accurate description of other liquid properties.

4.4 Conclusions

In this chapter an analytical method for calculating polarisation corrections for MD simulations has been presented. This is based on two competing corrections to describe the electronic changes that molecules undergo when leaving the gas phase and entering the liquid phase. These are, namely, the cost (always positive for hydration) of distorting the electron density from one appropriate for the gas phase to that present in the condensed liquid phase, and the energetic advantage (always negative for hydration) arising from interaction between the molecule and the surrounding polarised electron clouds of the other molecules in the fluid. Our example here is based on water due to the large amount of data available for this molecule. We confirm that the overall polarisation correction is a near cancellation of the above two fairly large competing factors, which explains why early water models have been able to achieve good performance while neglecting them.

In the present work, multipoles up to quadrupoles have been used, whereas usually only dipoles are applied. It has been shown that the contribution of quadrupoles is not negligible in comparison to that of the water dipole, ranging from 50 to 100% of the value of the dipole correction for both positive and negative corrections. We have also shown that the impact of the polarisability tensor on the value of the polarisation correction is relatively small, provided a high enough level of theory is used in the underlying QM calculations. Additionally, we have calculated the correction for a distribution of liquid-phase water molecules (based on the AIMD data of Schienbein and Marx [48]); in this way we have shown that using the average multipoles underestimates the value of the correction, making it more negative. In general, the near cancellation of the two contributions described above means that any errors in the calculation of either component will be magnified in the final result. Taking all these factors into account, we obtain an overall polarisation correction that is close to zero, with an estimated uncertainty of around $\pm$ 1 kJ/mol.

Although our treatment represents a clear advance over existing approaches, it still relies on several assumptions, the validity of which requires future testing. For instance,

we have assumed that the gas-phase polarisability tensor can also be applied in the liquid phase. Although this is a commonly used approach, it is unclear to what extent the polarisabilities will change due to the presence of surrounding molecules. Furthermore, we have truncated our multipole expansion at the quadrupoles and have shown that the contribution of the latter is significant. Similarly, it is possible that higher-order multipoles will have a non-negligible contribution to polarisation. Finally, we have employed a very simple continuum model to calculate the electronic polarisation contribution, motivated by the need for an analytical expression. This is likely a reasonable approximation for water, as it is a small and nearly spherical molecule. However, much more accurate continuum methods exist [350], and their application to the calculation of the electronic polarisation term should be the subject of future research.

We applied our theory to the results from MD simulations of 11 different non-polarisable water models. Our results showed that the usual Berendsen correction actually makes it more difficult to parametrise a model that accurately predicts both the heat of vaporisation and the hydration free energy of water. In contrast, a value of the correction close to zero, as determined using our approach, eliminates this systematic deviation, even when uncertainty is taken into account. This suggests that, at least in principle, it should be possible to parameterise a water model that can provide an accurate description of the liquid phase, while predicting both energetic properties accurately. Confirming this assertion is left for future work.

The main drawback of our procedure is that it relies on rather complex QM calculations - high levels of theory are required to obtain accurate gas-phase polarisability tensors, while obtaining a distribution of multipole moments in the liquid phase is highly non-trivial. Although accurate estimates of both quantities were available for water, this is very much an exception, arising from the practical relevance of this molecule. Extending the approach proposed here to other systems therefore remains rather challenging, particularly in the context of force field development. However, our results suggest that, for strongly polar molecules, the overall polarisation correction should be close to zero. As such, we propose that a reasonable approximation would be to parametrise such models by neglecting energetic polarisation corrections altogether. This is cer-

tainly better than including a positive correction based on the current state-of-the-art, which, as shown, leads to systematic deviations in phase-change energies. For less polar molecules, however, the electronic component will dominate over a much smaller distortion, leading to an overall negative correction [324]. In this case, a simple approach for estimating polarisation corrections needs to be applied.

Chapter 5

Self-Consistent Charge Determination

5.1 Population Analysis and Point-Charge Fitting Procedures

In the previous section, we focussed on polarisation effects in fixed-charge force fields due to their importance in modelling the principal solvent in silica synthesis (water). While this gave us some key insights into how to account for polarisation analytically, it does not provide us with any information *a priori* as to how polarised the point charges in any given force field should be, in order to reproduce its liquid-phase behaviour and take into account electronic-shielding effects implicitly. This chapter will therefore focus on developing and testing a procedure for deriving appropriate electrostatic parameters for liquid-phase molecular simulations.

A fundamental part of any force field are electrostatic interactions. Technically, all interactions are derived from the behaviour of electrons, but electrostatics deal with those arising from the interactions between two charged particles or atoms. In a force field, electrostatics are usually calculated using Coulomb's law, as stated in **Chapter 2**. In reality, however, the charges are not fixed, i.e. static, but vary over time with their environment, as interactions with other atoms or molecules cause distortions in the electron cloud, thereby redistributing charge. This is called polarisation and was

dealt with extensively in **Chapter 4**. However, there we were mainly concerned with multipoles, i.e. the distribution of charge, rather than the magnitude of the charges themselves. There is also a key difference between dipoles and partial charges (PC) - dipoles are an experimental observable, while PCs are not.

As such, there are two approaches for determining partial charges for a force field. The simpler, but more computationally burdensome, strategy is to treat the charges as a free parameter in the parametrisation. This allows the charges to be optimised against the data that are most relevant to the model. The downside of this method is that the charges that are optimal are not always chemically intuitive. A good example of this is acetone in the TraPPE force field - its charges correspond to a liquid-phase dipole (2.51 D) lower than that of the gas-phase molecule (2.91 D from experiment, 2.97 D from an MP2 level quantum calculation [351]), which is clearly unphysical.

The second strategy is to base the charges on some experimental observable that is related to the electrostatics - this field is called population analysis. Population-analysis techniques can be split into families. The first, known as either quantum theory of atoms in molecules (QTAIM) or quantum chemical topology (QCT), are approaches whereby parts of the electron density are associated with a particular atom and this quantifies the PCs.

The third approach is electrostatic potential derived (ESPD) partial charges which, under a particular set of constraints, fit PCs to atoms in order to reproduce the molecular electrostatic potential as accurately as possible. As such, this last approach is similar to that of empirically parametrising the PCs, except that QM rather than experimental data are used for the fitting.

While the population analysis approach has been used successfully for describing charge transfer in molecules, as well as PCs in buried atoms, it is generally unsuitable for determining PCs for use in parametrising force fields. Common population analyses in this class are Mulliken, Hirshfeld and Bader charges. Mulliken charges are based upon an analysis from linear combination of atomic orbitals (LCAO) theory. If a molecular orbital (ψ_i) is comprised of two atomic orbitals (ϕ_j and ϕ_k) [352]

$$\psi_i = c_{ij}\phi_j + c_{ik}\phi_k \tag{5.1}$$

where c represents the coefficient of the atomic orbital in question. The square of this gives us the probability density of the charge

$$\psi_i^2 = c_{ij}^2\phi_j^2 + c_{ik}^2\phi_k^2 + 2c_{ij}c_{ik}\phi_j\phi_k \tag{5.2}$$

The integral of this must be equal to one, since it is a normalised distribution, leading to

$$1 = c_{ij}^2 + c_{ik}^2 + 2c_{ij}c_{ik}S_{jk} \tag{5.3}$$

where S_{jk} is the overlap integral, which quantifies, as the name suggests, the degree of overlap of the two atomic orbitals. Mulliken interpreted this as follows: c_{ij}^2 and c_{ik}^2 are the electron populations of atomic orbitals, and hence atoms, j and k, respectively. The remaining terms are termed the overlap population and describe bonding between the atoms. This forms the basis of the method. As such, Mulliken charges are most suited to describing how molecular orbitals do or do not participate in bonding [353]. A major disadvantage of the Mulliken technique is that it is extremely basis-set-dependent since it relies on the atomic orbital coefficients. Diffuse functions are especially problematic since they decay slowly, i.e. contribute to the electron density far from the nucleus yet are still attributed fully to the nucleus [353].

In 1977, Hirshfeld [354] introduced his eponymous method, which is also sometimes known as the stockholder scheme. His idea was to use a weighting factor that divided the electron density among the atoms according to the same ratio as it would be for the individual atoms. Formally [353]

$$\rho_A = \omega^0\rho = \frac{\rho_A^0}{\sum_i \rho_i^0}\rho \tag{5.4}$$

where ρ is the electron density of the molecule, ρ_i^0 is the electron density of isolated atom i and ω^0 is the weighting factor. This has the benefit of having a much lower dependence on the choice of basis set and functional since the same ones are used for the isolated

atoms as for the molecule in question. Because of this, they also converge in the basis-set limit, unlike Mulliken charges. Nevertheless, like the Mulliken charges, the partitioning is based upon an arbitrary choice, namely assuming that the weighting of the isolated states is indicative of the weighting of the electron density in the molecule [355]. For example, this will not include effects arising from the Pauli exclusion principle since it is based upon isolated atoms. Other concerns are the choice of isolated atoms; using two neutral nitrogen atoms and a combination of one negative and one positive will produce different sets of charges. Finally, Hirshfeld charges tend to be significantly smaller than those from other methods [355], especially for covalent systems [353].

ESPD charges have proved far more successful for force-field parametrisation and widely used methods are the electrostatic potential (ESP)/restrained ESP (RESP), charges from electrostatic potentials (CHELP)/CHELP using a grid-based method (CHELPG), Merz-Singh-Kollman (MSK) and density derived electrostatic and chemical (DDEC) [356] approaches. ESP was the first method to be put forward, being proposed by Momany in 1978 [357]. With the rise of molecular mechanics (MM), MD and MC, and their reliance on force fields, ESP fitting became the preferred option. This was because they optimally reproduced the ESP energy surface, which is ultimately also the goal of force-field development. This method is successful in part because it is a semi-empirical technique; it marries the use of fitting (empirical) to a high-level quantum calculation (QC). An appropriate QC provides a highly accurate level of molecular detail, while the fitting optimises it in a simpler (i.e., faster) form for use in a force field.

MSK, CHELP and CHELPG are all similar techniques, diverging only in how the ESP is sampled [358]. MKS uses points located on a surface at multiples of 1.4, 1.6, 1.8 and 2.0 of the VDW radii [359], whereas CHELP uses points located at 2.5—6.5 Å from the VDW surface in 1 Å intervals. Finally, CHELPG eschews the shell-based sampling of the two former methods and simply samples from a regularly-spaced grid that encloses the area between the VDW surface and a distance 2.8 Å away [360]. It should be noted that these methods are limited to non-periodic systems.

However, due to the simplicity of the fitting method, the majority of work in this area has dealt with ironing out its kinks. These are that ESP charges are heavily

conformationally dependent, not transferable between different molecules for the same functional group and are too large for use in force fields [361]. Bayly et al. [361] noted that this is due to using a fitting without any restraints, leading to problems with the fitting of charges to buried atoms. The ESP grid points are located outside the van der Waals surface of the molecule and therefore buried atoms, i.e. those far from the surface, will have a minor effect on the ESP compared to those at the surface. As a result of this, buried atoms are responsible for the system suffering from overfitting; the buried atoms effectively act as extra parameters, even though they barely contribute to the fitting. In this way, unphysical charges can be achieved because they nevertheless produce the "best" fit, i.e. the fit is mathematical-statistical rather than physical.

Bayly et al. improved ESP point charges by adding restraints to produce RESP charges [361]. They did this by implementing a simple penalty function that restrained the charges around a given value when fitting to the ESP. In the ESP method, a least-squares fitting procedure is implemented and a figure-of-merit (χ) is minimised. The new target value is

$$\chi^2 = \chi^2_{ESP} + \chi^2_{res} \tag{5.5}$$

where χ_{ESP} is the figure-of-merit for the (unrestrained) ESP fitting and χ_{res} is the restraint. The restraint function can, in practice, be any function, although it is desirable that it is symmetrical, well-defined at all points and has one minimum [361]. As a simple example, Bayly et al., in their seminal paper, first tested a harmonic function, i.e.

$$\chi^2_{res} = a \sum_i (q_i - q_{i,0})^2 \tag{5.6}$$

where a is the "spring constant", i.e. defines the strength of the coupling to the restraint, and $q_{i,0}$ is the value of charge that atom i is coupled to. They demonstrated that a harmonic coupling is not the optimal restraining function, showing that a hyperbolic function was preferable due to the harmonic function restraining polar atoms too strongly. There is also the matter of which value $q_{i,0}$ should take. The two options used

by Bayly et al. in their original paper were Mulliken charges and simply setting $q_{i,0}$ to zero. The latter approach was shown to be more effective. To summarise, the RESP method reduces the unrealistically high charges of the ESP approach by introducing a restraint function that minimally degrades the ESP of the molecule while ensuring transferability. There is also a variant of this method called REPEAT [362] that is suitable for periodic systems, e.g. bulk solids.

Finally, the latest method to become widespread is the DDEC fitting procedure, as proposed by Manz and Sholl in their 2010 paper [363]. The motivation for DDEC was to amalgamate the faithful reproduction of the electrostatic potential and buried atoms in RESP with the chemical descriptiveness of AIM methods. To this end, they created a target function combining the two

$$\chi = aF_{chem} + (1-a)F_{ESP} + \int \lambda(\mathbf{r})\Theta(\mathbf{r})d^3\mathbf{r} \tag{5.7}$$

F_{chem} is a restraint function that is used to keep the electron density close to that of some reference, i.e. chemically meaningful, state (similar to the iterative stockholder method), whereas F_{ESP} represents the electrostatic potential as given by the electron density. The restraint parameter, a, should be kept small so that it only becomes meaningful for systems whose electrostatic potential is not uniquely defined, i.e. those with buried atoms. DDEC is suitable for both periodic and non-periodic systems.

5.2 Integrating Polarisation Effects into Charge-Determination Methods

As was aptly demonstrated in **Chapter 4**, all non-polarisable water models have a dipole that is intermediate between the observed and quantum-chemically calculated gas- and liquid-phase dipoles. This phenomenon stems from a combination of electronic screening and molecular distortion due to polarisation. Therefore, the point charges in solution are markedly different compared to those of the same molecule in vacuum. As such, the foundations of a popular approach for including these polarisation-solvation effects in QM calculations, namely continuum solvation models (CSM), will now be

briefly elucidated before moving onto a new hybrid quantum/classical method proposed in this work.

As the name suggests, in CSMs the explicit presence of the solvent is approximated by an implicit continuum; this means that the large number of degrees of freedom of the solvent molecules is replaced by a continuous function. This is often a distribution function [350]. We have already met two of the simplest of these, namely the Born and Onsager models, in **Chapter 4**. As was already stated, the advantage of those models is that their representations are simple enough to have analytical solutions. It should be noted that these formulations also only deal with a single molecule in a cavity. Since their development in the first half of the 20^{th} century, new solvation models have been developed that have harnessed the power of computers, which have partially obviated the necessity of analytical solutions. This has led to new, more complex models, which have been successfully embedded into QM calculations and which will be applied at times in the remainder of this work. As such, the principles behind three variants used here will be briefly introduced. These are namely the integral equation formalism polarisable continuum model (IEFPCM), the solvation model based on density (SMD) and the self-consistent continuum model (SCCS). For a more complete review of CSMs in general, the reader is directed to the review by Tomasi et al. [350].

5.2.1 General Principles of Continuum Solvation Models

Consider a single molecule inside a cavity surrounded by solvent molecules. The Hamiltonian of such a solute-solvent system can be written as [350]

$$H(\mathbf{f},\mathbf{r}) = H_{solute}(\mathbf{f}) + H_{solvent}(\mathbf{r}) + H_{interaction}(\mathbf{f},\mathbf{r}) \tag{5.8}$$

where $\mathbf{f}$ and $\mathbf{r}$ represent the degrees of freedom of the solute and solvent, respectively. Since we are not interested in the full Hamiltonian of the solvent, this term can be ignored

$$H_{effective}(\mathbf{f},\mathbf{r}) = H_{solute}(\mathbf{f}) + H_{interaction}(\mathbf{f},\mathbf{r}) \tag{5.9}$$

This reduces the number of terms in which **r** appears, leaving only the interaction term. The full removal of **r** is then achieved by substituting this interaction Hamiltonian for a solvent response function ($Q(\overrightarrow{r}, \overrightarrow{r}')$, where $\overrightarrow{r}$ is a position vector), i.e. a continuous function acting as a simplified representation of the interaction Hamiltonian, which replaces the discrete positions of the solvent molecules. This gives us

$$H_{effective}(\mathbf{f}) = H_{solute}(\mathbf{f}) + U_{interaction}[\mathbf{f}, Q(\overrightarrow{r}, \overrightarrow{r}')] \tag{5.10}$$

which is the general formulation of a CSM. This representation tells us that the system is split into two systems that interact with each other: a solute molecule and a polarisable continuum. However, this provides no information as to how to define the cavity in which the solute finds itself. As such, there are two important details to be decided when using a CSM: the solvent response function and the geometry of the cavity.

For best results, the cavity should be defined in some physically meaningful way [350]. However, there are two conditions that have to be satisfied: first of all, it should be at least as large as the solvent-excluded surface (SES), i.e. the solvent should not be included in the cavity; secondly, as much of the electronic density should be included within the cavity as possible. These two requirements have to be balanced; the charge density only decays to zero at an infinite distance from an isolated molecule but at such a condition all the solvent would also be enclosed. There are several commonly used approaches for how to achieve this, of which two of the more elaborate are surfaces based on a fixed charge density (isodensity approach) and a fixed energy (isoenergy) [350,364]. These approaches are sensible choices, however they require the evaluation of further calculations and are therefore more costly. The most common compromise involves a set of overlapping spheres, whose radii are equal to the van der Waals radii of the atoms multiplied by a solvent-dependent scaling factor.

The electrostatics are formulated at the classical level and are embedded in the QM calculation. These take the form of the Poisson equation, which is the basis of Coulomb's Law

$$\nabla[\epsilon \nabla U_{interaction}(\overrightarrow{r})] = 4\pi\rho_{solute}(\overrightarrow{r}) \tag{5.11}$$

where ρ_{solute} is the charge density of the solute and ϵ is the dielectric constant of the continuum. In practice, this equation means that ρ_{solute} polarises the continuum, which then in turn polarises ρ_{solute} via $U_{interaction}(\overrightarrow{r})$ (the potential energy from the solvent response function). This is what necessitates an iterative solution. Due to the different electronic environments in the two separate volumes, this equation is applied differently in each. Within the cavity, the solute exists in vacuum, i.e. ϵ is equal to unity, resulting in

$$\nabla^2 U_{interaction}(\overrightarrow{r}) = 4\pi\rho_{solute}(\overrightarrow{r}) \tag{5.12}$$

Meanwhile, under the assumption that all the charge density is enclosed in the cavity and knowing that ϵ is a constant, the Poisson equation outside the cavity is

$$\epsilon\nabla^2 U_{interaction}(\overrightarrow{r}) = 0 \tag{5.13}$$

Finally, the total energy of the CSM is the sum of these two contributions with a jump function to make the transition across the surface of the cavity smooth and continuous.

5.2.2 The Integral Equation Formalism Polarisable Continuum Model (IEFPCM)

PCM models are a subclass of apparent surface charge (ASM) models. They are called as such because they replace the whole dielectric volume with a charge distibution on the surface that has the same reaction potential. In this way, the three-dimensional problem is reduced to a two-dimensional one. Formally, this can be written as [350]

$$V_{interaction}(\overrightarrow{r}) = \int_{\Gamma} \frac{\sigma(\overrightarrow{s})}{|\overrightarrow{r} - \overrightarrow{s}|} d^2 s \tag{5.14}$$

where Γ represents the surface, s the dimensions of the surface and σ the charge distribution of the surface. IEFPCM belongs to the PCM family of models but is more generalised than other variants such as dielectric PCM (DPCM) [365] and COSMO (a.k.a. conductor-like PCM) [366]. IEFPCM can be formally written as a single-layer potential [350]

$$V_{interaction}(\overrightarrow{x}) = \int_{\Gamma} \frac{\sigma(\overrightarrow{y})}{|\overrightarrow{x} - \overrightarrow{y}|} dy \tag{5.15}$$

where $\overrightarrow{x}$ and $\overrightarrow{y}$ are real numbers representing two points in space. σ can then be calculated as the unique solution of the equation

$$A\sigma = -g \tag{5.16}$$

where A and g are integral operators. According to Tomasi et al. [350], this simplifies the PCM approach by making the surface charge only dependent on the potential and not the electric field, which makes the calculation faster. It also implicitly corrects for any charge that is left outside of the cavity. For more details, the reader is directed to the original papers [367,368].

5.2.3 Solvation Model - Density (SMD)

One common subset of CSMs is the generalised Born (GB) class of models. The name is derived from the Born solvation equation, which we met earlier

$$\Delta G_{solv} = -\frac{\epsilon - 1}{\epsilon} \frac{q}{2R} \tag{5.17}$$

where q is the charge on an ion, R is the radius of the cavity and ϵ is the dielectric constant of the solvent. GB methods extend the Born equation from the case of one ion in a cavity to multiple interacting charges, i.e. generalise it [350]. The generalisation is as follows:

$$\Delta G_{solv} = -\frac{\epsilon - 1}{\epsilon} \sum_{i,j=1} \frac{q_i q_j}{2 f_{GB}} \tag{5.18}$$

The term f_{GB} is an operator whose purpose is to make the solution of the equation smoothly transition between the Born equation, when the two charges overlap, and the Coulomb potential, when the cavities of the two charges no longer overlap, i.e. they become ions rather than parts of the same molecule. The advantage of replacing the charge density with partial charges is that all the charge remains explicitly within the

cavity, although at the expense of a less accurate treatment of charge [369]. Defining f_{GB} is the difficult part of this approach, although all versions of f_{GB} strongly depend on the Born radii of the individual atoms. Additionally, since the charges are represented as monopoles, i.e. point charges, these techniques are more sensitive to the choice of basis set [369].

The most widely used GB models are the SMx series, where "x" represents an identifier for a specific model. This series extends the purely electrostatic treatment outlined above and includes terms to describe other phenomena, including cavitation, dispersion and solvent structural effects [369] via a surface-tension term. However, the SMD model that has been used here is actually a reparametrisation of the IEFPCM method to include these non-electrostatic contributions to the free energy of solvation. For more details, the reader is directed to the original SMD paper [369], while for SMx models in general to the review by Cramer and Truhlar [370].

5.2.4 Self-Consistent Continuum Solvation (SCCS)

The SCCS method was developed due to problems that arise when using the previously described models for AIMD calculations; the use of cavities in AIMD leads to numerical singularities [364]. The solution to this was developed by Fattebert and Gygi, who introduced a method whereby the dielectric constant would transition smoothly between unity and the value for the solvent as a function of the solute's electron density [371]. This model was then reformulated by Andreussi et al. [364] so as to eliminate numerical instabilities; they achieved this by presenting the problem in terms of induced polarisation charges, as in IEFPCM. For full details the reader is directed to the original papers of Andreussi et al. [364,372], as well as of Fattebert and Gygi [371].

5.3 Trial of a Hybrid Quantum/Classical Iterative Charge Determination Method

In an ideal scenario, the liquid-phase structure would be optimised quantum-mechanically (which includes important polarisation effects) and the solution's ESP surface obtained

for fitting point charges. However, this is problematic since optimising the structure of a sufficiently large quantity of molecules to represent the solution is computationally taxing in the extreme. Indeed, this is why MD simulations are still used for optimising structures of such quantities of molecules; it is many orders of magnitude faster and hence can be carried out on realistic time scales. Since the force field that we wish to develop is exclusively for organosilica in solution (that is after all where the initial, and pivotal, stages of the synthesis take place), it was decided to first of all test whether a hybrid QM and MD method could be developed in order to more accurately determine liquid-phase charges generally. With this goal in mind, a flowchart demonstrating the developed procedure can found in **Figure 5.1**. Precise details of how the procedure was carried out are explained in the next section but it can be briefly summarised thus: from initial guess values of the point charges, a classical force field is parametrised that fits to some experimental properties; the configurations from the classical simulations matching the experimental data are then used as the input configurations for a set of QM calculations, from which new point charges are derived; the intermolecular parameters of the force field are reparametrised using the new point charges; once again, the configurations are used as input for a new set of QM calculations; this procedure then continues until the optimised charges for the classical simulations and those from the QM calculations converge. In this iterative manner, it is hoped to derive QC-quality charges for classical simulations.

5.3.1 Development of Method and Computational Details

First of all, it was necessary to devise an exact procedure. This meant characterising the effect of various aspects of the QM-calculation method, such as

- The largest computationally tractable basis set for QM calculations of several thousand atoms
- The effect of various exchange-correlation functionals
- Finite-size effects
- Different charge-fitting methods

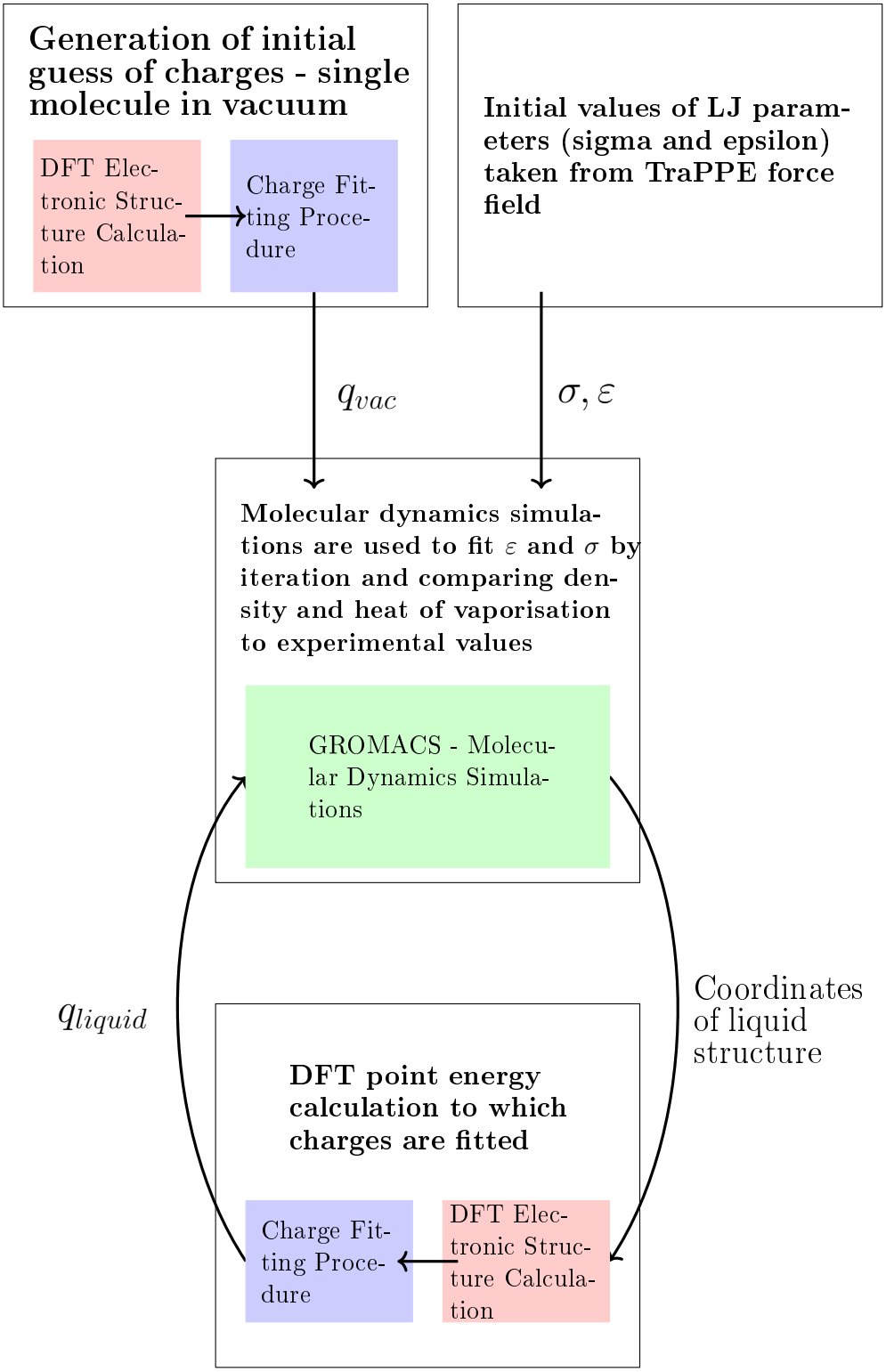


Figure 5.1: Procedure used to achieve self-consistent point charges via a combination of molecular-dynamics simulations and DFT calculations

For the sake of simplicity, it was decided to start with two molecules that are well-characterised: water and methanol.

The classical half of the procedure was carried out using simulations in the GROMACS [44] MD package (version 4.6). For these simulations of water, either 269, 300, 350, 600 or 900 molecules were simulated in four steps: energy minimisation, NVT-ensemble equilibration (0.1 ns), NPT equilibration (0.1 ns) and an NPT production simulation (1.0 ns). The electrostatics and van der Waals cut-offs were both 1.2 nm. PME [64] was used for long-range electrostatics. A Berendsen thermostat [87] and (for NPT simulations) a Parrinello-Rahman [94] barostat were employed. The same procedure was applied for methanol, but for only 180 molecules. For water, several different water force fields were tested (namely SPC [297], SPC/E [31], TIP3P [298] and TIP4P [299]) and for methanol the OPLS-AA force field [50] was initially used.

Having optimised the liquid structure in GROMACS, a Python script was used to extract all the frames from the MD trajectory that corresponded to the equilibrium density. Selected frames had to be at least 0.1 ns apart in time, so as to ensure that they were not correlated. The coordinates of all the atoms in these respective frames were then used as input for a QM calculation using CP2K (version 2.7) [373]. Upon attempting a calculation using an augmented triple-zeta (aug-TZ) basis set, it became clear that this was too computationally taxing and therefore only double-zeta (DZ) basis sets were used thereafter. The B3LYP and PBE exchange-correlation functionals were tested. The cut-off was 500 Ry, the relative cut-off 60 Ry and eight grids were used. The calculations were periodic. No optimisation of the liquid structure was attempted, rather a single-point energy calculation was carried out. REPEAT [362] charges were calculated directly by CP2K, whereas the DDEC charges were fitted using the DDEC [363] code, with the ".cube" file of the electron density from the QM calculation used as input. The results can be seen in **Table 5.1**.

From **Table 5.1**, several conclusions can be drawn. Firstly, finite-size effects are negligible or small; between the simulations of 269 (box size = 2.00 nm), 350 (2.18 nm) and 600 (2.62 nm) SPC/E molecules, the variations in the charges and dipoles determined using B3LYP and DDEC are minuscule. For the same charges obtained via

Table 5.1: Results for the average charge and its standard deviation, as well as for the dipole moment, for water molecules via the method described in **Figure 5.1** for two different charge-determination procedures

Calculation Details			DDEC			REPEAT		
Model	N_{mol}	XC	q_O	σ_q (%)	μ (D)	q_O	σ_q (%)	μ (D)
SPC	350	B3LYP	-0.7771	3.3	2.147	-0.8850	15	2.350
SPC/E	269	B3LYP	-0.7772	3.2	2.149	-0.8925	15	2.369
			-0.7761	3.2	2.150	-0.8970	16	2.381
			-0.7763	3.5	2.151	-0.9026	16	2.396
SPC/E	350	B3LYP	-0.7758	3.3	2.148	-0.8778	15	2.330
			-0.7764	3.7	2.147	-0.8761	14	2.326
			-0.7758	3.4	2.145	-0.8841	15	2.347
SPC/E	600	B3LYP	-0.7762	3.5	2.147	-	-	-
			-0.7759	3.5	2.146	-	-	-
SPC/E	900	B3LYP	-	-	-	-0.8693	18	2.308
			-	-	-	-0.8799	17	2.336
SPC/E	269	PBE	-0.7782	3.1	2.156	-	-	-
			-0.7784	3.4	2.157	-	-	-
SPC/E	350	PBE	-0.7785	3.2	2.156	-0.8679	16	2.304
			-0.7781	3.6	2.152	-0.8671	16	2.302
			-0.7780	3.3	2.151	-0.8747	16	2.322
TIP3P	350	B3LYP	-0.7858	3.0	2.203	-0.8815	15	2.340
			-0.7858	3.0	2.207	-0.8748	15	2.322
			-0.7858	3.0	2.206	-0.8753	16	2.324
TIP4P	300	B3LYP	-0.7781	3.0	2.186	-	-	-
			-0.7778	2.9	2.183	-	-	-
			-0.7779	3.2	2.185	-	-	-

the REPEAT procedure (and the final box containing 900 molecules instead of 600), the change is larger, albeit the dipole only decreases from an average of 2.38 to 2.32 D. The case for the influence of the exchange-correlation functional is similar; the DDEC charges are the same to the third decimal place for both B3LYP and PBE, whereas there is a small difference for the REPEAT charges. Finally, the influence of different models is best viewed by comparing SPC/E and TIP4P to SPC and TIP3P; SPC and TIP3P have noticeably lower densities (967 and 977 kg/m^3), which is a structural parameter, compared to their later generations (998 and 996 kg/m^3, respectively - much closer to the experimental value). However, it can be seen that the effect of this is also very small or negligible, although DDEC charges seem to be a little more sensitive to the density.

To conclude, the effect of various aspects of the computational procedure seems to be very minor, with the most important detail being the fitting method for the charges.

Examining the charges and their distributions is also instructive. The DDEC charges for SPC/E result in a dipole of around 2.15 D, whereas for REPEAT the value is approximately 2.35 D. The actual value for the SPC/E model is 2.35 D, so it would seem that the REPEAT charges are more consistent with the MD simulations. However, the dipole from the DDEC charges is a lot closer to that of the TIP4P model (2.18 D) so the agreement of the REPEAT and SPC/E dipoles could just be coincidental. Finally, the standard deviation of the charges clarifies how strongly an individual molecule is affected by its environment. This is especially important for determining atomic charges in solution where the environment has a large effect. From the AIMD dipole distribution (**Figure 4.9**, **Chapter 4**) calculated from the data of Schienbein and Marx [48], the standard deviation was computed and found to be 9%. Since the magnitude of the charges is directly related to the dipole, if we assume that the geometries are similar then it is possible to directly compare the standard deviation of the oxygen charges with the standard deviation of the dipole. What was found was that the standard deviations of the two methods lie almost exactly either side of the value from AIMD. For DDEC the value is around 3%, while for REPEAT it was approximately 15%. As such, it is not possible to say with any great confidence which method produces a more similar distribution of charges to that of the liquid phase. Nevertheless it can be said that the REPEAT method is much more responsive to its environment. In any case, it is surprising that the two sets of charges diverge so strongly. Since water is a small molecule, there are no buried atoms and hence it would have been expected that the two methods would have focussed on reproducing the ESP surface and have obtained a similar result.

In order to examine further whether the molecules in the multi-molecule calculation were affected by their environment, the charges were compared to their values in vacuum. This was done for both water and methanol. The results can be seen in **Table 5.2**.

The effect of increasing basis-set size in the vacuum, which is tested here as a proxy

Table 5.2: Comparison of charges derived from two different charge-determination methods in vacuum and in solution for water and methanol

Fitting Method	Basis Set	Water		Methanol			
		N_{mol}	q_O	N_{mol}	q_O	q_H	q_{CH3}
DDEC	aug-TZ	1	-0.771	1	-0.469	0.325	0.144
	DZ	1	-0.748	1	-0.474	0.336	0.138
	DZ	269	-0.777	180	-0.485	0.359	0.146
REPEAT	aug-TZ	1	-0.690	1	-0.614	0.376	0.238
	DZ	1	-0.721	1	-0.586	0.379	0.207
	DZ	269	-0.892	180	-0.593	0.400	0.193

for the same effect in the liquid, is unclear; in some cases this decreased the magnitude of the charges, whereas for others the opposite case was seen. The effect of calculating the charges in the many-molecule system however is more definitive - in almost all cases an increase is observed. However, the change in magnitude is surprisingly small in most cases - the only exception are the water charges obtained by means of the REPEAT method. These increase by about 25% compared to the calculation for a single molecule. Since the dipole moment of water increases by approximately 50% between the vapour and liquid phases, these results are the most realistic and hence it is disappointing that this behaviour is not repeated for methanol.

5.3.2 Test Case: Methanol

In order to test the effectiveness of this procedure, it was decided to start with methanol, since this is a simple molecule with a lot of data against which to parametrise a model. Since the objective of this work is to produce a united-atom (UA) force field for silicates, a UA representation of methanol was chosen, as shown in **Figure 5.2**.

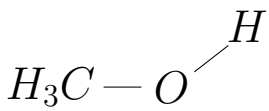


Figure 5.2: United atom representation of methanol used in this work. There are three interaction centres: methyl (CH_3), hydroxyl oxygen (O) and hydroxyl hydrogen (H)

This is the same representation used for TraPPE methanol [374,375]. Since we were

fitting to two properties (ρ and ΔH_{vap}), we could only allow for two degrees of freedom. These were chosen to be ε_O and σ_O. The methyl group parameters were taken to be the same as for the TraPPE model and the hydrogen atom had no LJ parameters, only a charge.

Initially, the charge of the molecule in question was obtained in vacuum for use as an initial guess value. This was achieved by carrying out a structure optimisation of methanol via CP2K (version 2.7) to obtain REPEAT charges along with the electron density. The electron density was then used to fit charges via the DDEC fitting procedure [363]. The B3LYP exchange-correlation functional was used alongside a DZ basis set, as was also used for the liquid-phase calculations. All other details were the same as described in the previous section.

Having generated a set of initial charges, the liquid structure was optimised via MD (using GROMACS 4.6 [44]). The well depth (ε) and the intersection of the x-axis (σ) of the LJ potential were fitted in order to reproduce the density (ρ) and heat of vaporisation (ΔH_{vap}) of the chosen molecule. The density was chosen as the structural parameter to fit against, whereas the heat of vaporisation was chosen so as to best fit the energetics of the model. The TraPPE [374] values for ε and σ were used as initial guesses.

One subtlety that had to be included were dummy hydrogen atoms on the methyl group. This was due to the fact that the model we were parametrising is a UA representation. While this presents no problems in MD simulations, a UA representation is incompatible with QM calculations, where all the nuclear centres have to be explicitly stated. In order to circumvent this, hydrogen atoms were added whose mass, charge and LJ parameters were set to zero. The only addition to the UA model were hence intramolecular parameters, so as to ensure that the positions of the dummy methyl hydrogen atoms were reasonable. This was not expected to affect the parametrisation since, although this would add extra energetic terms, we used the heat of vaporisation as the fitting property for the energetics. The heat of vaporisation effectively measures the strength of the intermolecular interactions present in the liquid-phase. Since the intramolecular interactions will be very similar in each phase, they will effectively cancel

out.

Finally, these new charges were used as new guesses for a new set of MD simulations. Once again, ε and σ were optimised in order to fit ρ and ΔH_{vap}. This cycle was then repeated until the charges from the MD simulations and DFT calculations agreed. An example of the convergence for REPEAT charges can be seen in **Table 5.3**. They demonstrate that the change from the vacuum charges is minimal; indeed, the charge on the oxygen converges to almost exactly the same value as in vacuum upon optimising the LJ parameters. The only change is a transfer of negative charge from the hydrogen in the hydroxyl group to the methyl group. This makes the hydroxyl group slightly more polar but this change is effectively negligible.

Table 5.3: Example of convergence for methanol charges using the iterative procedure outlined in **Figure 5.1** for the REPEAT fitting procedure

Iteration	N_{frame}	q_{CH_3}	q_O	q_H
0	1	0.207	-0.586	0.379
1	3	0.173	-0.571	0.398
2	3	0.186	-0.588	0.402
3	6	0.187	-0.585	0.398

The initial results are shown in **Table 5.4**.

Table 5.4: Results for charges of UA methanol models from an iterative charge-determination procedure. TraPPE charges are shown for comparison. The dielectric continuum available in CP2K was incompatible with the REPEAT charges and was therefore not used

Model	q_{CH_3}	q_O	q_H
DDEC - Vacuum	0.126	-0.461	0.335
DDEC	0.131	-0.485	0.354
DDEC - Continuum	0.158	-0.538	0.380
REPEAT - Vacuum	0.207	-0.586	0.379
REPEAT	0.187	-0.585	0.398
TraPPE	0.265	-0.700	0.435

From the results, it can clearly be seen that the DDEC charges were unsuccessful in capturing any of the effects of polarisation; the magnitude of the charges is barely greater than those from the vacuum calculation of methanol. Even with the application of a

dielectric continuum model (SCCS model [364,372], $\epsilon = 32$), the values of the obtained charges are a lot less than the REPEAT ones. The REPEAT charges were hence chosen for the model. Additionally, the polarisation corrections developed in **Chapter 4** were applied here too, albeit in a simpler form. This was due to the fact there were less data available for methanol, e.g. no dipole-quadrupole and quadrupole-quadrupole polarisability tensors. The cavity radius was calculated both from **Equation 4.16** in **Chapter 4** and estimated from the liquid density. Both results were very similar, 2.03 and 2.04 Å, respectively. Then the equation for the liquid dipole from Leontyev and Stuchebrukhov [30], namely

$$\mu_L = \frac{\mu_G}{1 - \frac{2(\epsilon-1)}{2\epsilon+1}\frac{\alpha}{R^3}} \tag{5.19}$$

was applied. As before, ϵ is the dielectric constant, α is the isotropic polarisability, μ_G and μ_L are the gas-phase dipole and liquid-phase dipole, respectively, and R is the cavity radius. The obtained value is 2.70 D. This value agrees well with a recent AIMD study by Sieffert et al. [376]. In their study they tested several AIMD procedures for liquid methanol and obtained values between 2.58 and 2.84 D, with the majority of values being approximately 2.7 D. Then the usual equations for each polarisation correction were used. This resulted in values of ΔG_{dist} and ΔG_{el} of 9.2 and -8.8 kJ/mol, respectively. This means that the overall polarisation correction for the free energy is 0.4 kJ/mol, i.e. -0.4 kJ/mol for correcting the heat of vaporisation.

Table 5.5: Lennard-Jones parameters for the hydroxyl oxygen in methanol fitted against experimental data. The charges used were those from the row entitled "REPEAT" in **Table 5.4** and the Lennard-Jones parameters for the methyl group were the same as those in the TRAPPE-UA representation

Model Parameter	σ_O (nm)	ε_O (kJ/mol)
Value	0.284	0.95
Property	ρ (kg/m^3)	ΔH_{vap} (kJ/mol)
Model	787.5 ± 0.6	37.20 ± 0.03
Experimental	787.8	37.6 ± 0.5

From **Table 5.5**, it can be seen that a good fit to the experimental data was found for the given charges; both fitted properties are within statistical error of their experimental

values. Compared to the TraPPE values, the value of σ is lower (0.284 to 0.302 nm) but ε is higher (0.950 to 0.773 kJ/mol), i.e. there is less short-range repulsion and more attractive long-range dispersion.

Subsequently, the model was validated against the TraPPE-UA and OPLS-AA models for several properties: the free energy of self-solvation, the dipole moment in solution and the dielectric constant. The main comparison here is between TraPPE-UA and our reparametrised version because they diverge in the way that the charges are obtained. TRAPPE-UA uses empirically-derived charges, whereas ours implements those from QM calculations. OPLS-AA was included so as to test the sensitivity of the given properties to a completely different formulation and parametrisation strategy, as well as due to it being a very popular model in the literature. This is, therefore, not for the sake of a direct comparison to an all-atom formulation; if that were the case, TraPPE-EH (explicit hydrogen) would have been a more appropriate choice [377]. The results are shown in **Table 5.6**.

Table 5.6: Results of the validation procedure for the developed model of methanol compared with the equivalent TraPPE-UA and OPLS-AA models

Model	ΔG_{solv} (kJ/mol)	μ (D)	ϵ
This Work	-18.1 ± 0.4	1.88 ± 0.08	16.8
TraPPE-UA	-20.3 ± 0.2	2.3 ± 0.1	23.4
OPLS-AA	-17.7 ± 0.4	2.3 ± 0.1	25.4
Experimental	-20.5	2.5	33.6

The results show that the proposed force field reproduced the free energy of self-solvation of methanol worse than TraPPE-UA, however better than OPLS-AA, although the statistical uncertainties of the two overlap. However, the agreement is significantly worse for the dipole moment and dielectric constant. As can be seen, the TraPPE model performs very well. However, for the application for which we intend to obtain charges (namely organosilica molecules in solution) a similar approach to TraPPE (empirical fitting of point charges) is not possible, since for organosilica molecules there are much fewer data compared to such a well-studied group as alcohols or similar organic molecules. This will be discussed in more detail in **Chapter 6**. Therefore, there are

not enough data points to fit to if point charges are left as free parameters.

It is interesting to note that both TraPPE and OPLS-AA have model dipoles (2.3 D) that sit approximately halfway between the experimental gas- (1.69 D) and liquid-phase (approximately 2.7 D) values (average value: 2.2 D). As was shown in **Chapter 4**, most widely used water models also have a dipole that lies between that lies between these extremes due to the effect of electronic screening in the medium. This once again shows the importance of the dipole and suggests that if a method was developed to estimate or measure the liquid-phase dipole, which is a non-trivial task, then it could be instrumental in parametrising point charges for molecular force fields.

In this section, a procedure was developed to try and obtain empirical-quality point charges using an iterative loop of MD simulations and QM calculations. The procedure was ultimately shown to be ineffective, with the UA methanol model developed being shown to be inferior to that of the TraPPE formulation. One reason for this could be that the TraPPE parameters were used for the methyl group in our model; if the LJ parameters for the methyl group had also been reparametrised, then a more coherent model might have been obtained. However, it is unlikely that this would have made a difference to the model's values for the dipole and dielectric constant since they are inherently electronic properties, i.e. related to the model's point charges. Additionally, it was seen that the charges obtained from a QM calculation on a box of molecules in their liquid structure were barely more polarised than in vacuum. This could have been due to the low level of theory employed, namely a DZ basis set. Indeed, Andreussi et al. provide several arguments [364] as to why this might be the case for water: standard DFT does not include hydrogen bonding and dispersion interactions; it also neglects the quantum-mechanical motion of the hydrogen nuclei. They state that the dielectric response, which causes the polarisation of the charges, is dominated by short-range hydrogen bonds, hence the neglect of this phenomenon in standard DFT most likely explains why almost all of the calculations here failed to achieve anything like the required polarisation/dielectric response. More advanced DFT methods could potentially solve this problem but would likely be too computationally arduous for use in parametrising force-field parameters en masse.

5.4 Sensitivity Analysis of Charge-Fitting Protocols

Due to the lack of success in obtaining effective point charges from QM calculations of bulk liquid configurations, it was decided to apply a simpler approach. This was to use a CSM together with the optimisation of individual molecules. Due to the switch to calculations of individual molecules, it was decided to use the Gaussian quantum-chemistry package, specifically Gaussian09. Furthermore, it was decided to test the influence of two charge-fitting procedures (CHELPG and DDEC) together with that of two CSMs, namely IEFPCM and SMD. Methanol was once more the test molecule. The basis set for all results was aug-QZ and the exchange-correlation functional employed was B3LYP. The results are summarised in **Table 5.7**.

Table 5.7: Results for methanol charges from various combinations of continuum solvation models and charge-fitting procedures

Charge Fitting	Solvation Model	q_{CH_3}	q_O	q_H	μ (D)
CHELPG	IEFPCM	0.277	-0.693	0.416	1.98
	SMD	0.301	-0.755	0.454	2.21
DDEC	IEFPCM	0.230	-0.645	0.415	2.09
	SMD	0.247	-0.693	0.446	2.27
TraPPE-UA		0.265	-0.700	0.435	-

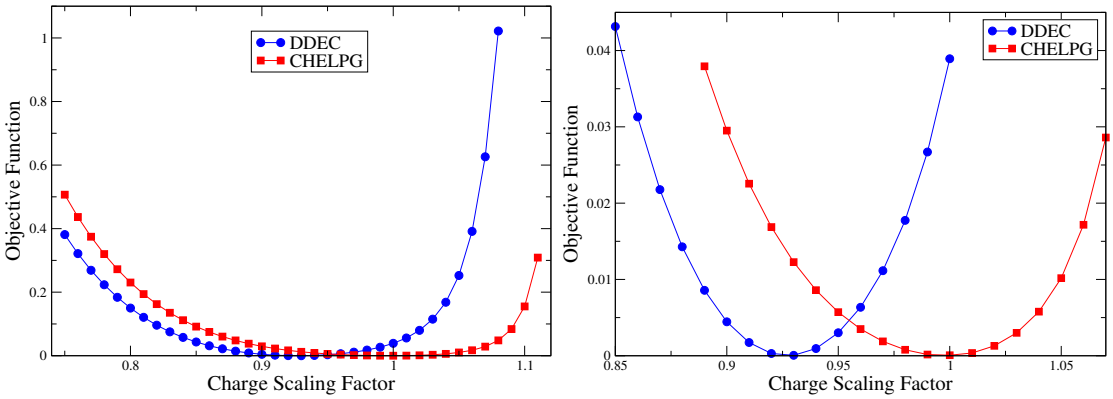
From **Table 5.7**, it is clear that both the charge-fitting method and the choice of continuum model have a noticeable effect on the result. The dipoles obtained in each case are about 10% higher for SMD than for IEFPCM; indeed, they are very close to the halfway point of the gas- and liquid-phase dipoles (2.2 D). Another interesting observation is that the SMD charges for each charge listed above are almost exactly a multiple of the IEFPCM charges. If the SMD charges are scaled by 0.92 then the IEFPCM charges are recovered. From these results it would seem that the addition of the surface tension to the solvation model increases the strength of the dielectric response, although it could also be due to the reparametrisation that was undertaken in the development of SMD. Finally, two sets of results (CHELPG/IEFPCM and DDEC/SMD) have point charges that are very close to those from the TraPPE-UA model of methanol, which is known to be a robust model. Curiously, the dipoles of these two sets of charges are the most different between all sets despite their similar charges; this tells us that the

obtained geometries are noticeably different.

Since the results in **Table 5.7** do not indicate conclusively that any one calculation set-up should be preferred, it was decided to carry out a sensitivity analysis of a UA methanol model with different sets of charges. The sensitivity was tested via a methanol metamodel that was developed and generously provided by a colleague, Ms. Maria Cecilia Barrera. The metamodel was developed by carrying out MD simulations for a large set of values of the intermolecular parameters, for which the density, heat of vaporisation and diffusion coefficient were calculated. This grid allowed for the quality of fit for these properties for different intermolecular parameters to be predicted quickly, in this way being able to rapidly suggest suitable parameters. The predicted results for each model were scored using an objective function (denoted as O). This function took the mathematical form

$$O = (\rho_{model} - \rho_{exp})^2 + (\Delta H_{vap,model} - \Delta H_{vap,exp})^2 + (D_{model} - D_{exp})^2 \qquad (5.20)$$

where ρ represents the density, ΔH_{vap} the heat of vaporisation and D the diffusion coefficient. The "$model$" and "exp" subscripts represent the predicted model and the experimental values of the observable in question, respectively. As such, a lower score is indicative of a better model. For this particular analysis, the charges were held constant and then suitable LJ parameters for the hydroxyl oxygen found (the methyl group was already robustly parametrised) that minimised the objective function. This was then carried out for different magnitudes of charges by scaling them by a constant and then carrying out the analysis again for the new charges. As was discussed before, the IEFPCM and SMD charges are seemingly linked by a scaling factor, hence only one set was chosen as input for the two different charge-fitting procedures. For CHELPG, this was IEFPCM, while for DDEC, SMD was used. The reason for this was, as previously discussed, these sets of charges were very close to the charges for TraPPE-UA methanol model and hence would likely provide good central points for the analysis, i.e. these values would likely be close to the optimum fit. The spacing for the scaling parameter for the charges, α, was 0.01. The results for this analysis can be seen in **Figure 5.3**.



(a) Full analysis (b) Enlargement of area around the minima

Figure 5.3: Sensitivity of a united-atom methanol model, as determined by a metamodel, to scaling of charges obtained using the CHELPG and DDEC fitting procedures

From **Figure 5.3**, it is clear that all the sets of charges obtained above would be suitable. However, it is also clear that IEFPCM seems to produce better charges; the minima for the CHELPG and DDEC charges are $\alpha = 1.00$ and $\alpha = 0.93$ (see **Table 5.8**), which correspond to those from calculations employing the IEFPCM CSM. The values of α representing the SMD continuum (1.07 and 0.99 for CHELPG and DDEC, respectively) are only mildly worse, however they sit right at the edge of the shallow region for the objective function. This tells us that any overestimation of these charges could lead to a rapid decay in the quality of the fit obtained from parametrising the LJ parameters of the hydroxyl oxygen with these sets of charges. These results clearly demonstrate why IEFPCM is such a widely used continuum solvation model. At least in the case of UA-methanol, it produces charges from two different fitting methods that are suitable for molecular force fields.

Although the results seem to clearly demonstrate which continuum model is more suitable for our purposes, the matter of which fitting procedure to use is less clear - both fitting procedures tested here are predicted to give highly satisfactory results. For this reason both fitting procedures were used going forward to organosilica molecules.

Table 5.8: Predicted optimal metamodel results and Lennard-Jones parameters for charges based on the CHELPG and DDEC charge-fitting procedures.

Optimum Values	CHELPG	DDEC
α	1.00	0.93
O	7×10^{-5}	6×10^{-5}
ε (kJ/mol)	0.815	0.945
σ (nm)	0.295	0.288
ρ (kg/m)	786.6	786.6
ΔH_{vap} (kJ/mol)	37.6	37.6
D (m^2/s)	2.35	2.36

5.5 Conclusions

The first half of this chapter proposed a method for identifying effective point charges for molecular force fields via an iterative method combining both classical (MD) and QM approaches. This approach was ultimately shown to be unsuccessful, producing point charges that were still significantly underpolarised, even in comparison to the charges from other force fields. This could either have been due to the low level of theory employed in the DFT calculations or due to deficiencies in standard DFT calculations, i.e. the lack of hydrogen bonding.

The second half of this section concerned itself with identifying the optimal procedure for determining the point charges for a chosen set of organosilica molecules. Two continuum solvation models (IEFPCM and SMD) and two charge-fitting procedures (CHELPG and DDEC) were analysed. On the basis of a sensitivity analysis, IEFPCM was deemed to be the most effective CSM method. On the other hand, no definite conclusion was reached on account of the effectiveness of the fitting methods.

Chapter 6

Parametrisation of a Force Field for Silicates in the Liquid Phase

6.1 Introduction

Up until now, we have focussed primarily on general aspects of force-field development for liquids and solutions: 1) in **Chapter 4**, we considered how to account for polarisation effects in aqueous media; 2) in **Chapter 5**, thought was given as to how to derive point charges that are representative of molecules in liquid-phase environments. We now bring our attention solely to silica-based molecules. More precisely, the parametrisation of a transferable model for organosilica molecules in the liquid phase.

Although some molecular simulation parametrisations for silica exist, these are inevitably tuned towards solid silica phases. Notable examples include the force fields of Cruz-Chu et al., Heinz et al. and Emami et al. [147,378,379]. However, we are not concerned with the ability of a force field to perfectly describe the solid phase, but rather its capability to correctly represent the liquid- and solution-phase behaviour during the synthesis of silica materials.

The literature introduced in **Chapter 3** acquainted the reader with studies that did indeed focus on the liquid-phase self-assembly. From this literature survey, it was clear that most methods were bespoke, employing techniques as diverse as lattice and kinetic Monte Carlo simulations, rather than applying a general force field. This holds

back the field from the standpoint of discovering new materials since more time is spent developing model systems than studying them. For this reason, we intend to start developing a force field for organosilica based on a transferable framework so that it will be easier to interchange different silica sources. In this manner, we intend to facilitate the study of a wider range of syntheses, with the hope of uncovering new, promising materials.

6.2 Experimental Data for Organosilica Molecules

One of the difficulties in parametrising a force field for organosilica molecules lies in the dearth of experimental liquid-phase data. Most studies focus on their use in the synthesis of a material, rather than the properties of the molecules themselves. However, in order to parametrise accurate potentials for these molecules, it will be necessary to have some bulk-liquid data against which to fit our force field parameters. Ideally, as broad a range of properties as possible would be available, so that we could fit the properties that are most relevant for our system of interest. Density is useful for parametrisations because it is simple to calculate and ensures that the model can reproduce at least one of the structural properties of the bulk phase. Additionally, if the LJ potential is being parametrised, the density is particularly sensitive to σ, making the density a good property to involve when parametrising this part of the potential. The heat of vaporisation and the free energy of solvation are both useful because they derive principally from the strength of the interactions between molecules. Since ε in the LJ potential is a measure of the relative strength of an interaction between atoms, the heat of vaporisation and/or free energy of solvation are obvious properties to include in the parametrisation. As mentioned in **Chapter 3**, the synthesis of mesoporous silica is strongly dependent on electronic, i.e. electrostatic and polarisation, effects. Therefore, it would be advantageous to have data for properties arising from such phenomena, such as dielectric constants and values of the dipoles in the gas and liquid phases, for the fitting of organosilica parameters. Finally, since the synthesis of silica nanomaterials relies on the aggregation and reaction of silicates, self-diffusion constants or diffusion constants in water would also be useful data points. A set of the abovementioned

experimental properties would represent a solid basis on which to parametrise or validate a model for organosilicates. Apart from the parameters themselves, a large training set of molecules is also crucial so that parameters for as many different functional groups as possible can be ascertained and validated.

The aforementioned lack of bulk-liquid data is another reason for choosing a transferable force field. The crucial molecule in the synthesis is silicic acid ($Si(OH)_4$), however, as noted in the literature survey, it is extremely reactive and polymerises on extremely short time scales. The polymerisation reaction was already given in **Chapter 3** (see **Equation 3.4**). For this reason, it is impossible to parametrise liquid-phase potentials for it directly, leaving transferable parameters from similar molecules as a potential avenue.

As a first step, it was necessary to collect as much data as possible for organosilica molecules, so that a parametrisation strategy could be defined. The data for 12 such molecules can be found in **Table 6.1**. Each of these molecules falls into one of four classes: (organo)silanes, containing only silica and alkane components; silanols, which are comprised of silica, alkane and alcohol groups; siloxanes, i.e. those containing an $Si-O-Si$ combination; and finally, alkoxysilanes, which feature an $Si-O-C$ chain as part of their structure.

Silanols were chosen since they contain $Si-O-H$ groups, as are found in silicic acid. Unfortunately, only the monoalcohol version is stable in the liquid phase; the dialcohol (dimethylsilanediol) is already solid at STP [380]. This means that it will not be possible to explicitly assess the effect of multiple, closely-located hydroxyl groups, however it is hoped that it will be possible to include this behaviour indirectly by comparison to similar behaviour in analogous organic molecules. Alternatively, the molecules tetramethoxysilane and tetraethoxysilane have four oxygen atoms surrounding a central silicon atom, albeit the oxygens are part of alkoxy, rather than hydroxyl groups. Nevertheless, this will provide some insight into how the behaviour of a given molecule changes when alkane chains are substituted for those with an oxygen connecting to the tetrahedrally-coordinated silicon atom.

Alkoxysilanes are chosen because they are usually present in the synthesis. Due

to the instability of silicic acid in solution, the silica source is usually stored as an alkoxysilane, whose $O-R$ bonds (where R represents any alkane group) are hydrolysed in aqueous solution, leading to the formation of $Si-O-H$ groups, as given by **Equation 3.3**.

The description of siloxanes is also valuable because they represent the effect of polymerisation. However, we could only uncover a suitable quantity of data for one molecule in this class in the liquid phase, namely the dimer of trimethylsilanol. Finally, silanes were chosen not because they are involved in silica-synthesis processes, but rather because they are the least complicated molecules in this class and hence provide a good starting point for decoupling the effect of other functional groups from the central atom.

Table 6.1: Experimental data of various physical properties collected for a selection of organosilica compounds

Compound	Chemical Formula	Group	ρ (kg/m^3)	ΔH_{vap} (kJ/mol)	P_{vap} (Pa)
Tetramethylsilane	$Si(Me)_4$	Silane	641 [381]	26.0 ± 0.6 [382]	94,470 [383]
Tetraethylsilane	$Si(Et)_4$	Silane	763.3 [381]	39.0 ± 0.7 [382]	705.1 [384]
Tetrapropylsilane	$Si(Pr)_4$	Silane		42.2 ± 0.7 [382]	
Trimethylsilanol	$(Me)_3-Si-OH$	Silanol	814.1 [380]	45.6 [380]	2,858 [385]
Triethylsilanol	$(Et)_3-Si-OH$	Silanol	863.8 [380]	50.6 [380]	202.7 [385]
Hexamethyldisiloxane	$(Me)_3-Si-O-Si-(Me)_3$	Ether	763.8 [380]	34.7 [380]	5,539 [386]
Tetramethoxysilane	$(MeO)_4Si$	Siloxane	1052.3 [380]	41.4 [382]	1,579 [387]
Tetraethoxysilane	$(EtO)_4Si$	Siloxane	934.6 [380]	52.0 [388]	249.8 [384]
Trimethylethoxysilane	$(Me)_3-Si-O-Et$	Siloxane	757 [380]	38.4 ± 0.6 [382]	
Dimethyl-ethyl-ethoxysilane	$(Me)_2,Et-Si-O-Et$	Siloxane	890 [380]		
Diethyl-methyl-ethoxysilane	$(Et)_2,Me-Si-O-Et$	Siloxane	938 [380]		
Triethyl-ethoxysilane	$(Et)_3-Si-O-Et$	Siloxane	895 [380]	38 [380]	889.4 [380]

From **Table 6.1**, it can be seen that the only properties available for the majority of molecules are the density (ρ) and the heat of vaporisation (ΔH_{vap}). While this represents only a small subset of our ideal set of experimental properties, it is nevertheless feasible to work with this pair of properties. Both properties are easy to calculate using MD and, as mentioned above, one represents a structural parameter, while the other is energetic in nature. Additionally, vapour pressure (P_{vap}) data are available for several molecules, from which it is possible to calculate the self-solvation free energy (ΔG_{solv}), as shown by the Minnesota Solvation Database [389, 390]. This is demonstrated in **Equation 6.1**

$$\Delta G_{solv} = -2.303RT\log_{10}(\frac{24.45\rho}{M_W P_{vap}}) \tag{6.1}$$

where R is the ideal gas constant, T is the temperature, ρ is the liquid density, M_W is the molecular weight and P_{vap} is the vapour pressure. From **Equation 6.1**, we obtain the free energy of self-solvation values in **Table 6.2**.

Table 6.2: Free energy of self-solvation (ΔG_{solv}) obtained from **Equation 6.1** for selected organosilica molecules

Compound	ΔG_{solv} (kJ/mol)
Tetramethylsilane	-13.0
Tetraethylsilane	-24.0
Trimethylsilanol	-22.2
Triethylsilanol	-27.9
Hexamethyldisiloxane	-20.4
Tetramethoxysilane	-22.8
Tetraethoxysilane	-26.5
Triethyl-ethoxysilane	-24.0

The free energy of self-solvation is a useful property for our parametrisation since solution thermodynamics play a key role in the synthesis of templated mesoporous silica. When parametrising a force field, it is advisable to use most of the data for fitting, while leaving some parameters for validation. As such, it was chosen to use ρ and ΔH_{vap} for the parametrisation, since they are a lot easier and quicker to compute than ΔG_{solv}. This means that we have two fitting parameters, meaning that we can parametrise two model parameters from one set of data for a given molecule.

6.3 Parametrisation Strategy

As already mentioned in **Chapter 2**, force field parameters can be split into two categories - bonded and non-bonded interactions. The former contain terms that describe interactions between chains of bonded atoms. Usually there are three of these terms, namely bond-stretching, angle-bending and torsional interactions. These describe the intramolecular interactions of chains of two, three and four atoms, respectively. Mathematically, due to our desire for transferability and ease of use, we will employ the formulations most commonly used in the GROMACS MD simulation package. These forms were already given in **Equations 2.14**, **2.15**, **2.17**. This means that we must fit two parameters for each bond type (k_B, r_0), two for each angle (k_θ, θ_0) and four terms

for each dihedral ($k_{T,1}$, $k_{T,2}$, $k_{T,3}$ and $k_{T,4}$).

Intermolecular interactions commonly consist of two terms, namely the LJ and Coulomb potentials. Mathematically, these are given in **Equations 2.19** and **2.22**. Due to the presence of combining rules (we will use the Lorentz-Berthelot rules [51] - see **Equations 2.20** and **2.21**) it is not necessary to derive ε and σ values for each interaction, but only for each atom type. As such, per atom type there are two parameter values to be fitted. Additionally, a charge for every group has to ascertained; however, these are more susceptible to the molecular environment, so these will be determined per molecule, rather than using one set charge for a specific atom type. As was described in **Chapter 5**, we have extensively tested QM methods for obtaining molecular point charges, meaning that we can exclude these parameters from the fitting. Not that any other choice is available given the limited number of data points.

This leaves us with ε, σ and the intramolecular parameters. Having surveyed the literature, we discovered an adequate quantity of data from quantum calculations for the bond-stretching and angle-bending parameters, obviating the need to calculate them in-house. However, this nevertheless leaves the torsion terms, by far the most complicated of the bonded interactions. For this reason, it was decided to calculate these from in-house QM calculations. One of the reasons for this is that some of the few torsional terms available in the literature include 1-4 interactions. This is when the intermolecular interaction between atoms at the end of a four atom chain is included, fully or partially, blurring the boundary between intermolecular and intramolecular terms. For example, OPLS scales the non-bonded interactions between 1-4 atoms by 0.5 [391], whereas in the CHARMM force field they have their full values, i.e. the scaling factor is 1.0 [392]. While this can improve a torsional potential, it also makes it less transferable since it includes the intermolecular terms, which may be different between different molecules. For this reason, we saw the chance to calculate our own torsional potentials as an opportunity to ensure their transferability.

Despite having now laid out how we will obtain each type of interaction in **Table 6.3**, we have not yet enumerated what these specifically are. Before this, it is important to note that we are intending to create a united-atom (UA) force field; this means that

Table 6.3: Table demonstrating how the parameters for each interaction for the force field will be obtained.

Potential	Components	Method of Parametrisation
Bond-Stretching	k_B, r_0	From the literature
Angle-Bending	k_θ, θ_0	From the literature
Torsional	$k_{T,1}$, $k_{T,2}$, $k_{T,3}$, $k_{T,4}$	Quantum-mechanical calculations
Electrostatic	q	Quantum-mechanical calculations
Lennard-Jones	ε, σ	Fitting to experimental data

most light atoms, i.e. hydrogen, will, where possible, be clumped together into one interaction centre with the heavy atom that they are connected to. For this reason, we will use the term interaction centre rather than atom, since some interaction centres will represent collections of atoms. This is generally appropriate where there is no, or relatively little, directional bias introduced by a hydrogen atom. For this reason, the hydrogen atoms in alkyl groups will be grouped into one interaction centre centred on the carbon atom, since they are distributed relatively symmetrically around the carbon atom and will produce near-identical interactions. However, this is less appropriate for the hydrogen in hydroxyl groups because there is only one hydrogen and it has a significant charge, whose effect will not be averaged out by other hydrogen atoms connected to the oxygen. Since these are the only two environments containing hydrogen in our list of training molecules, it means our only non-atomistic interaction centres will be CH_2 and CH_3. The advantage of this UA implementation is that it greatly reduces the number of interaction centres, and hence computational cost. For example, the molecule trimethylsilanol in our training set will go from having 15 to 6 interaction centres, while losing minimal chemical resolution.

The UA philosophy has been widely used, for example in the TraPPE force field [393,394], and shown to be very effective. The success of TraPPE with this approach is especially pertinent for our study since their philosophy is based upon ensuring the transferability of their parameters, something that we also wish to achieve. As such, it was decided to make use of the already existing alkane parameters instead of starting from scratch. However, for CH_2 and CH_3, we will use the modified TraPPE parameters for alkanes as introduced in the work of Jorge et al., where it was shown that a slight

modification of the LJ parameters can eliminate a systematic error in the self-solvation free energy of alkanes [327,328].

To summarise, the following interaction centres (as shown in **Table 6.4**) for LJ parameters would have to be parametrised in order to be able to describe all the molecules shown in **Table 6.1**. It should be noted that while hydroxyl hydrogen (H_O) is an interaction centre, due to its small size and in keeping with the philosophy used in many water models and TraPPE, it has a charge but no LJ parameters.

Table 6.4: Enumeration of the Lennard-Jones parameters that would be necessary to parametrise in order to describe all the organosilica molecules presented in **Table 6.1**

Interaction Centre	Symbol	Comments
Methyl	Me	Taken from modified TraPPE [328]
Methylene	Me_2	Taken from modified TraPPE [328]
Tetrahedral silicon	Si	
Hydroxyl oxygen	O_H	
Alkoxide oxygen	O_C	
Siloxane oxygen	O_{Si}	

From **Table 6.4**, it is apparent that there are four interaction centres for which the LJ parameters need to be ascertained. For each said interaction centre, at least one molecule from the training set needs to be used for the fitting. The simplest possible molecules for the fitting of each centre are shown in **Table 6.5**

Table 6.5: Enumeration of the molecules used in the training set to parametrise each Lennard-Jones interaction centre and any prerequisite Lennard-Jones parameters.

Interaction Centre	Molecule	Prerequisites
Si	$Si-(Me)_4$	Me
O_H	$(Me)_3-Si-OH$	Si, Me
O_C	$Si-(OMe)_4$	Si, Me
O_{Si}	$(Me_3)_3-Si-O-Si-(Me)_3$	Si, Me

Having now defined the global methodology for creating the force field, we can move on to the parameters themselves. First we will delve into the literature to find bond-stretching and angle-bending terms. Subsequently, the procedure behind the QM calculations for obtaining the data for dihedral torsion and deriving functional constants therefrom will be elucidated. The results for the point charges, once again from QM cal-

culations, will thereafter be presented and discussed. Finally, the procedure and results for the fitting of the LJ parameters using MD simulations for some of the molecules in the training set will be demonstrated.

6.4 Bond-Stretching and Angle-Bending Parameters

The bulk of the parameters were found in two papers by Grigoras and Lane [395] and Abraham and Grant [396]. These were the most comprehensive studies that could be found; they contained QM calculations for approximately 25 and 15 organosilica molecules, respectively, covering all of the different chemical families studied in this work. The only concern is that these studies are relatively antiquated, both having been published in 1988. For this reason, and due to the low level of theory employed (basis set: 3-21G*), it was considered necessary to cross-reference their values against more modern calculations, in order to check that their results are still valid.

Table 6.6: Enumeration of the relevant equilibrium bond lengths necessary for our organosilica force field. Adapted from Grigoras and Lane [395]

Bond	Length (nm)	Molecule
$Si-C$	0.1896	H_3Si-CH_3
$Si-O_H$	0.1653	H_3Si-OH
$Si-O_C$	0.1656	$(CH_3)_2SiH-O-CH3$
$Si-O_{Si}$	0.1645	$H_3Si-O-SiH_3$

In **Table 6.6**, the $Si-C$ bond length is given as 0.1896 nm, based on methylsilane. Nevertheless, methylsilane is not one of our test molecules (it is gaseous at STP). However, an experimental value for tetramethylsilane ($Si(CH_3)_4$), the first molecule for which we need to obtain force field parameters, is available and is 0.1875 $\pm$ 0.0002 nm [397]. Since this value is known experimentally, it was decided to use the bond-length value from this source for the force field.

Next, the values for $Si-O_H$ are compared. A later study by Koput [398] at a higher level of theory (CCSD(T), with basis sets up to cc-pV5Z) studied methanol and silanol. At the highest level of theory, the $Si-O_H$ bond length was given as 0.16481 nm, while in **Table 6.6** the value is 0.1653 nm. While there is some difference here,

the value for the calculation with the second-largest basis set (cc-pVQZ) is 0.16525 nm, which matches up almost exactly with the value of Grigoras and Lane. However, this value is for silanol, rather than trimethylsilanol, which is the simplest silanol that we're interested in. Trimethylsilanol is a much bulkier molecule and therefore the value of the bond length would be larger. This is shown in Grigoras and Lane's study [395]; apart from silanol, they also carried out calculations for methylsilanol ($CH_3 - SiH_2 - OH$) and found that it had an $Si - O_H$ bond length of 0.1657 nm, i.e. a longer one. Taking all this into consideration, we believe that the value of 0.1653 nm from Grigoras and Lane is a good approximation of the $Si - O_H$ bond length.

For the length of the siloxane bond, Grigoras and Lane report a value of 0.1645 nm [395], while Abraham and Grant cite 0.164 nm [396]. There are a pair of experimental values for disiloxane available in the literature [399,400] and these show that the length of the siloxane bond is either 0.1634 ± 0.0002 or 0.1631 nm. As such, we can see that the value of Abraham is closer to the experiment value. However, as was previously discussed, our molecules are more heavily substituted than the molecules studied in those papers; the molecule in our training set is hexamethyldisiloxane and not disiloxane. As such, we would expect the bond length to be larger. Indeed, if adding a methyl group made the same difference to the bond length as in silanols, then each methyl group would add 0.0003 nm. Therefore, we could tentatively extrapolate from the disiloxane values, of 0.1631 and 0.1634 nm, respectively, a bond length for hexamethyldisiloxane between 0.1640 and 0.1643 nm. Since the value of Abraham and Grant falls within this range, we shall use this value.

Finally, we move onto discussing the bond length of $Si - O_C$. The value from Grigoras and Lane [395] is 0.1656 nm for $(CH_3)_2SiH - O - CH_3$. They also provide a value for $CH_3SiH_2 - O - CH_3$, namely 0.1652 nm. This shows once again how adding methyl groups lengthens $Si - O$ bonds. Since $(CH_3)_2SiH - O - CH_3$ is chemically very similar to one of our training molecules (a substitution of the hydrogen bonded to the silicon atom with a methyl group would result in tetramethyl orthosilicate), we will use the value for this bond, i.e. 0.1656 nm.

Having decided on equilibrium bond lengths, it is now necessary to also decide on

the stretching constants. These were taken from the study of Grigoras and Lane [395] and modified as follows. The functional form applied in their work was anharmonic:

$$U_B = \frac{k_B}{2}(r - r_0)^2(1 + k_{CS}(r - r_0)) \tag{6.2}$$

The only difference here is the addition of a cubic stretch term, whose strength is dictated by k_{CS}. Since we intend to use a harmonic potential, we left out the cubic stretch term in order for it to have the same form as in **Equation 2.14**. A comparison of the full anharmonic and only the harmonic part of the bond-stretching potential of Si-C from the study of Grigoras and Lane [395] can be seen in **Figure 6.1**.

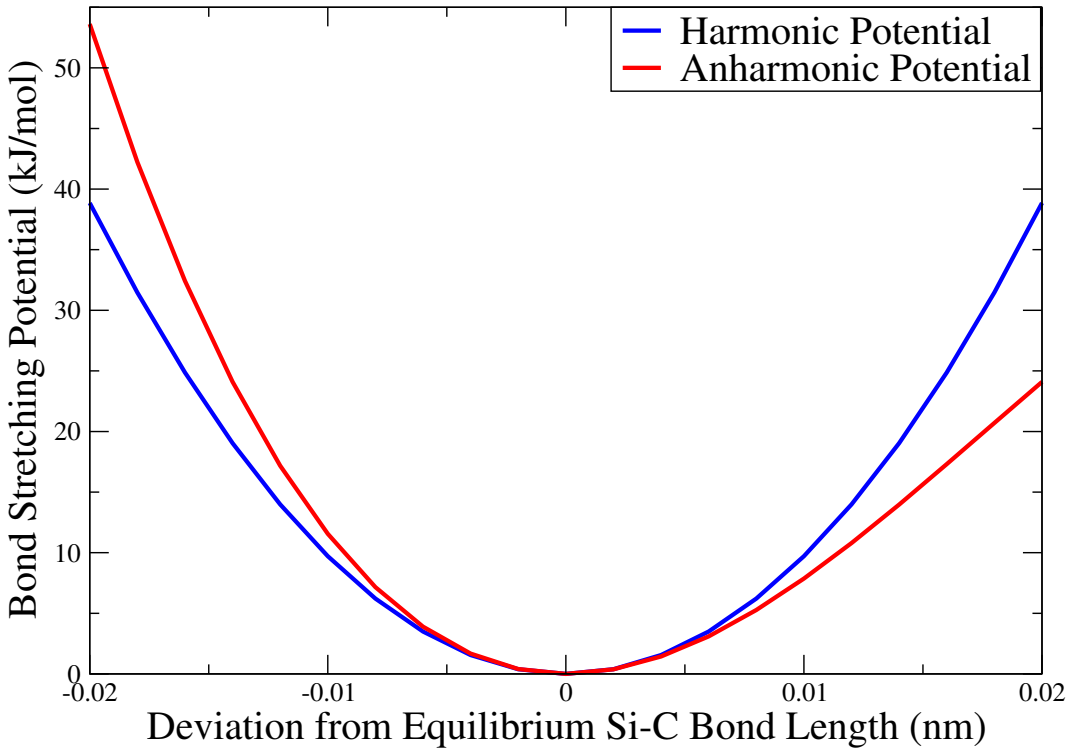


Figure 6.1: Comparison of the full anharmonic and only the harmonic component of the potential for stretching of the Si-C bond. Data taken from Grigoras and Lane [395]

As can be seen in **Figure 6.1**, the anharmonic part of the potential only has any significant effect for values far from the equilibrium distance. At a bond length deviation of 5%, i.e. a large deviation, the difference was around 15% and the harmonic potential

was hence seen as an acceptable simplification. Additionally, since there were only data available for the stretching constant of one $Si-O$ bond, the k_B value is the same for all three types of $Si-O$ bond. The final collection of bond-stretching parameters can be found in **Table 6.7**.

Table 6.7: Finalised bond-stretching parameters for our organosilica force field

Bond	Length (nm)	Force Constant (kJ/mol.nm^2)
$Si-C$	0.1875	194,200
$Si-O_H$	0.1653	381,300
$Si-O_C$	0.1656	381,300
$Si-O_Si$	0.1640	381,300

Having dealt with the bond-stretching parameters, we can now move on to the values for angle bending. The values for the angles in question can be seen in **Table 6.8**.

Table 6.8: Enumeration of the relevant equilibrium angles necessary for our organosilica force field. Data taken from Grigoras and Lane [395]

Angle	Value (deg)	Molecule
$C-Si-C$	112.0	$CH_3-SiH_2-CH_3$
$Si-C-C$	111.5	$H_3Si-CH_2-CH_3$
$C-Si-O_H$	107.4	H_3C-SiH_2-OH
$Si-O-H$	115.5	H_3CSiH_2-O-H
O_H-Si-O_H	104.4	$OH-SiH_2-OH$
$Si-O-Si$	149.5	$H_3Si-O-SiH_3$
$Si-O-C$	124.8	$(CH_3)_2SiH-O-CH_3$
O_C-Si-C	111.0	$CH_3-CH_3SiH-OCH_3$
O_C-Si-O_C	105.7	$OCH_3-SiH_2-OCH_3$

Due to the much higher number of angles, we will focus our discussion on the accuracy of three of them for the sake of clarity: $C-Si-C$, $Si-O-H$ and $Si-O-Si$. For the first of these, an experimental value from an electron diffraction study [397] is available and is 110.5 $\pm$ 0.4 deg. However, this value is for hexamethyldisilane, whereas the value in **Table 6.8** is from dimethylsilane ($C-Si-C$ angle = 112 deg). Despite the fact that hexamethyldisilane is a much bulkier molecule, the two values are fairly close, just over one degree once experimental error is taken into account. As such, the value of 112 deg is most likely a suitable value.

Secondly, as already mentioned above in our discussion of bonds, silanol has been

studied at a much higher level of theory, namely CCSD(T) [398]. The $Si-O-H$ angle from this study has a value of 117.91 deg, compared to Grigoras and Lane's 115.5 deg for a slightly bulkier silanol (methylsilanol). Since the values are fairly close and the molecule that Grigoras and Lane studied is closer to the molecules that we wish to parametrise, we chose to stick with their value for the $Si-O-H$ angle.

Finally, we considered the $Si-O-Si$ angle. This is one of the more important angles, since it is known to be extremely large and have a very low bending energy. In fact, the energy needed to linearise the molecule, i.e. distort the equilibrium angle to 180 deg, is 1.3 kJ/mol [395,401], well below the energy of thermal fluctuations at room temperature ($\approx$ 2.5 kJ/mol). This means that there can be a large range of angles far from the equilibrium value in solution. Keeping this in mind, if there is one angle for which it may be prudent to consider using a more complex angle-bending function, it would be the siloxane angle. Indeed, Grigoras and Lane themselves used a different angle functional for $Si-O-Si$ (cubic rather than to the power of six). However, we will stick with the harmonic angle-bending function in the first instance and will only implement a more complex function if strictly necessary. Grigoras and Lane's value of 149.5 deg compares well with an experimental value of 151.2 deg and more recent theoretical work pointing to a value of 150.2 [402].

As was the case for the bond potentials, the form of the angle potential in the study of Grigoras and Lane [395] includes an anharmonic term. More precisely, it has the form

$$U_B = \frac{k_\theta}{2}(\theta - \theta_0)^2(1 + k_6(\theta - \theta_0)^4) \tag{6.3}$$

where k_6 is the force constant for the sixth-order anharmonic term. We present a comparison of the harmonic part and the full anharmonic potential in **Figure 6.2** for the angle $C-Si-C$.

In **Figure 6.2**, it can be seen that the harmonic potential fits well up to a deviation of the equilibrium angle of about 10 degrees, where the energetic barrier is approximately 10 kJ/mol. Since this is a good deal larger than the scale of thermal fluctuations at room temperature, this is most likely a reasonable approximation. As such, the final

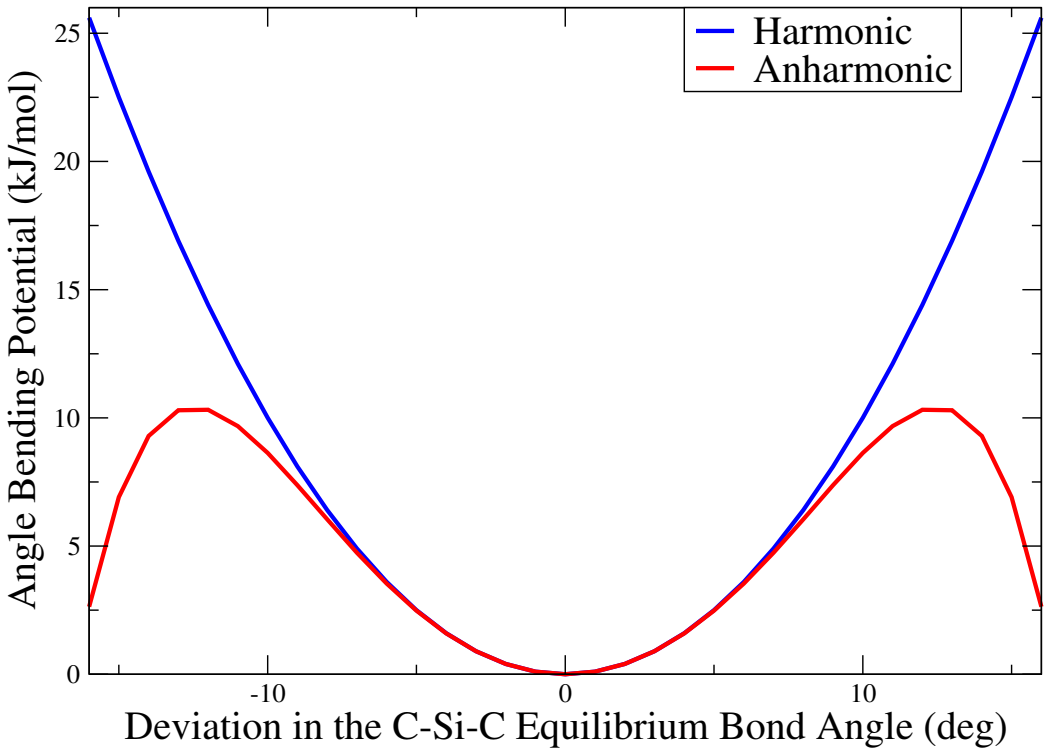


Figure 6.2: Comparison of the full anharmonic and only the harmonic component of the potential for bending of the $C-Si-C$ angle from Grigoras and Lane [395]

collection of parameters can be seen in **Table 6.9**. One point to note is that although Grigoras and Lane provided an angle for $C-Si-O_C$, they did not include a force constant. We have therefore assumed that the force constant for $C-Si-O_H$ is an acceptable substitute.

This concludes our survey of the bond-stretching and angle-bending intramolecular parameters and we will now proceed to our own calculations for the final intramolecular potential, namely dihedral torsion.

6.5 Dihedral Potentials

The work of Grigoras and Lane [395] did not include any dihedral potentials, so it was left to us to calculate these in-house. The full list of necessary dihedrals can be seen in **Table 6.10**. For the sake of simplicity, we have treated CH_3 and CH_2 as the same

Table 6.9: Enumeration of the relevant equilibrium angles necessary for our organosilica force field. Data taken from Grigoras and Lane [395]

Angle	Value (deg)	Force Constant ($kJ/mol.rad^2$)
$C-Si-C$	112.0	656.2
$Si-C-C$	111.5	726.5
$C-Si-O_H$	107.4	774.6
$Si-O-H$	115.5	257.8
O_H-Si-O_H	104.4	872.6
$Si-O-Si$	149.5	61.3
$Si-O-C$	124.8	298.1
$C-Si-O_C$	111.0	774.6
O_C-Si-O_C	105.7	795.0

environment for the purposes of the dihedrals; the difference will likely be minimal and this will appreciably decrease the number of necessary calculations. In the first part of this section, we will explain our methodology. In the second half, we will present a case study for the $O-Si-O-H$ dihedral.

Table 6.10: Enumeration of the relevant dihedral potentials necessary for our organosilica force field

Dihedral	Molecule
$H-O-Si-C$	$CH_3-Si(CH_3)_2-O-H$
$H-O-Si-O_H$	$H-O-Si(CH3)_2-OH$
$C-C-Si-O_H$	$CH_3-CH_2-Si(CH_3CH_2)_2-OH$
$C-C-Si-C$	$CH_3-CH_2-SiCH_3CH_2OH-CH_2CH_3$
$C-C-C-Si$	From TraPPE
$C-Si-O-Si$	$CH_3-Si(CH_3)_2-O-Si(CH_3)_3$
$C-Si-O-C$	$CH_3-Si(CH_3)_2-O-CH_2CH_3$
$C-O-Si-O_C$	$CH_3-O-Si(OCH_3)_2-OCH_3$
$C-C-O-Si$	$CH_3-CH_2-O-Si(OCH_2CH_3)_3$

In an ideal situation, we would be able to recover the same energy for dihedral torsion from our force field as from a QM calculation. Specifically, a DFT calculation in our case. Mathematically, we can write this as

$$\Delta E_{DFT} = \Delta E_{FF} = \Delta E_B + \Delta E_\theta + \Delta E_T + \Delta E_{LJ} + \Delta E_{ES} \tag{6.4}$$

where ΔE represents the energy barrier compared to the dihedral angle corresponding to the energetic minimum and FF represents our classical force field. We can then

rearrange for the dihedral energy

$$\Delta E_T = \Delta E_{DFT} - \Delta E_B - \Delta E_\theta - \Delta E_{LJ} - \Delta E_{ES} \tag{6.5}$$

Subtracting these other interactions is important, since they can be of a higher magnitude than the energetic barrier of the dihedral itself, thereby cloaking the actual behaviour of the dihedral. We present an example of this from our work in **Figure 6.3**. This calculation represents the $O - Si - O - H$ dihedral in dimethylsilanediol. In the DFT calculation in **Figure 6.3**, the bonds and angles were held constant; we can therefore state that ΔE_B and ΔE_θ are always equal to zero. Additionally, the only intermolecular interactions within the molecule involve hydroxyl hydrogen atoms. Since in our force field hydroxyl hydrogens only have a point charge and no LJ parameters, we can also say that $E_{LJ} = 0$ for all dihedral angles and this is then also true for ΔE_{LJ}. Taking these considerations into account, **Equation 6.5** becomes

$$\Delta E_T = \Delta E_{DFT} - \Delta E_{ES} \tag{6.6}$$

All three of the components in **Equation 6.6** are shown in **Figure 6.3**. We can see how the electrostatic interactions alone overwhelm the energy from the DFT dihedral scan. As such, it can be seen that this changes the locations of the maximum and minimum of the curve in comparison to those from the DFT energy alone. This demonstrates the importance of taking the other force field components into account when attempting to obtain the dihedral energy barrier.

However, there are also often multiple dihedral angles involved, hence the total dihedral energy is actually the sum of all of them

$$\Delta E_T = \sum_{i=1} \Delta E_{T,i} \tag{6.7}$$

For all the dihedrals here, where Si is one of the central atoms, there will be three dihedrals because silicon is tetrahedral, i.e. one branch will be rotating past the other three. If these dihedrals are not of the same type, then the others will also need to be subtracted off first. For example, if we were to parametrise a molecule with two A

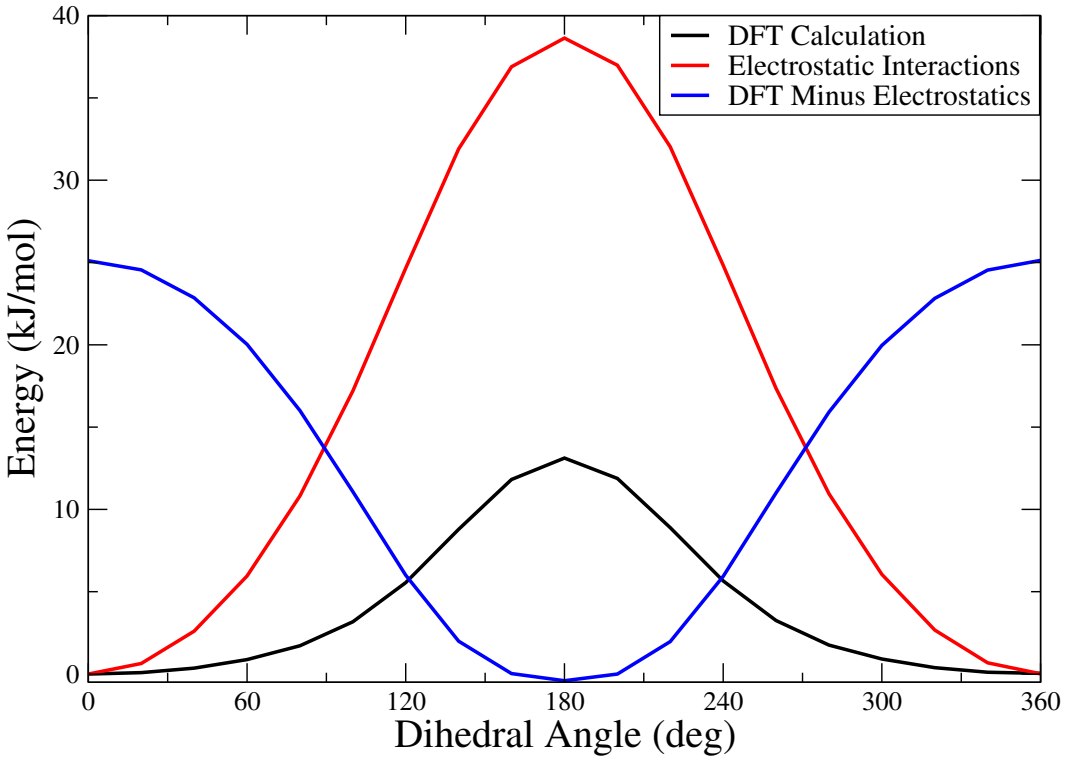


Figure 6.3: Comparison of the magnitude of the relative torsional and electrostatic forces at different dihedral angles in the dihedral scan of $O-Si-O-H$ in dimethylsilanediol

dihedrals as well as one of type B, then the value for a single dihedral A will be

$$\Delta E_{T,A} = \frac{\Delta E_T - \Delta E_{T,B}}{2} \tag{6.8}$$

Having explained our methodology, we will now move on to a case study. Our case study focusses on two molecules, namely dimethylsilanediol ($Si(CH_3)_2(OH)_2$) and methylsilanetriol ($SiCH_3(OH)_3$). While these molecules are not present in our training set, since both are solid at STP, they act as a continuum between trimethylsilanol, which is in our training set, and silicic acid, which we want to be able to describe via the transferability of our force field. Therefore, we want a dihedral that can describe the torsional behaviour of both these molecules. Additionally, there are two sets of parameters for this dihedral already available in the literature. These are summarised in **Table 6.11**. The comparison of the effectiveness of the fitted dihedral function

between two different molecules (dimethylsilanediol and methylsilanetriol), as well as against other two sets of dihedral parameters already existing in the literature, should provide a rigorous test for our calculations.

Table 6.11: The Ryckaert-Bellemans potential for the $O-Si-O-H$ dihedral from the studies of Pereira et al. [161] and Jorge et al. [35]. N.B. The values for Pereira et al. are converted from the functional form given in reference [161]. All values are given in kJ/mol

Study	$k_{T,0}$	$k_{T,1}$	$k_{T,2}$	$k_{T,3}$
Pereira et al. [161]	10.73	10.00	4.36	1.027
Jorge et al. [35]	14.8473	9.1554	-3.6233	2.0686

For the quantum calculations, the Gaussian09 software package was used. The DFT calculations employed an aug-cc-pVTZ basis set and used the B3LYP exchange-correlation functional. The scan of the dihedral consisted of 18 steps of 20 degrees around the dihedral, i.e. one full revolution. All bonds and angles were frozen, i.e. were not allowed to change after the initial optimisation. As mentioned above, we did this so that we would not have to subtract off these energies from the DFT energy. The methyl hydrogen atoms were an exception; their positions were not fixed and were optimised. The reason for this is that we intend to parametrise a united-atom force field, where methyl groups are represented by one interaction centre. Therefore, by allowing the positions of the methyl hydrogens to be optimised, we are able to take some of this behaviour into account in the force field implicitly. The atom being rotated around one of the $Si-O$ bonds was a hydroxyl hydrogen. For subtracting the intermolecular interactions (of which there were only electrostatic components in these two systems), we used the CHELPG charges, which will be introduced in the next section, on the atoms in trimethylsilanol as representative of the two silanols given here. See **Section 6.6** and **Table 6.14** for further information. For the dihedrals of Pereira et al. and Jorge et al., we used the charges given in their papers [35, 161], respectively.

In these molecules there is also a second dihedral type present, namely $C-Si-O-H$. This was parametrised from trimethylsilanol (see fit in **Figure 6.4**), since this is a well-behaved system; there is only one type of dihedral in this molecule (at least in our united-atom representation - we are not interested in any dihedrals containing methyl

hydrogens) and the three of them are arranged symmetrically. The same QM calculation procedure was described above, with the exception that we used 36 steps of 10 degrees instead of 18 20-degree ones. The values obtained for the Ryckaert-Bellemans dihedral potential were (in kJ/mol): $k_{T,0} = 0.87; k_{T,1} = 2.6; k_{T,2} = 0; k_{T,3} = -3.47$. In the diol there are two $C-Si-O-H$ dihedrals and one $O-Si-O-H$, whereas the triol has one $C-Si-O-H$ and two $O-Si-O-H$ dihedrals.

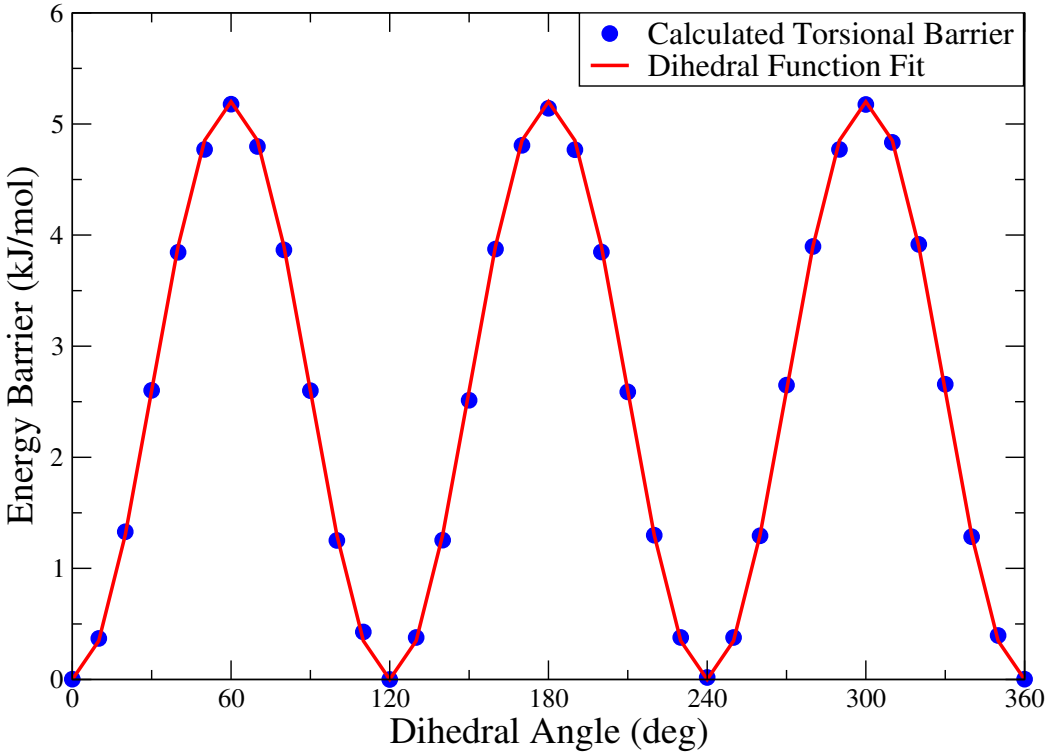


Figure 6.4: Fit of the Ryckaert-Belleman dihedral function to the calculated $C-Si-O-H$ torsional barrier of trimethylsilanol

The dihedral parameters were then fitted so that the dihedral functional plus the electrostatic interactions reproduced the curves from both the DFT calculations as well as possible with the same parameters. The fits for our parameters, as well as those from Pereira et al. and Jorge et al., can be seen in **Figures 6.5** and **6.6** for the diol and triol, respectively. The values for the dihedral parameters that minimised the total squared error for both the diol and triol were obtained via a Python script. The optimum values were (in kJ/mol): $k_{T,0} = 10.92; k_{T,1} = 8.41; k_{T,2} = 1.36; k_{T,3} = 4.35$. The values for the

total squared error for the diol and triol over all 19 points were 40 and 30 kJ^2/mol^2, respectively.

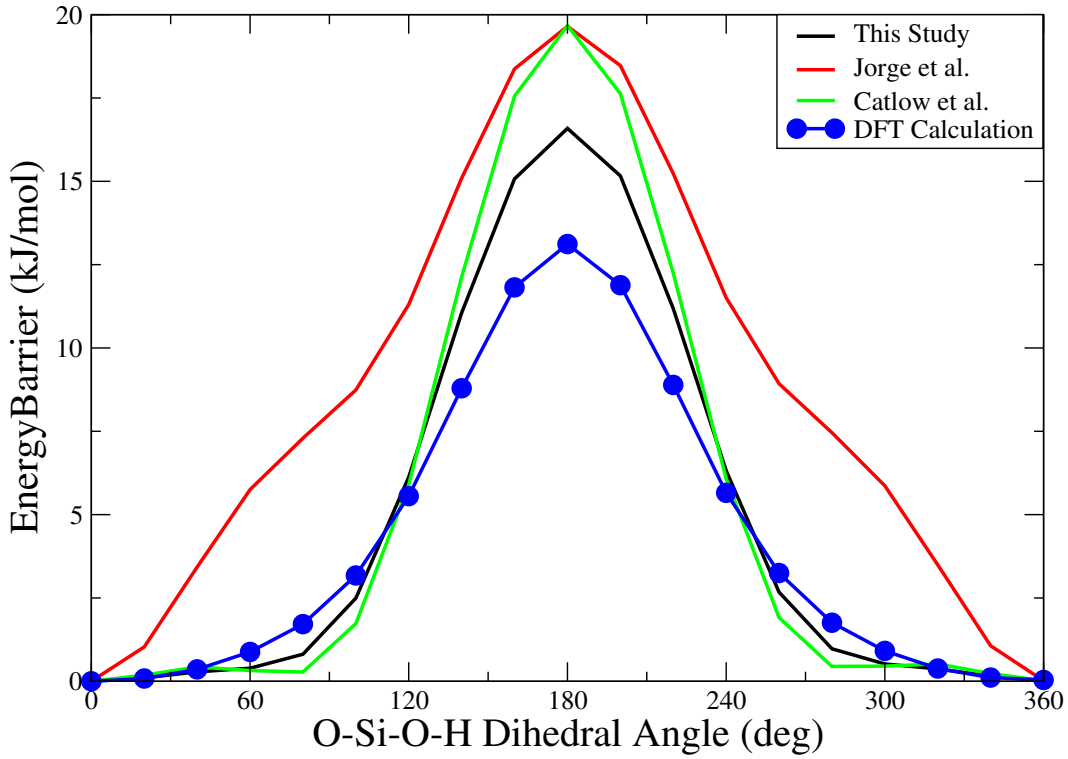


Figure 6.5: Comparison the $O - Si - O - H$ dihedral parameters fitted in this work against the result from a DFT dihedral scan of dimethylsilanediol and the parameters from two other works, namely those of Jorge et al. [35] and Pereira et al. [161]

From **Figure 6.5**, it is clear that the parameters from this study provide the best result. The study of Jorge et al. performs worst, not getting the shape of the curve correct and having the maximum furthest from the DFT value. The dihedral of Pereira et al. has about the same maximum as that of Jorge et al. but the shape of the curve is much more accurate. However, this is to the point of exaggeration; the values close to the minimum are too shallow, whereas the values near the maximum increase too rapidly. Near the minimum and the maximum, the values obtained in this study perform best. However, this is not surprising, since they were in part parametrised against the target DFT energies for this molecule. The same is true for the triol results; the values from this study are the most accurate, although none of the three results

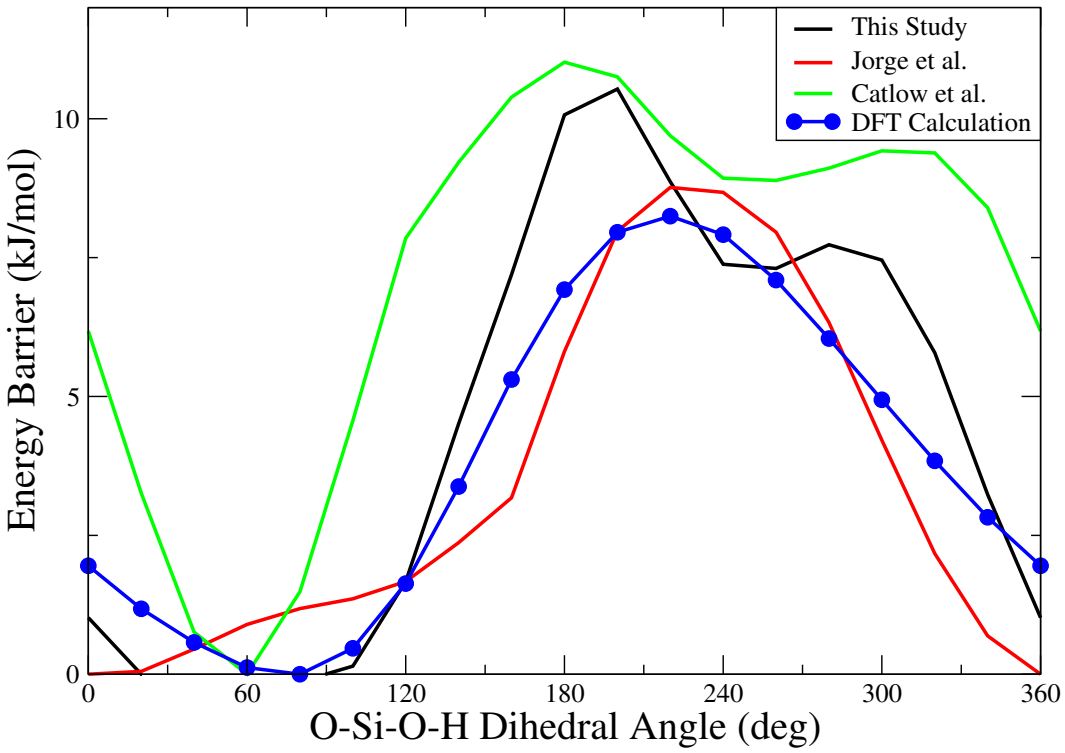


Figure 6.6: Comparison the $O-Si-O-H$ dihedral parameters fitted in this work for against the result from a DFT dihedral scan of methylsilanetriol and the parameters from two other works, namely those of Jorge et al. [35] and Pereira et al. [161]

can capture the behaviour of the curve fully. The results of Jorge et al. capture the behaviour of the maximum best, but are easily the worst for the minimum. The shapes of the curve for this study and that of Pereira et al. are very similar, having an unexpected local minimum between two local maxima, whereas the DFT calculation only has one. However, the curve of Pereira et al. has a very sharp minimum too, i.e. much sharper than that of the DFT calculation, so is probably the worst result. The dihedral parameters from this study do the best job of describing the minimum, which is the most important region. Altogether, the fits from study, while far from being perfect, do describe the dihedral in these two complicated molecules adequately and demonstrate the efficacy of the calculation and fitting procedure. The parameters for the two dihedrals parametrised in this work are shown in **Table 6.12**. The raw data for all the DFT calculations in this section can be found in **Appendix A**.

Table 6.12: Summary of the Ryckaert-Bellemans parameters obtained for the dihedrals studied in this work. All values are given in kJ/mol

Dihedral	$k_{T,0}$	$k_{T,1}$	$k_{T,2}$	$k_{T,3}$
$C-Si-O-H$	0.87	2.60	0.00	-3.47
$O-Si-O-H$	10.92	8.41	1.36	4.35

Having now developed and validated a procedure for calculating the torsional parameters, thereby completing the intramolecular parameters, we can now move onto obtaining parameters for the intermolecular interactions.

6.6 Charge Determination

The quantum calculations for the organosilica molecules were also carried out using Gaussian09. The geometries of the molecules were optimised using an aug-TZ basis set and the B3LYP exchange-correlation functional in the IEFPCM CSM. For those molecules whose dielectric constant was known (which was only the case for tetramethylsilane (7.17 F/m) and tetraethylsilanol (2.66 F/m)), a solvent of near-identical dielectric constant was chosen as the continuum background. Due to the lack of organosilica solvents available in Gaussian09, the solvent for the rest of the calculations was chosen to be the most similar organic molecule and these are listed in **Table 6.13**.

Table 6.13: List of the molecules used as the solvent in the IEFPCM continuum solvation model for the quantum-mechanical calculation of each organosilica molecule. ϵ and ϵ_∞ represent the static and high-frequency dielectric constants of the solvent, respectively

Organosilica Molecule	Continuum Molecule	ϵ	ϵ_∞
Tetramethylsilane	n-Nonane	1.9605	1.975149
Tetraethylsilane	n-Nonane	1.9605	1.975149
Trimethylsilanol	2,6-Dimethyl pyridine	7.1735	2.235922
Triethylsilanol	Pentanoic acid	2.6924	1.983872
Silicic Acid	Water	78.3553	1.777849
Hexamethyldisiloxane	Dibutyl ether	3.0473	1.957761
Tetramethylorthosilicate	Diethyl ether	4.2400	1.829527
Tetraethylorthosilicate	Diethyl ether	4.2400	1.829527
Trimethylethoxysilane	Diisopropyl ether	3.3800	1.871150
Dimethyl,ethylethoxysilane	Diisopropyl ether	3.3800	1.871150
Diethyl,methylethoxysilane	Dibutyl ether	3.0473	1.957761
Triethylethoxysilane	Dibutyl ether	3.0473	1.957761

From these calculations CHELPG and DDEC charges were fitted. The results are shown in **Table 6.14**. Since we intend to parametrise a united-atom force field for these molecules, the charges for the atoms in alkane groups have been summed into an overall charge located on the carbon atom. The molecules were selected in order to represent as wide a range of organosilica molecules as possible. The chosen molecules encompass silane, silanol, alkoxysilane and siloxane environments, as discussed previously.

In **Table 6.14**, several trends are readily apparent. First of all, the charges for the DDEC fitting are significantly larger; this is surprising since this trend was not observed for methanol in **Chapter 5**. In some cases, the charges for the tetrahedral silicon atom are 2-4 times larger than the equivalent CHELPG charges. This is most likely due to the fact that the silicon atom is never close to the molecular surface; as explained earlier, one of the motivations for the DDEC procedure was to parametrise buried atoms more consistently. This can be clearly seen from the data. The last four molecules in **Table 6.14** constitute a series where the central silicon atom becomes progressively more buried; the first molecule has three methyl groups in addition to its ethoxy group, whereas the last has three ethyl groups. For the CHELPG charges, there is a marked decay in the magnitude of the silicon charge as the number of ethyl groups increases, while this effect is significantly smaller for DDEC.

However, it is not only the buried atoms whose charges are larger. Looking at the silanol molecules (tetramethyl-, tetraethyl- and silicic acid), which have exposed hydroxyl groups, the charge on the hydroxyl oxygen is around 10—20% larger for DDEC than CHELPG. Although the DDEC fitting procedure seems to result in point charges that are more chemically consistent, they seem to be rather large. For example, they are noticeably larger than those present in the INTERFACE force field [379] that deals with solid surfaces, where one would expect that the point charges would be higher. Since there is no clear consensus as to which fitting method is the most effective for our purposes, the charges from both methods will be carried forward into the initial parametrisation of our force field.

Table 6.14: Point charges obtained using both the CHELPG (indicated by C in the "fit" column) and DDEC (D) fitting procedures for 12 organosilica molecules. The calculations were carried out using an aug-TZ basis set and the B3LYP functional in Gaussian09. The IEFPCM continuum solvation model was applied to the calculations. The molecule chosen for the continuum model for each organosilica molecule are listed in **Table 6.13**

Molecule	Fit	q_{Si}	q_{CH_3}	-	-	-	-	-
Tetramethyl-	C	0.888	-0.222	-	-	-	-	-
silane	D	0.960	-0.240	-	-	-	-	-
Molecule	Fit	q_{Si}	q_{CH_2}	q_{CH_3}	-	-	-	-
Tetraethyl-	C	0.260	-0.0091	-0.0559	-	-	-	-
silane	D	0.975	-0.2875	0.04375	-	-	-	-
Molecule	Fit	q_{Si}	q_{CH_3}	q_O	q_H	-	-	-
Tetramethyl-	C	1.086	-0.224	-0.827	0.413	-	-	-
silanol	D	1.451	-0.329	-0.908	0.444	-	-	-
Molecule	Fit	q_{Si}	q_{CH_2}	q_{CH_3}	q_O	q_H	-	-
Tetraethyl-	C	0.646	-0.110	0.001	-0.7157	0.3967	-	-
silanol	D	1.237	-0.333	0.0605	-0.849	0.4295	-	-
Molecule	Fit	q_{Si}	q_O	q_H	-	-	-	-
Silicic	C	1.448	-0.7933	0.4313	-	-	-	-
Acid	D	2.10	-1.01	0.485	-	-	-	-
Molecule	Fit	q_{Si}	q_O	q_{CH_3}	-	-	-	-
Hexamethyl-	C	1.056	-0.744	-0.228	-	-	-	-
disiloxane	D	1.458	-0.9150	-0.3335	-	-	-	-
Molecule	Fit	q_{Si}	q_O	q_{CH_3}	-	-	-	-
Tetramethyl-	C	1.148	-0.477	0.190	-	-	-	-
orthosilicate	D	1.852	-0.719	0.256	-	-	-	-
Molecule	Fit	q_{Si}	q_O	q_{CH_2}	q_{CH_3}	-	-	-
Tetraethyl-	C	1.872	-0.534	0.312	-0.246	-	-	-
orthosilicate	D	1.878	-0.758	0.326	-0.0375	-	-	-
Molecule	Fit	q_{Si}	q_O	$q_{CH_2-(O)}$	$q_{CH_3-(O)}$	q_{CH_3}	-	-
Trimethyl-	C	0.974	-0.629	0.332	-0.062	-0.205	-	-
ethoxysilane	D	1.390	-0.678	0.294	-0.046	-0.320	-	-
Molecule	Fit	q_{Si}	q_O	$q_{CH_2-(O)}$	$q_{CH_3-(O)}$	q_{CH_2}	$q_{CH_3-(CH_2)}$	q_{CH_3}
Dimethyl,ethyl-	C	0.831	-0.592	0.336	-0.0683	-0.126	-0.0085	-0.1861
ethoxysilane	D	1.329	-0.664	0.2915	-0.0448	-0.332	0.0583	-0.319
Molecule	Fit	q_{Si}	q_O	$q_{CH_2-(O)}$	$q_{CH_3-(O)}$	q_{CH_2}	$q_{CH_3-(CH_2)}$	q_{CH_3}
Diethyl,methyl-	C	0.6378	-0.567	0.332	-0.0712	-0.0304	-0.0515	-0.168
ethoxysilane	D	1.258	-0.645	0.289	-0.0404	-0.3282	0.0519	-0.309
Molecule	Fit	q_{Si}	q_O	$q_{CH_2-(O)}$	$q_{CH_3-(O)}$	q_{CH_2}	q_{CH_3}	-
Triethyl-	C	0.434	-0.482	0.323	-0.0746	-0.0414	-0.0254	-
ethoxysilane	D	1.194	-0.638	0.287	-0.0432	-0.322	0.0554	-

6.7 Fitting Lennard-Jones Parameters from Molecular Dynamics Simulations

Finally, we move on to the parametrisation of the LJ parameters. This was done by manipulating the ε and σ parameters until the model matched the experimental density (ρ_{exp}) and heat of vaporisation ($\Delta H_{vap,exp}$) for each target molecule. Practically, this was done by creating one or multiple two-dimensional grids linking input parameters and predicted properties. This was implemented in a Python script previously developed by a colleague, Ms. Maria Cecilia Barrera.

The liquid-phase simulations were carried out in GROMACS (version 5.1.2) and consisted of: 1) an energy minimisation step; 2) an equilibration step in the NVT ensemble (0.1 ns); 3) an equilibration step in the NPT ensemble (0.1 ns); 4) a production simulation in the NPT ensemble (10 ns). There were 200 organosilica molecules in each simulation box. The model density (ρ_{model}) was taken from the GROMACS inbuilt function "GMX energy" and was equal to the average density over all the frames in the production simulation. The density of a given frame is equal to the total mass of the molecules, which is constant in a non-reactive NPT simulation, divided by the volume of the cubic simulation cell, which fluctuated. Additionally, the total energy of the liquid phase was extracted, also via "GMX energy", and was simply the sum of all the potential terms introduced above, i.e. bond-stretching, angle-bending, torsional, electrostatic and LJ potentials, averaged over all frames of the trajectory. The Berendsen thermostat [87] was used (time constant = 0.1 ps) to maintain a constant temperature of 298.15 K. For the NPT simulations, an average pressure of 1.0 bar was maintained by the Parrinello-Rahman barostat [94] (time constant = 2.0 ps). Long-range electrostatic interactions were treated using the Particle-Mesh Ewald technique [64]. All the data in the trajectory were saved every 10 ps. Dispersion corrections and cubic boxes with periodic boundary conditions were also employed.

For the gas-phase simulations, an energy minimisation step and then a 50 ns simulation in the NVT ensemble were carried out. There was one organosilica molecule in each simulation. This is to represent an ideal gas; if there is only one molecule present

in a sufficiently large box then there will be no intermolecular interactions, i.e. it will be as close to an ideal gas as possible. From these simulations, only the energy term was extracted. This was done in the same way as for the liquid-phase simulations. The heat of vaporisation for the model ($\Delta H_{vap,model}$) was then calculated as

$$\Delta H_{vap,model} = E_{liq} - E_{vap} + RT \tag{6.9}$$

where E_{liq} and E_{vap} are the total energies of the liquid- and gas-phase simulations, respectively. The first molecule to be parametrised was tetramethylsilane ($Si(CH_3)_4$). Initially, the CHELP charges (q_{Si} = 0.888, see **Table 6.14**), the modified TraPPE parameters [327,328] for the methyl groups (σ_{CH_3} = 0.379 nm, ε_{CH_3} = 0.833 kJ/mol) and the bonded parameters from the previous sections were employed. The grid consisted of: $0.55 \leq \sigma_{Si} \leq 0.70$ nm with a spacing of 0.005 nm; $0.04 \leq \varepsilon_{Si} \leq 0.30$ kJ/mol with a spacing of 0.02 kJ/mol. The modified TraPPE model for neopentane (the organic analogue of tetramethylsilane) has LJ parameters of σ_C = 0.646 nm and ε_C = 0.042 nm [328], so this was an appropriate region of the parameter space to probe. The optimisation algorithm found an optimum at σ_{Si} = 0.585 nm and ε_{Si} = 0.16 nm. For these values, ρ_{model} = 642.5 ± 0.8 kg/m^3 (cf. ρ_{exp} = 641 kg/m^3 in **Table 6.1**), while $\Delta H_{vap,model}$ = 25.7 ± 0.1 kJ/mol (cf. $\Delta H_{vap,exp}$ = 26.0 ± 0.6 kJ/mol in **Table 6.1**). Here we can see that the model density is, taking into account statistical uncertainty, slightly higher than the experimental density, while the model heat of vaporisation is well within the experimental error of the experimental value. The same procedure was also carried out with the DDEC charges and the optimum values for σ_{Si} and ε_{Si} were 0.59 nm and 0.1675 kJ/mol, respectively. The density for these values of σ_{Si} and ε_{Si} was 641.4 ± 0.5 kg/m^3, while the heat of vaporisation was 26.0 ± 0.1 kJ/mol.

Subsequently, we can move onto our second molecule, namely trimethylsilanol. Trimethylsilanol substitutes one of tetramethylsilane's methyl groups for a hydroxyl group and hence has the chemical formula $Si(CH_3)_3OH$. From **Table 6.14**, the CHELPG charges for this molecule are: q_{Si} = 1.086; q_O = -0.827; q_{H_O} = 0.413. Since the hydroxyl hydrogen has only a point charge located on it and is not a LJ centre, the only unknowns are σ_{O_H} and ε_{O_H}. Nevertheless, this parametrisation turned out to be

more challenging than expected and we will demonstrate this with the aid of **Figure 6.7**. It should also be noted that we applied polarisation corrections for this molecule since it is a polar molecule. We used the same corrections as in **Chapter 4**, but only up to dipoles due to data for the quadrupoles being unavailable. The data needed for the calculation were the gas-phase dipole (2.01 D [380]), the isotropic polarisability (10.4 ^{3} [403]), the refractive index (1.3889 [380]), the static dielectric constant (7.17 [380]) and the liquid-phase density (814.1 kg/m^3 [380]). From the input data, the correction turned out to be +1.18 kJ/mol for the heat of vaporisation and therefore -1.18 kJ/mol for the free energy of self-solvation.

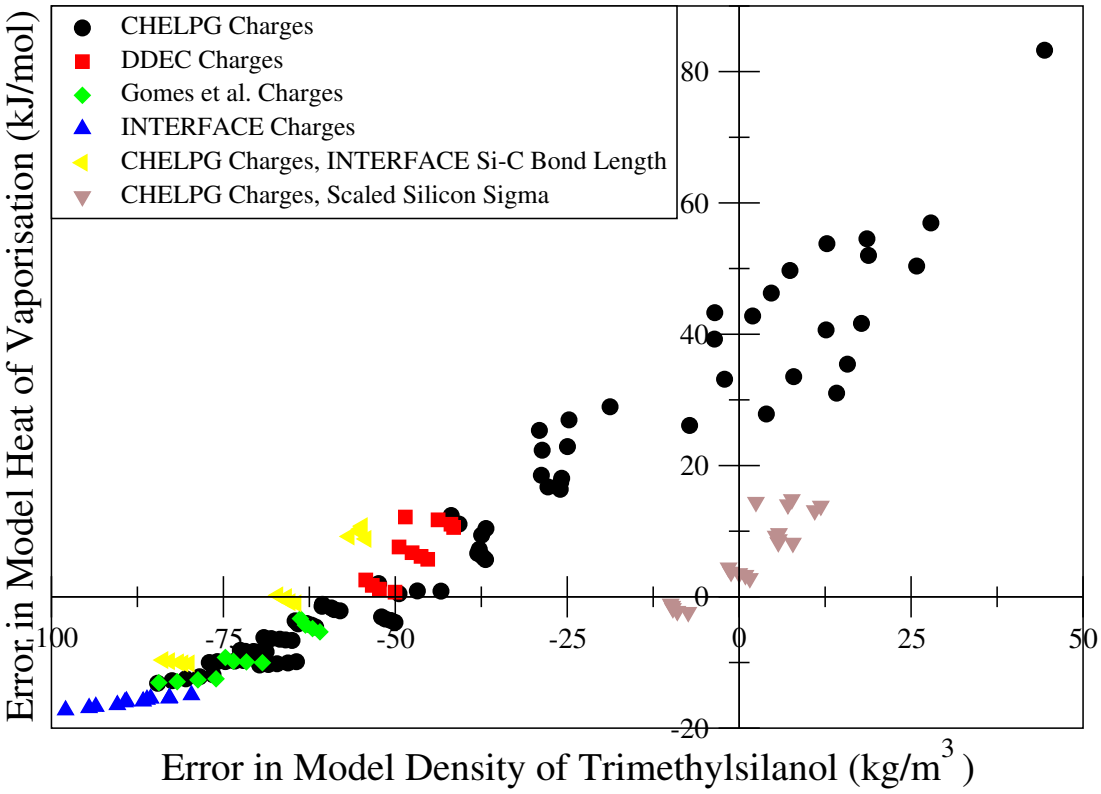


Figure 6.7: Graphical representation of all the different combinations used for the Lennard-Jones parameters of trimethylsilanol. Each point represents the error in the density and heat of vaporisation of the model for a certain combination of Lennard-Jones parameters for the hydroxyl oxygen

Initially, we used only the CHELPG charges for the parametrisation. However, despite using increasingly extreme values of σ_{O_H} and ε_{O_H}, we were unable to find

any combination thereof that could simultaneously reproduce the experimental density and heat of vaporisation, as shown by the black circles in **Figure 6.7**. The values of σ_{O_H} ranged from 0.20 to 0.34 nm, while those of ε_{O_H} went from 0.7 to 8.0 kJ/mol. Considering that the TraPPE parameters are $\sigma_{O_H} = 0.302$ nm and $\varepsilon_{O_H} = 0.773$ kJ/mol [374], the lower limit of $\sigma_{O_H} = 0.20$ nm and the higher limit of $\varepsilon_{O_H} = 8.0$ kJ/mol, in particular, are extreme. The values in this region represent the subset of black circles in the top right-hand corner of **Figure 6.7**. What the black points do show, however, is that no combination of parameters for σ_{O_H} and ε_{O_H} can even come close to reproducing ρ_{exp} and $\Delta H_{vap,exp}$ at the same time, since they all lie broadly along a single curve far from the origin. **Figure 6.7** therefore reveals that either our other potential parameters are inaccurate or that our value for the LJ parameters for the central tetrahedral silicon are not transferable from tetramethylsilane to trimethylsilanol.

The challenge is, therefore, not to find the parameter that will significantly change both density and heat of vaporisation, but rather the parameter upon which one property depends much more strongly than the other. Our most reasonable estimates for the σ_{O_H} and ε_{O_H}, i.e. those similar to the TraPPE parametrisation, produce values for the heat of vaporisation that are not all too far from the experimental value. However, the values for the density deviate significantly, i.e. are approximately 50 kg/m^3 too low. This tells us that the main property of interest here is the density. As such, we are looking for the potential that will have a much stronger effect on the density, a structural property, than on the heat of vaporisation, an energetic one.

To try overcome this problem, other combinations of potentials were employed. The remaining options are the electrostatic and intramolecular potentials. Since we are looking for an increase in the density, we would expect that a higher value of q_{O_H} would be appropriate. This would increase the electrostatic contribution, making this interaction between the molecules more attractive. The only difference between trimethylsilanol and tetramethylsilane is the substitution of a methyl for an alcohol group; this substitution significantly increases the density from 650 to 810 kg/m^3, i.e. by 25%. Since the main difference that this makes to the molecule is to make it dipolar and significantly increase the Coulombic interaction, electrostatics are, most likely, the

origin of the large increase in the density. Therefore, it is worth testing to see whether our charges, and thereby the electrostatic part of the potential, are simply not large enough to produce our desired density. However, as already stated, the root of this problem is to find the parameter that significantly influences the density but not the heat of vaporisation. Bearing this in mind, it is less likely that electrostatics are the origin of our fitting problem. The increase in the charges, and thereby in the level of attraction between the molecules, will also increase the heat of vaporisation, since more energy will need to be added to the system to prise them apart and create a gas. As such, it would seem that the magnitude of our charges is not the fundamental reason why our simulations of trimethylsilanol cannot match the experimental density.

To test this hypothesis, CHELPG (q_{O_H} = -0.827) charges were replaced with DDEC (q_{O_H} = -0.908) charges, as well as with the charges from the works of Gomes et al. [243] (q_{O_H} = -0.799) and the INTERFACE force field [379] (q_{O_H} = -0.660). The trimethylsilanol was then reparametrised with each of these new sets of charges for the same range of LJ parameters (σ_{O_H} = [0.28,0.29,0.30], ε_{O_H} = [0.7,0.8,0.9,1.0]). They are represented by the red, green and blue data points, respectively. As expected, these new parametrisations still fall along the same curve as the black points, i.e. the electrostatics do not fundamentally change the shape of the curve, but rather only move the data points along the curve. Therefore, the electrostatics are not the part of our system that is stopping us from matching both the experimental density and heat of vaporisation simultaneously.

Since we have already ruled out the influence of the LJ and coulombic parameters in being able to shift us away from the general trend of the curve in **Figure 6.7**, this only leaves us with the intramolecular parameters. The most basic way to test the effect of the intramolecular parameters is to alter the bond length. Since it appears three times in the molecule, the $Si-C$ bond is the most obvious candidate for this test. We would expect a decrease in the length of this bond to enact an increase on the density due to the overall size of the molecule effectively being shrunk. This may pull the repulsive part of the LJ potential closer to the surface of the molecule, thereby allowing molecules to approach more closely to one another. However, it may also weaken the attractive

part of the LJ potential; since these groups will now be closer to the central Si in the molecule, they will also be further from other molecules. Since the attractive part of the LJ potential also decays fairly rapidly, this could also result in a lowering of the intermolecular attraction.

To test this, we implemented the $Si - C$ bond length (0.1809 nm) that appears within the INTERFACE force field but with the same force constant as used in the other iterations. It should be stressed that this was more to test the influence of the intramolecular parameters rather than because this bond length is physically sensible; the INTERFACE force field was developed principally for solids. This is the reason why the value is about 3.5% smaller than the experimental value of 0.1875 nm that we otherwise use here. The CHELPG charges were used in this parametrisation. Since we are using a different bond length, it was also decided to reparametrise the LJ parameters for the tetrahedral silicon atom from tetramethylsilane. This resulted in new LJ parameters of $\sigma_{Si} = 0.605$ nm and $\varepsilon_{Si} = 0.14$ kJ/mol. This resulted in the yellow points in the **Figure 6.7**. We can see that this fairly drastic change in one of the intramolecular parameters produces only a very minor deviation away from the general trend of the curve. Indeed, rather than the decreased bond length increasing the density of the fluid, as was originally expected, there is an increase; this is probably due to the concomitant increase in σ_{Si}.

Finally, having eliminated most of the other potential functions in the model, we look at the transferability of the LJ parameters for the central silicon atom. Interestingly, there is a parallel between our work and the work of Chen et al. [374], who developed parameters for neopentane and tert-butyl alcohol, the alkane analogues of our molecules in question. They also found that it was not possible to represent both these molecules with the same LJ parameters for the central carbon atom and therefore scaled their value of σ. More precisely, the σ for the tetrahedral carbon in neopentane was 0.64 nm, while this was reduced to 0.58 nm for tert-butyl alcohol. The effect of this is to move the repulsive region closer to the surface of the molecule. This will also mean that the minimum, i.e. the most attractive part of the potential, will also move closer to the surface of the molecule, allowing them to pack more closely. Additionally, and

importantly, this will not necessarily change the heat of vaporisation. Drawing the LJ potential closer to the surface of the molecule will mean that the same intermolecular potential can be achieved, but with a closer packing. As such, this would be seem to be a promising avenue to explore.

If we apply the same ratio as Chen et al., namely 0.906 [374], then, from our value of 0.585 nm for tetramethylsilane, we end up with a value of 0.530 nm for trimethylsilanol. Reparametrising with this value of σ_{Si}, but remaining with our previous value of ε_{Si} (0.16 kJ/mol), we produce the brown points in **Figure 6.7**. From these results, it is immediately clear that this approach is transferable from alkanes to organosilicates, since we almost instantly end up with a curve that goes through the origin. As such, we can see that our method for the parametrisation of our model has been successful in maintaining the transferable nature of the TraPPE model inasmuch as this simple scaling rule produces almost identical results. Additionally, this result provides an indication that many of the trends seen in the parametrisation of organic molecules may also hold true for their organosilica analogues.

Subsequently, using the scaled σ_{Si} value, the LJ parameters of the hydroxyl oxygen in trimethylsilanol were optimised for both the charges from the CHELPG and DDEC fitting schemes. These parametrisations, as well as that for the central silicon atom in tetramethylsilane, are summarised in **Table 6.15**.

Table 6.15: Summary of the parametrisation of the models for tetramethylsilane (TMS) and trimethylsilanol (TMSOL) for both CHELPG and DDEC charges. For a full list of the organosilica charges, please see **Table 6.14**

Molecule	Charges	Optimum			
		σ_{Si} (nm)	ε_{Si} (kJ/mol)	σ_{O_H} (nm)	ε_{O_H} (kJ/mol)
TMS	CHELPG	0.58	0.16	-	-
	DDEC	0.59	0.1675	-	-
TMSOL	CHELPG	0.53	0.16	0.305	1.05
	DDEC	0.535	0.1675	0.318	1.00
Molecule	Charges	ρ (kg/m^3)		ΔH_{Vap} (kJ/mol)	
		Model	Exp.	Model	Exp.
TMS	CHELPG	642.5 ± 0.8	641	25.7 ± 0.1	26.0
	DDEC	641.4 ± 0.5		26.0 ± 0.1	
TMSOL	CHELPG	814.3 ± 0.4	814.1	46.2 ± 0.1	45.6
	DDEC	811.7 ± 0.4		46.0 ± 0.1	

Having now parametrised our models for tetramethylsilane and trimethylsilanol, the last step of our procedure for these molecules is to validate our models against experimental data. As mentioned earlier, our chosen property for validation is the free energy of self-solvation. This was calculated using the Bennett Acceptance Ratio (BAR) method [83] via the "gmx bar" module in GROMACS. The values of λ for the decoupling were [0.0, 0.3, 0.6, 0.7, 0.8, 0.9, 1.0] and [0.00, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 0.55, 0.60, 0.70, 0.80, 0.90, 1.00] for the electrostatic and LJ potentials, respectively. Each simulation was run for 5.0 ns and the BAR analysis began after 0.5 ns. All other details were the same as those used in the simulations for calculating the density and heat of vaporisation. The results for the free energy of self-solvation for the models of tetramethylsilane and trimethylsilanol can be found in **Table 6.16**.

Table 6.16: Summary of the free energy results for tetramethylsilane (TMS) and trimethylsilanol (TMSOL) for the models parametrised in this work for both CHELPG and DDEC charges. For a full list of the charges, please see **Table 6.14**. *LJ* refers to the Lennard-Jones interactions, *ES* to electrostatic, i.e. Coulombic interactions. *Corr* and *Exp* refer to the polarisation correction and the experimental value, respectively

Molecule	Charges	ΔG_{LJ}	ΔG_{ES}	ΔG_{corr}	ΔG_{total}	ΔG_{exp}
TMS	CHELPG	-15.52 ± 0.07	-0.190 ± 0.005	-	-15.71 ± 0.08	-13.0
	DDEC	-15.8 ± 0.1	-0.28 ± 0.01	-	-16.1 ± 0.1	
TMSOL	CHELPG	-13.7 ± 0.4	-9.7 ± 0.1	-1.138	-24.5 ± 0.5	-22.2
	DDEC	-12.7 ± 0.2	-10.2 ± 0.1		-24.0 ± 0.3	

From **Table 6.16**, we can draw some important conclusions about our models. First of all, considering the results for tetramethylsilane, we can see that the LJ interactions dominate. This is expected because the molecule is bulky as well as being non-polar due to symmetry and comparatively weakly charged. From our results, we can see that both our models nevertheless overestimate the LJ interactions. This is apparent because the electrostatic interactions are negligible, yet both models, i.e. those parametrised with CHELPG and DDEC charges, still overestimate the self-solvation free energy by -2.7 and -3.1 kJ/mol, respectively. While this prediction is far from perfect, and indicates that some refinement of our models is required, it is nevertheless an encouraging result. The results for trimethylsilanol are moderately better, the errors for the CHELPG and

DDEC versions of the model being -2.3 and -1.8 kJ/mol, respectively. This means that we still cannot definitively state which charge-fitting procedure is better; the CHELPG charges perform better for tetramethylsilane, while the model fitted with DDEC charges has a lower error for the free energy of self-solvation for trimethylsilanol. However, since electrostatics only have a minor influence in the model of tetramethylsilane, it would be logical to assign more weight to the performance of the model of trimethylsilanol when determining which charge-fitting procedure should be preferred. In that case, our recommendation would be to use the DDEC charges. Nevertheless, the difference is minimal; indeed, the uncertainties of the two simulation values for the self-solvation free energy of trimethylsilanol overlap. Therefore, we can state that both sets of charges produce acceptable results. It would be interesting in further work to test whether the performance of both charge-fitting methods is still comparable for even bulkier molecules, for which DDEC is specialised.

Having validated our parameters, we now have all the necessary components to propose a model for silicic acid by transferability. The parameters for this proposed model are given in **Table 6.17**.

Table 6.17: Enumeration of the parameters for silicic acid from those developed in our model. Where two options are listed, the value on the left-hand side is that parametrised using CHELPG charges, whereas those from DDEC charges are on the right-hand side

Atom Type	σ (nm)	ε (kJ/mol)	q (e)	
Si	0.530 / 0.535	0.16 / 0.1675	+1.448 / +2.10	
O	0.302 / 0.318	0.90 / 1.00	-0.7933/ -1.01	
H	0.00	0.00	+0.4313 / +0.485	
Bond	r_0 (nm)	k_B (kJ/mol.nm^2)		
$Si-O$	0.1653	381,300		
$O-H$	0.0945	462,800		
Angle	θ_0 (nm)	k_θ (kJ/mol.rad^2)		
$Si-O-H$	115.5	257.8		
$O-Si-O$	104.4	872.6		
Dihedral	$k_{T,0}$ (kJ/mol)	$k_{T,1}$ (kJ/mol)	$k_{T,2}$ (kJ/mol)	$k_{T,3}$ (kJ/mol)
$H-O-Si-O$	10.92	8.41	1.36	4.35

6.8 Conclusions

In this chapter, a framework for the parametrisation of a force field has been proposed, developed and validated for the organosilica precursors of mesoporous silica materials. The first stage of this was to identify which classes of organosilicas were most important for studying the self-assembly of mesoporous silica and for which reliable experimental data were available. The chosen molecular classes were silanes, silanols, siloxanes and alkoxysilanes. From our survey of the literature, data for a few molecules belonging to each of these families were collated. These data were mainly comprised of values for the liquid-phase density, heat of vaporisation and vapour pressure. Additionally, for each of these molecules it was possible to locate data pertaining to the geometry of these molecules, namely for bond stretching and angle bending.

During this part of the work, it was noted that there is a relative dearth of experimental data available for organosilica molecules. Despite the intense interest in solid silica materials and their properties, it would seem that this interest has not extended to their (potential) precursors. A suggestion, from an experimental point of view, would be to undertake a systematic study of liquid-phase properties of such molecules that would be relevant for parametrising molecular models of them. Our suggestions for these properties are the liquid-phase density, vapour pressure, heat of vaporisation, free energy of self-solvation, diffusion coefficients, dielectric constants and the magnitudes of the liquid-phase dipoles. Were these properties to be available for a large training set of organosilica molecules, they would form a solid basis for a wide-ranging and transferable force field suitable for the study of the synthesis of mesoporous silica.

Following on from the literature survey, QM calculations were carried out in order to ascertain parameters for the dihedrals and point charges for the molecules included in our test set. The dihedrals proved to be the most difficult part of the force field to fit. The fitted parameters were shown to not be able to describe the full 360-degree rotational energy with a high level of accuracy. However, the method in this work was shown to be fairly accurate at the most important part of the torsional curve, namely around the minimum. Additionally, the fit was shown to be a significant improvement

on other values of the dihedral parameters in the literature. Nevertheless, there is still room for improvement in this part of the parametrisation.

The point charges were fitted to data from QM calculations using a continuum solvation model. In this way, it was possible to achieve charges that, crucially, were more representative of the liquid phase, where the synthesis of mesoporous silicas takes place, than simply using vacuum calculations. As in **Chapter 5**, two different charge-fitting procedures were employed. The DDEC method produced charges that were higher than CHELPG. Additionally, they were more consistent, with similar atoms having approximately the same charge rather than experiencing strong deviations. However, the DDEC charges are very large, higher even than some equivalent atoms in force fields of solids. Nevertheless, both parametrisations employing these charge-fitting procedures produced reasonable results. The DDEC charges were shown to be slightly better and would likely perform even better in bulkier molecules, for which DDEC was developed.

Finally, we used these insights to help us fit LJ parameters for two molecules (tetramethylsilane and trimethylsilanol) related to silicic acid. We chose these molecules because, unlike silicic acid, they are stable in the liquid-phase and our goal was therefore to parameterise a model for silicic acid via the transferability of our parameters. It was possible to fit the LJ parameters so as to match the density and heat of vaporisation of these molecules accurately. Subsequent validation showed that the fitted models were able to match the free energy of self-solvation to within 10—20%, demonstrating the efficacy of our modelling approach. Another important result in this part of the work was that we were able to establish the transferability of trends seen for models of organic molecules to their organosilica analogues. This could be invaluable for solving problems associated with future parametrisations of this molecular class.

Although we have therefore completed one of the principal objectives of this work, namely to parametrise a model of silicic acid, our new model still requires more work to validate its efficacy. As such, our future work will consist of the following steps:

- Validate the transferability of our models of tetramethylsilane and trimethylsilanol by testing how well they can predict the properties of their ethyl analogues

- Testing our model of silicic acid in aqueous solution by calculating its hydration free energy
- Fit the LJ parameters for the siloxane- and alkoxide-oxygen atom types

To conclude, we believe that the work contained within this chapter demonstrates the validity of our approach and that extending our model to a greater variety of organosilica molecules will result in a useful tool for simulations of the synthesis of mesoporous silica materials.

Chapter 7

Conclusions

Having surveyed the literature of computational studies of mesoporous silica synthesis in **Chapter 3**, we identified two areas that we believed constituted a hindrance to the advancement of this field. The first of these was the fact that, despite the outcome of the synthesis being strongly dependent on electronic phenomena, polarisation effects were effectively ignored. The second conclusion that we drew was that the models reviewed in our discussion of mesoporous silica synthesis simulations tended to fall into two categories: either they were extremely generic, and thereby lacking in any precise correspondence to any real-world synthesis; or very specialised, making them suitable only for a very limited range of applications. This lack of transferability means that it is a relatively arduous task to introduce new organosilica molecules into simulations, inhibiting these models' ability to easily study the *in situ* functionalisation of mesoporous silica. As a consequence of these assessments, it was decided that our objectives were twofold:

- To investigate and thereby incorporate liquid-phase polarisation effects into force fields of silica synthesis as cheaply as possible
- With the insights gathered from carrying out our initial objective, to parametrise a transferable force field for organosilica molecules

Our investigation of polarisation in the liquid phase was further subdivided into two areas of interest. The first touched upon polarisation corrections for the energy,

particularly the free energy, upon changes of phase; this was deemed important because the properties of the material are strongly influenced by phase equilibria. This topic was extensively dealt with in **Chapter 4**. Due to its role as by far the most common solvent in mesoporous silica synthesis, as well as the abundance of theoretical data available for it, our study focussed primarily on water. The second objective was to develop a procedure for deriving point charges for molecular simulation models that had the correct degree of polarisation. This was dealt with in **Chapter 5**.

In **Chapter 4**, we demonstrated that the currently accepted "gold standard" for polarisation corrections, the Berendsen correction, is inaccurate for describing polarisation effects during phase changes. Furthermore, it was shown that it was not only insufficient, but even detrimental to a fixed-charge water model being able to simultaneously match the heat of vaporisation and free energy of self-solvation. This initial study posited that the overall polarisation correction should be negative in order to cancel out this discrepancy. Building on the work of Leontyev and Stuchebrukhov, who showed that the relative success of the Berendsen correction was based upon cancellation of error, rather than good physics, we showed that, by extending the polarisation correction from only dipoles up to quadrupoles, the posited overall negative correction was obtained. This demonstrated that the inclusion of higher-order multipoles is necessary for the correct mathematical treatment of polarisation. Subsequently, it was clearly demonstrated that the use of average multipole moments, rather than the whole multipole distribution, has a considerable effect on the overall value of the correction. Finally, having integrated these new phenomena into the polarisation corrections, the resulting value was shown to be very close to the corrective needed to unify the predictions of the heat of vaporisation and free energy of self-solvation in a single fixed-charge model. We believe that this is a convincing indicator of the validity and rigour of our approach.

Despite the success of our approach, it is not without its limitations. The first of these is that a large amount of data need to be known *a priori* in order to calculate these corrections. While this fact was not overly burdensome for water, due to its ubiquity, this could be problematic for applying our calculations more generally to force-field development. One way to circumvent this in future work could be to calculate these

corrections for families of common molecules and search for trends that could be used to generalise these effects. This has the potential to have a profound effect on the parametrisation of new force fields in two ways: 1) it would allow us to identify for which molecules it is particularly necessary to include these corrections; 2) it would allow these corrections to be estimated in a simple manner, thereby allowing them to be included *en masse* for parametrisation of future force fields. An obvious candidate family for a future study in this direction would be alcohols.

Secondly, while our study has raised the level of sophistication of these corrections, they are nevertheless still truncated forms of a longer series. While we have shown the importance of including quadrupoles in our calculations, it would be necessary to extend this even further, were we to definitively demonstrate that this is far enough to capture the bulk of the phenomenon. For future work, a rigorous test of this would be to calculate the individual corrections at the quantum-mechanical level of theory. Since such a calculation would include every set of multipoles, if the corrections from this method could be shown to be approximately the same as those from the quantum calculation, then this would clearly demonstrate that the influence of multipoles above quadrupoles is negligible for water.

Finally, our chosen continuum solvation model, the Born model, can hardly be considered "cutting-edge", and was instead chosen for mathematical simplicity. Since we have clearly demonstrated that the distortion and electronic corrections are the crucial properties of interest for calculating the missing polarisation contribution to the free energy of solvation, an attractive future avenue would be to move from calculating these properties by classical approximations to computing them via quantum-mechanical calculations, i.e. the highest available level of theory. This would substantially reduce the barriers to the implementation of these corrections because it would obviate the need for the large volume of input data required to calculate them via classical approximations.

In **Chapter 5**, we first proposed a procedure for calculating liquid-phase point charges. This procedure consisted of using molecular dynamics simulations to optimise the liquid structure of a collection of molecules and then using coordinates from frames corresponding to the average density as input for a single-point DFT calcu-

lation. This process was then iterated in the hope that the resulting charges would correspond to those of the liquid state. The results, however, demonstrated that the charges obtained were poor and were much lower than would be expected for the real liquid phase. This was shown to be true for both DDEC and REPEAT charges, with both producing charges for methanol that were barely more polarised than their vacuum equivalents. There are a few possible causes for the poor results. Firstly, a small basis set (double-zeta without polarisation corrections) had to be used in order to facilitate DFT calculations of a few thousand atoms. Secondly, standard DFT calculations do not include hydrogen bonding and dispersion interactions. The lack of polarisation corrections in the basis set and hydrogen bonding are probably the most important factors here for water and methanol. The principal dilemma here is that a large quantity of molecules are needed to represent the liquid phase in the quantum calculation, however this prohibits the optimisation of all these molecules with a suitably high level of theory. One avenue that should be explored in future, in order to escape this bind, is electrostatic embedding. This would replace all but one of the molecules with charges and then optimise only the remaining one. If this were to be carried out for a large number of snapshots, this could preclude the need for a large number of molecules.

Subsequently, a continuum solvation model approach was applied. In this part of the project, two continuum solvation models (IEFPCM and SMD) and two charge-fitting procedures (RESP and DDEC) were tested. Methanol was once more selected as the test system. By using a sensitivity analysis, it was shown that charges calculated using the IEFPCM continuum solvation approach were superior to those obtained via the SMD method. Indeed, the sensitivity analysis demonstrated that the charges from either RESP or DDEC in combination with IEFPCM both lie near the degree of polarisation (modelled as a scaling factor) that is optimal for a united-atom methanol model. As such, we clearly demonstrated, at least for methanol, that IEFPCM is the better of the two continuum solvation models for determining liquid-phase charges. The obvious deficiency here is that we only demonstrated this for methanol. While methanol was a reasonable candidate, since we are interested in organosilica molecules, it would nevertheless be instructive to test whether IEFPCM is generally superior for obtain-

ing liquid-phase charges for a larger test set of molecules. If it could be demonstrated that IEFPCM, or the MD/QM-with-electrostatic-embedding approach described above, consistently produces charges that are the optimal set of point charges, i.e. correctly take into account polarisation and electronic screening, then this would be a powerful tool for the parametrisation of force fields. It is often necessary to parametrise the point charges as a free parameter, which greatly increases the amount of data required for fitting. Obviating this would therefore facilitate the development of more accurate liquid-phase force fields. This would be especially useful when developing models for chemical families where the quantity of available experimental data points for fitting is low, such as in our case with organosilica.

In **Chapter 6**, we addressed our second objective, namely the parametrisation of a generic, transferable model for organosilica molecules. The first step of this process was to identify the data available in the literature against which we could parametrise our model. From this initial stage, it was clear that there is a relative dearth of experimental data for organosilica. One recommendation we would therefore make for the continuation of this work would be to experimentally measure more properties that would aid the further parametrisation of this class of molecules. The properties that we have identified as potentially most useful for this project, due to their importance in the synthesis of mesoporous silica materials, are the liquid-phase density, heat of vaporisation, free energy of (self-)solvation, vapour pressure, diffusion constant, dielectric constant and liquid-phase dipole.

We started the parametrisation by tracking down and evaluating the accuracy of bond-stretching and angle-bending parameters available in the literature. The majority of these were from a study by Grigoras and Lane in 1988. Due to the age of the calculations, the level of theory used was not especially high. However, we demonstrated that they were nevertheless of good quality by cross-referencing them against more recent quantum calculations employing a higher level of theory and experimental data. Next, we calculated torsional parameters for our collection of test molecules and attempted to demonstrate their transferability. This was met with mixed results; although our parameters were shown to capture the behaviour of two dihedrals appropriately and to

be an improvement over previous attempts, they demonstrated a non-negligible degree of error. While this probably did not have a large impact on the parametrisation of the molecules in this work, due to their relatively small size, this could nevertheless be important when this approach is extended to larger precursor molecules, for which dihedral torsion may be significant. As such, this part of the parametrisation should be reviewed and improved.

Finally, we parametrised Lennard-Jones parameters for the first two molecules in our training set, namely tetramethylsilane and trimethylsilanol. One important conclusion that stemmed from this part of the work was that parametrisation trends seen in the organic analogues of organosilicas are likely to be transferable. This was clearly demonstrated from the fact that only by using an identical scaling factor for the σ-value of the tetrahedral silicon to that of the central carbon atom in tertiary alcohols was it possible to parametrise an accurate model for both the density and heat of vaporisation of trimethylsilanol. This insight will undoubtedly be useful for overcoming similar problems with the parametrisation of other models in this chemical series. Validation of these parameters was carried out by comparing their free energies of self-solvation to the experimental values. This demonstrated the efficacy of our parametrisation strategy; we were able to match these values to within 2 kJ/mol. Considering the lack of data available for the parametrisation, this is an excellent result. Nevertheless, this still leaves room for improvement. Since the ultimate goal of this work would be to use this model for studying the synthesis of porous silica materials and since the self-assembly process relies on such a delicate balance of chemical and physical phenomena, it will likely be necessary to match the free energy of self-solvation more accurately. One way to do this could be to introduce another parameter, namely a scaling parameter for the charges, and use this to fit the free energy of self-solvation as well. Although we demonstrated that the point charges that we obtained from our fitting procedure were near ideal for methanol, we do not know for certain that this also holds true for organosilicas. Since the synthesis routes that we wish to explore are heavily determined by electrostatics, it would definitely be worth exploring the efficacy of our charge-fitting strategy for organosilicas, in particular. The implementation of a scaling parameter for the charges,

as was implemented in **Chapter 5** for methanol, in the parametrisation would be an effective way to achieve this. However, this would leave us without any data for the validation procedure for the majority of our target molecules; this highlights once more the importance of obtaining a larger and more varied set of experimental data points in future work.

The ultimate test of our transferable force field will, however, be its ability to correctly predict the properties of organosilica molecules for which it has not been directly parametrised, something which we have not yet attempted. As a consequence of parametrising the two aforementioned molecules, we have obtained all the necessary parameters to propose a model for silicic acid, the key precursor for mesoporous silica materials, via transferability, therefore completing one of the main objectives of this work. Therefore, an appropriate next step would be to replicate studies of silicic acid in solution and compare them with our new model. This would provide a stringent benchmarking of the transferability of our model due to the complexity of silicic acid. Were this to be successful, it would represent significant progress towards opening up simulations of the synthesis of mesoporous silica materials to a more generalist audience, thereby facilitating the discovery of further novel materials in this class.

To conclude, in the first part of this work, significant advances were made in our understanding and the implementation of polarisation corrections in molecular simulations. It was demonstrated that the *de facto* method for calculating polarisation corrections, the Berendsen correction, was flawed due to its neglect of a second correction of opposite sign and of higher-order multipoles. The inclusion of these properties was shown to explain and fix a systematic error in the parametrisation of fixed-charge water models, opening the door to more accurate versions of such models. Subsequently, a fitting strategy was developed for obtaining point charges that were shown to be optimal for representing polarisation in liquid-phase molecular simulations. Finally, a parametrisation strategy employing the aforementioned insights was developed and validated for a force field for the precursors of functionalised mesoporous silica materials. Further development of this model has the potential to facilitate a wider range of computational studies into mesoporous silica, thereby aiding the discovery of new materials for a broad

range of technological applications.

Bibliography

[1] M. Kah, N. Tufenkji, and J. C. White, "Nano-enabled strategies to enhance crop nutrition and protection," *Nature Nanotechnology*, vol. 14, pp. 532–540, jun 2019.

[2] J. P. Giraldo, H. Wu, G. M. Newkirk, and S. Kruss, "Nanobiotechnology approaches for engineering smart plant sensors," *Nature Nanotechnology*, vol. 14, pp. 541–553, jun 2019.

[3] N. Yousefi, X. Lu, M. Elimelech, and N. Tufenkji, "Environmental performance of graphene-based 3D macrostructures," *Nature Nanotechnology*, vol. 14, pp. 107–119, feb 2019.

[4] M. Gibertini, M. Koperski, A. F. Morpurgo, and K. S. Novoselov, "Magnetic 2D materials and heterostructures," *Nature Nanotechnology*, vol. 14, pp. 408–419, may 2019.

[5] J. S. Beck, J. C. Vartuli, W. J. Roth, M. E. Leonowicz, C. T. Kresge, K. D. Schmitt, C. T.-W. W. Chu, D. H. Olson, E. W. Sheppard, S. B. McCullen, J. B. Higgins, J. L. Schlenker, and J. L. Schlenkert, "A new family of mesoporous molecular sieves prepared with liquid crystal templates," *J. Am. Chem. Soc.*, vol. 114, pp. 10834–10843, dec 1992.

[6] C. T. Kresge, M. E. Leonowicz, W. J. Roth, J. C. Vartuli, and J. S. Beck, "Ordered mesoporous molecular sieves synthesized by a liquid-crystal template mechanism," *Nature*, vol. 359, pp. 710–712, 1992.

[7] M. Jorge, A. Milne, O. Sobek, A. Centi, G. Pérez-Sánchez, and J. Gomes, "Modelling the self-assembly of silica-based mesoporous materials," *Molecular Simulation*, vol. 44, no. 6, 2018.

[8] S. M. Auerbach, W. Fan, and P. A. Monson, "Modelling the assembly of nanoporous silica materials," *International Reviews in Physical Chemistry*, vol. 34, pp. 35–70, jan 2015.

[9] T. Asefa, M. J. MacLachlan, N. Coombs, and E. al., "Periodic mesoporous organosilicas with organic groups inside the channel walls," *Nature*, no. 402, pp. 867–871, 1999.

[10] B. J. Melde, B. T. Holland, C. F. Blanford, and A. Stein, "Mesoporous Sieves with Unified Hybrid Inorganic/ Organic Frameworks," *Chem. Mater.*, no. 11, pp. 3302–3308, 1999.

[11] S. Inagaki, S. Guan, Y. Fukushima, T. Ohsuna, and O. Terasaki, "Novel Mesoporous Materials with a Uniform Distribution of Organic Groups and Inorganic Oxide in Their Frameworks," *J. Am. Chem. Soc.*, no. 121, pp. 9611–9614, 1999.

[12] F. Hoffmann, M. Cornelius, J. Morell, and M. Fröba, "Silica-Based Mesoporous OrganicâĂŞInorganic Hybrid Materials," *Angewandte Chemie International Edition*, vol. 45, pp. 3216–3251, may 2006.

[13] A. Patti, A. D. Mackie, V. Zelenak, and F. R. Siperstein, "One-pot synthesis of amino functionalized mesoporous silica materials: using simulations to understand transitions between different structures," *Journal of Materials Chemistry*, vol. 19, no. 6, p. 724, 2009.

[14] A. Patti, A. D. Mackie, and F. R. Siperstein, "Monte Carlo simulations of self-assembling hexagonal and cage-like bifunctional periodic mesoporous materials," *Journal of Materials Chemistry*, vol. 19, no. 42, p. 7848, 2009.

[15] L. Jin, S. M. Auerbach, and P. A. Monson, "Simulating the formation of surfactant-templated mesoporous silica materials: A model with both surfactant self-assembly and silica polymerization," *Langmuir*, vol. 29, pp. 766–780, jan 2013.

[16] G. L. Hornyak, H. F. Tibbals, J. Dutta, and J. J. Moore, *Introduction to Nanoscience and Nanotechnology.* CRC Press, 2008.

[17] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of State Calculations by Fast Computing Machines," *Studies in Molecular Dynamics. I. General Method The Journal of Chemical Physics*, vol. 21, no. 6, p. 459, 1953.

[18] B. J. Alder and T. E. Wainwright, "Phase Transition for a Hard Sphere System," *Citation: J. Chem. Phys*, vol. 27, p. 1208, 1957.

[19] S. H. Fuller and L. I. Millett, "Computing Performance: Game Over or Next Level?," *Computer*, vol. 44, pp. 31–38, jan 2011.

[20] M. Govoni and G. Galli, "Large Scale GW Calculations," *J. Chem. Theory Comput.*, vol. 11, no. 6, pp. 2680–2696, 2015.

[21] S. I. Allec and B. M. Wong, "Inconsistencies in the Electronic Properties of Phosphorene Nanotubes: New Insights from Large-Scale DFT Calculations," *J. Phys. Chem. Lett*, vol. 7, pp. 4340–4345, 2016.

[22] A. Erba, J. Baima, I. Bush, R. Orlando, and R. Dovesi, "Large-Scale Condensed Matter DFT Simulations: Performance and Capabilities of the CRYSTAL Code," *Journal of Chemical Theory and Computation*, vol. 13, pp. 5019–5027, oct 2017.

[23] D. R. Bowler, R. Choudhury, M. J. Gillan, and T. Miyazaki, "Recent progress with large-scale ab initio calculations: the CONQUEST code," *phys. stat. sol. (b)*, vol. 243, no. 5, pp. 989–1000, 2006.

[24] B. Engquist and E. Weinan, "Multiscale Modeling and Computation," *Notices of the AMS*, vol. 50, no. 9, pp. 1062–1070, 2003.

[25] G. S. Ayton, W. G. Noid, and G. A. Voth, "Multiscale modeling of biomolecular systems: in serial and in parallel," *Current Opinion in Structural Biology*, vol. 17, pp. 192–198, apr 2007.

[26] Q. Zeng, A. Yu, and G. Lu, "Multiscale modeling and simulation of polymer nanocomposites," *Progress in Polymer Science*, vol. 33, pp. 191–269, feb 2008.

[27] F. Siperstein, A. L. Myers, and O. Talu, "Long range corrections for computer simulations of adsorption," *MOLECULAR PHYSICS*, vol. 100, no. 13, pp. 2025–2030, 2002.

[28] S. Piana, K. Lindorff-Larsen, R. M. Dirks, J. K. Salmon, R. O. Dror, and D. E. Shaw, "Evaluating the Effects of Cutoffs and Treatment of Long-range Electrostatics in Protein Folding Simulations," *PLoS ONE*, vol. 7, p. e39918, jun 2012.

[29] J. Gelpi, A. Hospital, R. Goñi, and M. Orozco, "Molecular dynamics simulations: advances and applications," *Advances and Applications in Bioinformatics and Chemistry*, vol. 8, p. 37, nov 2015.

[30] I. V. Leontyev and A. Stuchebrukhov, "Electronic Polarizability and the Effective Pair Potentials," *Journal of Chemical Theory and Computation*, vol. 6, no. 6, pp. 3153–3161, 2010.

[31] H. J. Berendsen, J. R. Grigera, and T. P. Straatsma, "The missing term in effective pair potentials," *J. Phys. Chem.*, vol. 91, pp. 6269–6271, 1987.

[32] A. Milne and M. Jorge, "Polarization Corrections and the Hydration Free Energy of Water," *Journal of Chemical Theory and Computation*, vol. 15, no. 2, 2019.

[33] H. Yu and W. F. van Gunsteren, "Accounting for polarization in molecular simulation," *Computer Physics Communications*, vol. 172, pp. 69–85, nov 2005.

[34] C. M. Baker and R. B. Best, "Matching of Additive and Polarizable Force Fields for Multiscale Condensed Phase Simulations," *Journal of Chemical Theory and Computation*, vol. 9, pp. 2826–2837, 2013.

[35] M. Jorge, J. R. B. Gomes, M. N. D. S. Cordeiro, and N. a. Seaton, "Molecular dynamics simulation of the early stages of the synthesis of periodic mesoporous silica.," *J. Phys. Chem. B*, vol. 113, pp. 708–718, jan 2009.

[36] G. Pérez-Sánchez, J. R. B. Gomes, and M. Jorge, "Modeling self-assembly of silica/surfactant mesostructures in the templated synthesis of nanoporous solids," *Langmuir*, vol. 29, pp. 2387–2396, feb 2013.

[37] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*. Elsevier, 2nd2002 ed., 2002.

[38] D. C. Rapaport, *The Art of Molecular Dynamics Simulation*. Cambridge University Press, 2nd ed., 2004.

[39] B. H. M.J. Abraham, D. van der Spoel, E. Lindahl and The GROMACS development team, "GROMACS User Manual version 2016," 2018.

[40] U. D. Schiller, "An Overview of Integration Schemes for Molecular Dynamics Simulations," tech. rep., Johannes Gutenberg University Mainz, 2008.

[41] J. Liouville, "Liouville's Theorem," *Journal de mathématiques pures et appliquées*, vol. 3, p. 349, 1836.

[42] L. Verlet, "Computer Experiments on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules," *Physical Review*, vol. 159, pp. 98–103, jul 1967.

[43] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson, "A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters," *The Journal of Chemical Physics*, vol. 76, pp. 637–649, jan 1982.

[44] H. J. Berendsen, D. van der Spoel, and R. van Drunen, "GROMACS: A message-passing parallel molecular dynamics implementation," *Computer Physics Communications*, vol. 91, no. 1-3, pp. 43–56, 1995.

[45] A. Stone, *The Theory of Intermolecular Forces.* Oxford University Press, jan 2013.

[46] S. A. Clough, Y. Beers, G. P. Klein, and L. S. Rothman, "Dipole moment of water from Stark measurements of H2O, HDO, and D2O," *J. Chem. Phys.*, vol. 59, pp. 2254–2259, sep 1973.

[47] Y. S. Badyal, M.-L. Saboungi, D. L. Price, S. D. Shastri, D. R. Haeffner, and A. K. Soper, "Electron distribution in water," *J. Chem. Phys.*, vol. 112, p. 9206, may 2000.

[48] P. Schienbein and D. Marx, "LiquidâĂŞVapor Phase Diagram of RPBE-D3 Water: Electronic Properties along the Coexistence Curve and in the Supercritical Phase," *The Journal of Physical Chemistry B*, vol. 122, pp. 3318–3329, apr 2018.

[49] F. London, "The general theory of molecular forces," *Transactions of the Faraday Society*, vol. 33, p. 8b, jan 1937.

[50] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, "Development and Testing of the OLPS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids," *J. Am. Chem. Soc.*, vol. 118, no. 15, pp. 11225–11236, 1996.

[51] H. A. Lorentz, "Ueber die Anwendung des Satzes vom Virial in der kinetischen Theorie der Gase," *Annalen der Physik*, vol. 248, pp. 127–136, jan 1881.

[52] J. Delhommelle and P. Millie, "Inadequacy of the Lorentz-Berthelot combining rules for accurate predictions of equilibrium properties by molecular simulation," *Molecular Physics*, vol. 99, pp. 619–625, apr 2001.

[53] D. Boda and D. Henderson, "The effects of deviations from LorentzâĂŞBerthelot rules on the properties of a simple mixture," *Molecular Physics*, vol. 106, pp. 2367–2370, oct 2008.

[54] W. Song, P. J. Rossky, and M. Maroncelli, "Modeling alkane+perfluoroalkane interactions using all-atom potentials: Failure of the usual combining rules," *The Journal of Chemical Physics*, vol. 119, pp. 9145–9162, nov 2003.

[55] M. Waldman and A. Hagler, "New combining rules for rare gas van der waals parameters," *Journal of Computational Chemistry*, vol. 14, pp. 1077–1084, sep 1993.

[56] B. E. F. Fender and G. D. Halsey, "Second Virial Coefficients of Argon, Krypton, and ArgonâĂŘKrypton Mixtures at Low Temperatures," *The Journal of Chemical Physics*, vol. 36, pp. 1881–1888, apr 1962.

[57] C. L. Kong, "Combining rules for intermolecular potential parameters. II. Rules for the LennardâĂŘJones (12âĂŞ6) potential and the Morse potential," *The Journal of Chemical Physics*, vol. 59, pp. 2464–2467, sep 1973.

[58] B. A. Wells and A. L. Chaffee, "Ewald Summation for Molecular Simulations," *Journal of Chemical Theory and Computation*, vol. 11, pp. 3684–3695, aug 2015.

[59] P. P. Ewald, "Die Berechnung optischer und elektrostatischer Gitterpotentiale," *Annalen der Physik*, vol. 369, pp. 253–287, jan 1921.

[60] A. Y. Toukmaji and J. A. Board, "Ewald summation techniques in perspective: A survey," *Computer Physics Communications*, vol. 95, pp. 73–92, jun 1996.

[61] V. Ballenegger, J. J. Cerdà, and C. Holm, "How to convert SPME to P3M: Influence functions and error estimates," *Journal of Chemical Theory and Computation*, vol. 8, pp. 936–947, mar 2012.

[62] K. P. Esler, *Long-Range Potentials and the Ewald Method*. PhD thesis, University of Illinois at Urbana-Champaign.

[63] O. N. Osychenko, G. E. Astrakharchik, and J. Boronat, "Ewald method for polytropic potentials in arbitrary dimensionality," *Molecular Physics*, vol. 110, pp. 227–247, feb 2012.

[64] T. Darden, D. York, and L. Pedersen, "Particle mesh Ewald: An N.log(N) method for Ewald sums in large systems," *J. Chem. Phys.*, vol. 98, pp. 10089–10092, jun 1993.

[65] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A smooth particle mesh Ewald method," *The Journal of Chemical Physics*, vol. 103, no. 19, pp. 8577–8593, 1995.

[66] R. W. Hockney and J. Eastwood, *Computer Simulation Using Particles*. New York: McGraw-Hill, 1981.

[67] C. M. Baker, "Polarizable force fields for molecular dynamics simulations of biomolecules," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 5, pp. 241–254, mar 2015.

[68] P. Ball, "Water âĂŤ an enduring mystery," *Nature*, vol. 452, pp. 291–292, mar 2008.

[69] J. Chen, *THEORY AND APPLICATIONS OF FLUCTUATING-CHARGE MODELS*. Doctoral, University of Illinois at Urbana-Champaign, 2009.

[70] S. W. Rick, S. J. Stuart, and B. J. Berne, "Dynamical fluctuating charge force fields: Application to liquid water," *J. Chem. Phys.*, vol. 101, pp. 6141–6156, oct 1994.

[71] A. K. Rappe and W. A. Goddard, "Charge equilibration for molecular dynamics simulations," *The Journal of Physical Chemistry*, vol. 95, pp. 3358–3363, apr 1991.

[72] P. Ren and J. W. Ponder, "Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations," *Journal of Computational Chemistry*, vol. 23, pp. 1497–1506, oct 2002.

[73] J. Huang, P. E. M. Lopes, B. Roux, and A. D. MacKerell, "Recent Advances in Polarizable Force Fields for Macromolecules: Microsecond Simulations of Proteins Using the Classical Drude Oscillator Model," *The Journal of Physical Chemistry Letters*, vol. 5, pp. 3144–3150, sep 2014.

[74] 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, "Current Status of the AMOEBA Polarizable Force Field," *The Journal of Physical Chemistry B*, vol. 114, pp. 2549–2564, mar 2010.

[75] H. R. Glyde, "PHYS813 Class Notes," tech. rep., University of Delaware, 2010.

[76] W. Heisenberg, "Ueber den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik," *Zeitschrift fuer Physik*, vol. 43, pp. 172–198, mar 1927.

[77] E. H. Kennard, "Zur Quantenmechanik einfacher Bewegungstypen," *Zeitschrift fuer Physik*, vol. 44, pp. 326–352, apr 1927.

[78] A. P. Bhati, S. Wan, D. W. Wright, and P. V. Coveney, "Rapid, Accurate, Precise, and Reliable Relative Free Energy Prediction Using Ensemble Based Thermodynamic Integration," *Journal of Chemical Theory and Computation*, vol. 13, pp. 210–222, jan 2017.

[79] A. P. Bhati, S. Wan, Y. Hu, B. Sherborne, and P. V. Coveney, "Uncertainty Quantification in Alchemical Free Energy Methods," *Journal of Chemical Theory and Computation*, vol. 14, pp. 2867–2880, jun 2018.

[80] C. D. Christ, A. E. Mark, and W. F. Van Gunsteren, "Basic ingredients of free energy calculations: A review," *Journal of Computational Chemistry*, vol. 31, no. 8, pp. 1569–1582, 2010.

[81] T. Steinbrecher, I. Joung, and D. A. Case, "Soft-core potentials in thermodynamic integration: comparing one- and two-step transformations.," *Journal of computational chemistry*, vol. 32, pp. 3253–63, nov 2011.

[82] T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber, and W. F. van Gunsteren, "Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations," *Chemical Physics Letters*, vol. 222, pp. 529–539, jun 1994.

[83] C. H. Bennett, "Efficient estimation of free energy differences from Monte Carlo data," *Journal of Computational Physics*, vol. 22, pp. 245–268, oct 1976.

[84] R. W. Zwanzig, "High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases," *Journal of Chemical Physics*, vol. 22, no. 8, pp. 1420–1426, 1954.

[85] G. M. Torrie and J. P. Valleau, "Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling," *Journal of Computational Physics*, vol. 23, no. 2, pp. 187–199, 1977.

[86] Z. Yanxiang, "Brief Introduction to The Thermostats," tech. rep., University of California San Diego, 2008.

[87] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," *The Journal of Chemical Physics*, vol. 81, pp. 3684–3690, oct 1984.

[88] H. C. Andersen, "Molecular dynamics simulations at constant pressure and/or temperature," *The Journal of Chemical Physics*, vol. 72, pp. 2384–2393, feb 1980.

[89] S. Nosé, "A molecular dynamics method for simulations in the canonical ensemble," *Molecular Physics*, vol. 52, pp. 255–268, jun 1984.

[90] W. G. Hoover, "Canonical dynamics: Equilibrium phase-space distributions," *Physical Review A*, vol. 31, pp. 1695–1697, mar 1985.

[91] S. Nosé, "A unified formulation of the constant temperature molecular dynamics methods," *The Journal of Chemical Physics*, vol. 81, pp. 511–519, jul 1984.

[92] H. A. Posch, W. G. Hoover, and F. J. Vesely, "Canonical dynamics of the Nosé oscillator: Stability, order, and chaos," *Physical Review A*, vol. 33, pp. 4253–4265, jun 1986.

[93] J. Finnerty, "Molecular dynamics meets the physical world: Thermostats and barostats," tech. rep., German Research School for Simulation Sciences, 2011.

[94] M. Parrinello and A. Rahman, "Crystal Structure and Pair Potentials: A Molecular-Dynamics Study," *Physical Review Letters*, vol. 45, pp. 1196–1199, oct 1980.

[95] A. Malani, S. M. Auerbach, and P. A. Monson, "Probing the Mechanism of Silica Polymerization at Ambient Temperatures using Monte Carlo Simulations," *The Journal of Physical Chemistry Letters*, vol. 1, pp. 101025102748084—-3224, 2010.

[96] K. A. Fichthorn and W. H. Weinberg, "Theoretical foundations of dynamical Monte Carlo simulations," *The Journal of Chemical Physics*, vol. 95, pp. 1090–1096, jul 1991.

[97] M. Born and P. Jordan, "Zur Quantenmechanik," *Zeitschrift fuer Physik*, vol. 34, pp. 858–888, dec 1925.

[98] A. Einstein, "Zur Elektrodynamik bewegter Körper," *Annalen der Physik*, vol. 322, pp. 891–921, jan 1905.

[99] A. Einstein, *UÌĹber die spezielle und die allgemeine RelativitaÌĹtstheorie*. Springer Spektrum, 24 ed., 1916.

[100] D. R. Hartree, "The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part II. Some Results and Discussion," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 24, pp. 111–132, jan 1928.

[101] V. Fock, "Naeherungsmethode zur Loesung des quantenmechanischen Mehrkoerperproblems," *Zeitschrift fuer Physik*, vol. 61, pp. 126–148, jan 1930.

[102] J. C. Slater, "The Self Consistent Field and the Structure of Atoms," *Physical Review*, vol. 32, pp. 339–348, sep 1928.

[103] J. A. Gaunt, "A Theory of Hartree's Atomic Fields," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 24, pp. 328–342, apr 1928.

[104] D. R. Hartree and W. Hartree, "Self-consistent field, with exchange, for beryllium," *Proceedings of the Royal Society of London. Series A - Mathematical and Physical Sciences*, vol. 150, pp. 9–33, may 1935.

[105] M. Born and R. Oppenheimer, "Zur Quantentheorie der Molekeln," *Annalen der Physik*, vol. 389, pp. 457–484, jan 1927.

[106] J. C. Slater, "The Theory of Complex Spectra," *Physical Review*, vol. 34, pp. 1293–1322, nov 1929.

[107] D. S. Sholl and J. A. Steckel, *Density functional theory: a practical introduction.* Wiley, 2009.

[108] R. M. Eisberg and R. Resnick, *Quantum physics of atoms, molecules, solids, nuclei, and particles.* Wiley, 1985.

[109] H. J. Monkhorst and J. D. Pack, "Special points for Brillouin-zone integrations," *Physical Review B*, vol. 13, pp. 5188–5192, jun 1976.

[110] E. S. Sachs, J. Hinze, and N. H. Sabelli, "Frozen core approximation, a pseudopotential method tested on six states of NaH," *The Journal of Chemical Physics*, vol. 62, p. 3393, sep 1975.

[111] J. C. Slater, "Atomic Shielding Constants," *Physical Review*, vol. 36, p. 57, jul 1930.

[112] T. Kato, "On the eigenfunctions of manyâĂŘparticle systems in quantum mechanics," *Communications on Pure and Applied Mathematics*, vol. 10, pp. 151–177, jan 1957.

[113] S. Boys, "Electronic wave functions - I. A general method of calculation for the stationary states of any molecular system," *Proceedings of the Royal Society of London. Series A, Mathematical and physical sciences.*, vol. 200, no. 1063, 1950.

[114] R. Ditchfield, W. J. Hehre, and J. A. Pople, "SelfâĂŘConsistent MolecularâĂŘOrbital Methods. IX. An Extended GaussianâĂŘType Basis for MolecularâĂŘOrbital Studies of Organic Molecules," *The Journal of Chemical Physics*, vol. 54, pp. 724–728, jan 1971.

[115] V. Hanninen, "Introduction to Computational Chemistry," tech. rep., University of Helsinki, 2012.

[116] T. H. Dunning, "Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen," *The Journal of Chemical Physics*, vol. 90, pp. 1007–1023, jan 1989.

[117] P. Manninen and J. Vaara, "Systematic Gaussian Basis-Set Limit Using Completeness-Optimized Primitive Sets. A Case for Magnetic Properties," *J Comput Chem*, vol. 27, pp. 434–445, 2006.

[118] F. Jensen, "Polarization consistent basis sets: Principles," *The Journal of Chemical Physics*, vol. 115, pp. 9113–9125, nov 2001.

[119] T. Kupka and C. Lim, "Polarization-Consistent versus Correlation-Consistent Basis Sets in Predicting Molecular and Spectroscopic Properties," *J. Phys. Chem. A*, vol. 111, pp. 1927–1932, 2007.

[120] F. B. van Duijneveldt, J. G. C. M. van Duijneveldt-van de Rijdt, and J. H. van Lenthe, "State of the Art in Counterpoise Theory," *Chemical Reviews*, vol. 94, pp. 1873–1885, nov 1994.

[121] I. Mayer and P. Valiron, "Second order Mo/llerâĂŞPlesset perturbation theory without basis set superposition error," *The Journal of Chemical Physics*, vol. 109, p. 3360, nov 1998.

[122] D. P. Craig, "Configurational interaction in molecular orbital theory. A higher approximation in the non-empirical method," *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 200, pp. 474–486, feb 1950.

[123] J. Čížek, "Origins of coupled cluster technique for atoms and molecules," *Theoretica Chimica Acta*, vol. 80, pp. 91–94, mar 1991.

[124] C. Møller and M. S. Plesset, "Note on an Approximation Treatment for Many-Electron Systems," *Physical Review*, vol. 46, pp. 618–622, oct 1934.

[125] V. Magnasco, *Elementary Molecular Quantum Mechanics: Mathematical Methods and Applications.* Elsevier Science, 2013.

[126] K. Sahan Thanthiriwatte, "Coupled-Cluster Theory: Lecture Series in Electronic Structure Theory," tech. rep., Georgia Institute of Technology, 2010.

[127] V. Hanninen, "Introduction to Computational Chemistry," tech. rep., University of Helsinki, 2013.

[128] P. Hohenberg and W. Kohn, "Inhomogeneous Electron Gas," *Physical Review*, vol. 136, pp. B864–B871, nov 1964.

[129] W. Kohn and L. J. Sham, "Self-Consistent Equations Including Exchange and Correlation Effects," *Physical Review*, vol. 140, pp. A1133–A1138, nov 1965.

[130] D. C. Langreth and M. J. Mehl, "Beyond the local-density approximation in calculations of ground-state electronic properties," *Physical Review B*, vol. 28, pp. 1809–1834, aug 1983.

[131] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized Gradient Approximation Made Simple," *Physical Review Letters*, vol. 77, pp. 3865–3868, oct 1996.

[132] Y. Wang and J. P. Perdew, "Spin scaling of the electron-gas correlation energy in the high-density limit," *Physical Review B*, vol. 43, pp. 8911–8916, apr 1991.

[133] A. D. Becke, "Density-functional exchange-energy approximation with correct asymptotic behavior," *Physical Review A*, vol. 38, pp. 3098–3100, sep 1988.

[134] C. Lee, W. Yang, and R. G. Parr, "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density," *Physical Review B*, vol. 37, pp. 785–789, jan 1988.

[135] R. Car and M. Parrinello, "Unified Approach for Molecular Dynamics and Density-Functional Theory," *Physical Review Letters*, vol. 55, pp. 2471–2474, nov 1985.

[136] J. Hutter, "Car-Parrinello molecular dynamics," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 2, pp. 604–612, jul 2012.

[137] I. Dmochowski and R. Eckenhoff, *Chemical and Biochemical Approaches for the Study of Anesthetic Function, Part A*. Academic Press, 602 ed., 2018.

[138] K. M. Visscher and D. P. Geerke, "Deriving Force-Field Parameters from First Principles Using a Polarizable and Higher Order Dispersion Model," *Journal of Chemical Theory and Computation*, vol. 15, pp. 1875–1883, mar 2019.

[139] O. M. Becker, A. D. MacKerell, B. Roux, and M. Watanabe, *Computational biochemistry and biophysics*. Taylor & Francis Ltd, 1 ed., 2001.

[140] S. Grimme, "Density functional theory with London dispersion corrections," *WIREs Computational Molecular Science*, vol. 1, pp. 211–228, mar 2011.

[141] S. Grimme, "A general quantum mechanically derived force field (QMDFF) for molecules and condensed phase simulations," *Journal of Chemical Theory and Computation*, vol. 10, pp. 4497–4514, oct 2014.

[142] D. J. Cole, J. Z. Vilseck, J. Tirado-Rives, M. C. Payne, and W. L. Jorgensen, "Biomolecular Force Field Parameterization via Atoms-in-Molecule Electron Density Partitioning," *Journal of Chemical Theory and Computation*, vol. 12, no. 5, pp. 2312–2323, 2016.

[143] Q. Huo, D. I. Margolese, U. Ciesla, D. G. Demuth, P. Feng, T. E. Gier, P. Sieger, A. Firouzi, B. F. Chmelka, F. Schuth, and G. D. Stucky, "Organization of Organic Molecules with Inorganic Molecular Species into Nanocomposite Biphase Arrays," *Chem. Mater.*, vol. 6, pp. 1176–1191, 1994.

[144] R. Mellaerts, C. A. Aerts, J. V. Humbeeck, P. Augustijns, G. V. den Mooter, and J. A. Martens, "Enhanced release of itraconazole from ordered mesoporous SBA-15 silica materials," *Chemical Communications*, no. 13, p. 1375, 2007.

[145] A. Katiyar, S. Yadav, P. G. Smirniotis, and N. G. Pinto, "Synthesis of ordered large pore SBA-15 spherical particles for adsorption of biomolecules," *Journal of Chromatography A*, vol. 1122, pp. 13–20, jul 2006.

[146] S. C. Chien, S. M. Auerbach, and P. A. Monson, "Modeling the self-assembly of silica-templated nanoparticles in the initial stages of zeolite formation," *Langmuir*, vol. 31, pp. 4940–4949, may 2015.

[147] F. S. Emami, V. Puddu, R. J. Berry, V. Varshney, S. V. Patwardhan, C. C. Perry, and H. Heinz, "Force field and a surface model database for silica to simulate interfacial properties in atomic resolution," *Chemistry of Materials*, vol. 26, pp. 2647–2658, apr 2014.

[148] J. D. Deetz and R. Faller, "Parallel optimization of a reactive force field for polycondensation of alkoxysilanes," *Journal of Physical Chemistry B*, vol. 118, pp. 10966–10978, sep 2014.

[149] M. Kley, A. Kempter, V. Boyko, and K. Huber, "Mechanistic studies of silica polymerization from supersaturated aqueous solutions by means of time-resolved light scattering," *Langmuir*, vol. 30, pp. 12664–12674, oct 2014.

[150] S. Bhattacharyya, G. Lelong, and M.-L. Saboungi, "Recent progress in the synthesis and selected applications of MCM-41: a short review," *Journal of Experimental Nanoscience*, vol. 1, pp. 375–395, sep 2006.

[151] A. Malani, S. M. Auerbach, and P. A. Monson, "Monte Carlo simulations of silica polymerization and network formation," *Journal of Physical Chemistry C*, vol. 115, pp. 15988–16000, aug 2011.

[152] G. M. Whitesides and M. Boncheva, "Beyond molecules: Self-assembly of mesoscopic and macroscopic components," in *PNAS*, vol. 16, pp. 4769–4774, 2002.

[153] D. Lombardo, M. A. Kiselev, S. Magazù, and P. Calandra, "Amphiphiles Self-Assembly: Basic Concepts and Future Perspectives of Supramolecular Approaches," *Advances in Condensed Matter Physics*, vol. 2015, no. 151683, pp. 1–22, 2015.

[154] Y. Mai and A. Eisenberg, "Self-assembly of block copolymers," *Chemical Society Reviews*, vol. 41, pp. 5969–5985, sep 2012.

[155] L. Magnus Bergstrom, "Thermodynamics of Self-Assembly," in *Application of Thermodynamics to Biological and Materials Science*, InTech, jan 2011.

[156] Z. Lin, J. J. Cai, L. E. Scriven, and H. T. Davis, "Spherical-to-wormlike micelle transition in CTAB solutions," *Journal of Physical Chemistry*, vol. 98, no. 23, pp. 5984–5993, 1994.

[157] E. M. Johansson, *Controlling the Pore Size and Morphology of Mesoporous Silica.* PhD thesis, Linkoping University, 2010.

[158] M. Manzano, V. Aina, C. O. Areán, F. Balas, V. Cauda, M. Colilla, M. R. Delgado, and M. Vallet-Regí, "Studies on MCM-41 mesoporous silica for drug delivery: Effect of particle morphology and amine functionalization," *Chemical Engineering Journal*, vol. 137, pp. 30–37, mar 2008.

[159] H. E. Bergna, "Colloid Chemistry of Silica," in *The Colloid Chemistry of Silica*, pp. 1–47, ACS Publications, may 1994.

[160] D. J. Belton, O. Deschaume, and C. C. Perry, "An overview of the fundamentals of the chemistry of silica with relevance to biosilicification and technological advances," *FEBS Journal*, vol. 279, pp. 1710–1720, may 2012.

[161] J. C. Pereira, C. R. Catlow, and G. D. Price, "Molecular dynamics simulation of methanolic and ethanolic silica-based sol-gel solutions at ambient temperature and pressure," *Journal of Physical Chemistry A*, vol. 106, no. 1, pp. 130–148, 2002.

[162] T. Matsoukas and E. Gulari, "Dynamics of growth of silica particles from ammonia-catalyzed hydrolysis of tetra-ethyl-orthosilicate," *Journal of Colloid And Interface Science*, vol. 124, no. 1, pp. 252–261, 1988.

[163] G. H. Bogush and C. F. Zukoski IV, "Studies of the kinetics of the precipitation of uniform silica particles through the hydrolysis and condensation of silicon alkoxides," *Journal of Colloid And Interface Science*, vol. 142, pp. 1–18, mar 1991.

[164] V. Chaudhary and S. Sharma, "An overview of ordered mesoporous material SBA-15: synthesis, functionalization and application in oxidation reactions," *Journal of Porous Materials*, vol. 24, pp. 741–749, jun 2017.

[165] S. A. Idris, C. M. Davidson, C. McManamon, M. A. Morris, P. Anderson, and L. T. Gibson, "Large pore diameter MCM-41 and its application for lead removal from aqueous media," *Journal of Hazardous Materials*, vol. 185, pp. 898–904, jan 2011.

[166] S. Bhattacharya, B. Coasne, F. R. Hung, and K. E. Gubbins, "Modeling micelle-templated mesoporous material SBA-15: Atomistic model and gas adsorption studies," *Langmuir*, vol. 25, no. 10, pp. 5802–5813, 2009.

[167] K. Cassiers, T. Linssen, M. Mathieu, M. Benjelloun, K. Schrijnemakers, P. Van Der Voort, P. Cool, and E. F. Vansant, "A detailed study of thermal, hydrothermal, and mechanical stabilities of a wide range of surfactant assembled mesoporous silicas," *Chemistry of Materials*, vol. 14, no. 5, pp. 2317–2324, 2002.

[168] R. Narayan, U. Nayak, A. Raichur, and S. Garg, "Mesoporous Silica Nanoparticles: A Comprehensive Review on Synthesis and Recent Advances," *Pharmaceutics*, vol. 10, p. 118, aug 2018.

[169] R. E. Morsi and R. S. Mohamed, "Nanostructured mesoporous silica: Influence of the preparation conditions on the physical-surface properties for efficient organic dye uptake," *Royal Society Open Science*, vol. 5, mar 2018.

[170] M. Leclerc and R. Gauvin, *Functional Materials : For Energy, Sustainable Development and Biomedical Sciences*. De Gruyter, 2014.

[171] S. Che, Y. Sakamoto, O. Terasaki, and T. Tatsumi, "Control of crystal morphology of SBA-1 mesoporous silica," *Chemistry of Materials*, vol. 13, no. 7, pp. 2237–2239, 2001.

[172] M. J. Kim and R. Ryoo, "Synthesis and pore size control of cubic mesoporous silica SBA-1," *Chemistry of Materials*, vol. 11, no. 2, pp. 487–491, 1999.

[173] M. Pérez-Mendoza, J. Gonzalez, P. A. Wright, and N. A. Seaton, "Structure of the mesoporous silica SBA-2, determined by a percolation analysis of adsorption," *Langmuir*, vol. 20, pp. 9856–9860, oct 2004.

[174] D. Zhao, Y. Wan, and W. Zhou, *Ordered Mesoporous Materials*. Weinheim, Germany: Wiley-VCH, jan 2013.

[175] A. E. Garcia-Bennett, S. Williamson, P. A. Wright, and I. J. Shannon, "Control of structure, pore size and morphology of three-dimensionally ordered mesoporous silicas prepared using the dicationic surfactant [CH3(CH2)15N(CH3)2(CH 2)3N(CH3)3]BR2," *Journal of Materials Chemistry*, vol. 12, pp. 3533–3540, dec 2002.

[176] R. Guillet-Nicolas, R. Ahmad, K. A. Cychosz, F. Kleitz, and M. Thommes, "Insights into the pore structure of KIT-6 and SBA-15 ordered mesoporous silica-recent advances by combining physical adsorption with mercury porosimetry," *New Journal of Chemistry*, vol. 40, pp. 4351–4360, may 2016.

[177] B. Zhou, C. Y. Li, N. Qi, M. Jiang, B. Wang, and Z. Q. Chen, "Pore structure of mesoporous silica (KIT-6) synthesized at different temperatures using positron as a nondestructive probe," *Applied Surface Science*, vol. 450, pp. 31–37, aug 2018.

[178] G. Gu, P. P. Ong, and C. Chu, "Thermal stability of mesoporous silica molecular sieve," *Journal of Physics and Chemistry of Solids*, vol. 60, no. 7, pp. 943–947, 1999.

[179] H. Chen and Y. Wang, "Preparation of MCM-41 with high thermal stability and complementary textural porosity," *Ceramics International*, vol. 28, pp. 541–547, jan 2002.

[180] A. Matsumoto, T. Sasaki, N. Nishimiya, and K. Tsutsumi, "Thermal stability and hydrophobicity of mesoporous silica FSM-16," *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, vol. 203, pp. 185–193, apr 2002.

[181] K. J. Edler, "Description of MCM-41 Structure," tech. rep., Australian National University, 2011.

[182] M. Vallet-Regí, M. Colilla, I. Izquierdo-Barba, and M. Manzano, "Mesoporous silica nanoparticles for drug delivery: Current insights," *Molecules*, vol. 23, no. 47, pp. 1–19, 2018.

[183] R. K. Iler, *The chemistry of silica: solubility, polymerization, colloid and surface properties, and biochemistry.* Wiley, 1979.

[184] J. S. Beck, J. C. Vartuli, G. J. Kennedy, C. T. Kresge, W. J. Roth, and S. E. Schramm, "Molecular or Supramolecular Templating: Defining the Role of Surfactant Chemistry in the Formation of Microporous and Mesoporous Molecular Sieves," *Chemistry of Materials*, vol. 6, pp. 1816–1821, oct 1994.

[185] C. Huiyong, W. Jianghong, L. Shaojuan, X. Hongxia, and Q. Yu, "MesoDyn and Experimental Approach to the Structural Fabrication and Pore-size Adjustment of SBA-15 Molecular Sieves," *Adsorpt. Sci. Technol.*, vol. 27, pp. 579–592, 2009.

[186] Z. He and P. Alexandridis, "Micellization thermodynamics of Pluronic P123 (EO20PO70EO20) amphiphilic block copolymer in aqueous Ethylammonium nitrate (EAN) solutions," *Polymers*, vol. 10, dec 2017.

[187] W. Sangchoom and R. Mokaya, "High temperature synthesis of exceptionally stable pure silica MCM-41 and stabilisation of calcined mesoporous silicas via refluxing in water," *Journal of Materials Chemistry*, vol. 22, pp. 18872–18878, sep 2012.

[188] P. Khodaee, N. Najmoddin, and S. Shahrad, "The effect of ethanol and temperature on the structural properties of mesoporous silica synthesized by sol-gel method," *Journal of Tissues and Materials*, vol. 1, no. 1, pp. 10–17, 2018.

[189] B. Lopez, L. D. Perez, S. Urrego, and M. Mesa, "Mesoporous Silica Applications: Characterization of the interactions between mesoporous silicas and rubbery matrices," *Macromol. Symp.*, vol. 258, pp. 129–141, 2007.

[190] T. L. Chew, A. L. Ahmad, and S. Bhatia, "Ordered mesoporous silica (OMS) as an adsorbent and membrane for separation of carbon dioxide (CO2)," *Advances in Colloid and Interface Science*, vol. 153, pp. 43–57, jan 2010.

[191] H.-J. Kim, H.-C. Yang, D.-Y. Chung, I.-H. Yang, Y. J. Choi, J.-K. Moon, and J. M. Coronado, "Functionalized Mesoporous Silica Membranes for CO2 Separation Applications," *Journal of Chemistry*, vol. 202867, pp. 1–9, 2015.

[192] P. N. Diagboya and E. D. Dikio, "Silica-based mesoporous materials; emerging designer adsorbents for aqueous pollutants removal and water treatment," *Microporous and Mesoporous Materials*, vol. 266, pp. 252–267, aug 2018.

[193] A. Taguchi and F. Schüth, "Ordered mesoporous materials in catalysis," *Microporous and Mesoporous Materials*, vol. 77, pp. 1–45, jan 2005.

[194] R. Luque, A. M. Balu, J. M. Campelo, M. D. Gracia, E. Losada, A. Pineda, A. A. Romero, and J. C. Serrano-Ruiz, "Catalytic applications of mesoporous silica-based materials," in *Catalysis*, vol. 24, pp. 253–280, RSC Publishing, feb 2012.

[195] J. Liang, Z. Liang, R. Zou, and Y. Zhao, "Heterogeneous Catalysis in Zeolites, Mesoporous Silica, and MetalâĂŞOrganic Frameworks," *Advanced Materials*, vol. 29, p. 1701139, aug 2017.

[196] S. Alberti, G. J. Soler-Illia, and O. Azzaroni, "Gated supramolecular chemistry in hybrid mesoporous silica nanoarchitectures: Controlled delivery and molecular transport in response to chemical, physical and biological stimuli," *Chemical Communications*, vol. 51, pp. 6050–6075, apr 2015.

[197] N. K. Mal, M. Fujiwara, and Y. Tanaka, "Photocontrolled reversible release of guest molecules from coumarin-modified mesoporous silica," *Nature*, vol. 421, pp. 350–353, jan 2003.

[198] Q. Fu, G. V. Rao, L. K. Ista, Y. Wu, B. P. Andrzejewski, L. A. Sklar, T. L. Ward, and G. P. López, "Control of molecular transport through stimuli-responsive or-

dered mesoporous materials," *Advanced Materials*, vol. 15, pp. 1262–1266, aug 2003.

[199] Q. Yang, S. Wang, P. Fan, L. Wang, Y. Di, K. Lin, and F. S. Xiao, "pH-responsive carrier system based on carboxylic acid modified mesoporous silica and polyelectrolyte for drug delivery," *Chemistry of Materials*, vol. 17, pp. 5999–6003, nov 2005.

[200] C. Y. Lai, B. G. Trewyn, D. M. Jeftinija, K. Jeftinija, S. Xu, S. Jeftinija, and V. S. Lin, "A mesoporous silica nanosphere-based carrier system with chemically removable CdS nanoparticle caps for stimuli-responsive controlled release of neurotransmitters and drug molecules," *Journal of the American Chemical Society*, vol. 125, pp. 4451–4459, apr 2003.

[201] F. Sancenõn, L. Pascual, M. Oroval, E. Aznar, and R. Martínez-Máñez, "Gated Silica Mesoporous Materials in Sensing Applications," *ChemistryOpen*, vol. 4, pp. 418–437, aug 2015.

[202] V. Meynen, P. Cool, and E. F. Vansant, "Verified syntheses of mesoporous materials," *Microporous and Mesoporous Materials*, vol. 125, pp. 170–223, oct 2009.

[203] S. Brunauer, P. H. Emmett, and E. Teller, "Adsorption of Gases in Multimolecular Layers," *Journal of the American Chemical Society*, vol. 60, pp. 309–319, feb 1938.

[204] F. Ambroz, T. J. Macdonald, V. Martis, and I. P. Parkin, "Evaluation of the BET Theory for the Characterization of Meso and Microporous MOFs," *Small Methods*, vol. 2, p. 1800173, nov 2018.

[205] S. Scaife, P. Kluson, and N. Quirke, "Characterization of Porous Materials by Gas Adsorption: Do Different Molecular Probes Give Different Pore Structures?," *Journal of Physical Chemistry B*, vol. 104, pp. 313–318, jan 2000.

[206] K. S. Sing, "Characterization of porous materials: Past, present and future," *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, vol. 241, pp. 3–7, jul 2004.

[207] E. A. Ustinov, "Nitrogen adsorption on silica surfaces of nonporous and mesoporous materials," *Langmuir*, vol. 24, pp. 6668–6675, jul 2008.

[208] E. P. Barrett, L. G. Joyner, and P. P. Halenda, "The Determination of Pore Volume and Area Distributions in Porous Substances. I. Computations from Nitrogen Isotherms," *Journal of the American Chemical Society*, vol. 73, pp. 373–380, jan 1951.

[209] R. Bardestani, G. S. Patience, and S. Kaliaguine, "Experimental methods in chemical engineering: specific surface area and pore size distribution measurements - BET, BJH, and DFT," *The Canadian Journal of Chemical Engineering*, vol. 97, pp. 2781–2791, nov 2019.

[210] Y. Ishii, Y. Nishiwaki, A. Al-Zubaidi, and S. Kawasaki, "Pore-size determination in ordered mesoporous materials using powder X-ray diffraction," *Journal of Physical Chemistry C*, vol. 117, pp. 18120–18130, sep 2013.

[211] C. Chen, S. L. Burkett, H. Li, and M. E. Davis, "Studies on mesoporous materials II. Synthesis mechanism of MCM-41," *Microporous Mater.*, vol. 2, pp. 27–34, 1993.

[212] A. Monnier, F. Schuth, Q. Huo, D. Kumar, D. Margolese, R. S. Maxwell, G. D. Stucky, M. Krishnamurty, P. Petroff, A. Firouzi, M. Janicke, and B. F. Chmelka, "Cooperative Formation of Inorganic-Organic Interfaces in the Synthesis of Silicate Mesostructures," *Science (80-.).*, vol. 261, pp. 1299–1303, 1993.

[213] A. Firouzi, F. Atef, A. G. Oertli, G. D. Stucky, and B. F. Chmelka, "Alkaline Lyotropic Silicate-Surfactant Liquid Crystals," *J. Am. Chem. Soc.*, vol. 119, pp. 3596–3610, 1997.

[214] Y. Lee, D. Surjadi, and J. F. Rathman, "Effects of Aluminate and Silicate on the Structure of Quaternary Ammonium Surfactant Aggregates," *Langmuir*, vol. 12, pp. 6202–6210, 1996.

[215] A. Albuquerque, C. Vautier-Giongo, and H. O. Pastore, "Physical chemistry of nanostructured molecular sieves by the study of phase diagrams: the case of

the cetyltrimethylammonium bromideâĂŞtetramethylammonium silicateâĂŞwater system," *J. Colloid Interface Sci.*, vol. 284, pp. 687–693, 2005.

[216] C. Vautier-Giongo and H. O. Pastore, "Micellization of CTAB in the presence of silicate anions and the exchange between bromide and silicate at the micelle surface: A step to understand the formation of mesoporous molecular sieves at extremely low surfactant and silicate concentrations," *J. Colloid Interface Sci.*, vol. 299, pp. 874–882, 2006.

[217] J. C. Vartuli, K. D. Schmitt, C. T. Kresge, W. J. Roth, M. E. Leonowicz, S. B. Mccullen, S. D. Hellring, J. S. Beck, J. L. Schlenker, D. H. Olson, and E. W. Sheppard, "Effect of Surfactant/Silica Molar Ratios on the Formation of Mesoporous Molecular Sieves: Inorganic Mimicry of Surfactant Liquid-Crystal Phases and Mechanistic Implications," *Chem. Mater.*, vol. 6, pp. 2317–2326, 1994.

[218] R. Zana, J. Frasch, M. Soulard, B. Lebeau, and J. Patarin, "Fluorescence Probing Investigations of the Mechanism of Formation of Organized Mesoporous Silica," *Langmuir*, vol. 15, pp. 2603–2606, 1999.

[219] J. Frasch, B. Lebeau, M. Soulard, J. Patarin, and R. Zana, "In Situ Investigations on Cetyltrimethylammonium Surfactant/Silicate Systems, Precursors of Organized Mesoporous MCM-41-Type Siliceous Materials," *Langmuir*, vol. 16, pp. 9049–9057, 2000.

[220] A. Galarneau, F. Di Renzo, F. Fajula, L. Mollo, B. Fubini, and M. F. Ottaviani, "Kinetics of Formation of Micelle-Templated Silica Mesophases Monitored by Electron Paramagnetic Resonance," *J. Colloid Interface Sci.*, vol. 201, pp. 105–117, 1998.

[221] D. Baute, V. Frydman, H. Zimmermann, S. Kababya, and D. Goldfarb, "Properties of the Silica Layer during the Formation of MCM-41 Studied by EPR of a Silica-Bound Spin Probe," *J. Phys. Chem. B*, vol. 109, pp. 7807–7816, 2005.

[222] O. Regev, "Nucleation Events during the Synthesis of Mesoporous Materials Using Liquid Crystalline Templating," *Langmuir*, vol. 12, pp. 4940–4944, 1996.

[223] S. Sadasivan, C. E. Fowler, D. Khushalani, and S. Mann, "Nucleation of MCM-41 Nanoparticles by Internal Reorganization of Disordered and Nematic-Like Silicaâ ĂŞSurfactant Clusters We thank the University of Bristol for the support of an ORS award to S.S., and Dr. S. A. Davis for help with electron microscopy.," *Angew. Chemie Int. Ed.*, vol. 41, p. 2151, jun 2002.

[224] I. Beurroies, P. A. Gren, G. Bu, J. B. Rosenholm, H. Amenitsch, R. Denoyel, and M. Linden, "Detailed in Situ XRD and Calorimetric Study of the Formation of Silicate/Mixed Surfactant Mesophases under Alkaline Conditions. Influence of Surfactant Chain Length and Synthesis Temperature," *J. Phys. Chem. B*, vol. 110, pp. 16254–16260, 2006.

[225] B. P. Feuston and J. B. Higgins, "Model Structures for {MCM-41} Materials: A Molecular Dynamics Simulation," *J. Phys. Chem.*, vol. 98, no. 16, pp. 4459–4462, 1994.

[226] C. D. Williams, K. P. Travis, N. A. Burton, and J. H. Harding, "A new method for the generation of realistic atomistic models of siliceous MCM-41," *Microporous Mesoporous Mater.*, no. 228, pp. 215–223, 2016.

[227] C. A. Koh, T. Montanari, R. I. Nooney, S. F. Tahir, and R. E. Westacott, "Experimental and Computer Simulation Studies of the Removal of Carbon Dioxide from Mixtures with Methane Using AlPO4-5 and MCM-41," *Langmuir*, no. 15, pp. 6043–6049, 1999.

[228] Y. Jing, L. Wei, Y. Wang, and Y. Yu, "Molecular simulation of MCM-41: Structural properties and adsorption of CO2 , N2 and flue gas," *Chemical Engineering Journal*, vol. 220, pp. 264–275, 2013.

[229] F. R. Siperstein and K. E. Gubbins, "Synthesis and characterization of templated mesoporous materials using molecular simulation," *Molecular Simulation*, vol. 27, pp. 339–352, nov 2001.

[230] F. R. Siperstein and K. E. Gubbins, "Phase separation and liquid crystal self-assembly in surfactant-inorganic-solvent systems," *Langmuir*, vol. 19, no. 6, pp. 2049–2057, 2003.

[231] R. G. Larson, L. E. Scriven, and H. T. Davis, "Monte Carlo simulation of model amphiphile-oilâĂŞwater systems Monte Carlo simulation of model amphiphile-oil-water systems," *The Journal of Chemical Physics*, vol. 83, no. 10, pp. 2411–2420, 1985.

[232] S. Bhattacharya and K. E. Gubbins, "Modeling triblock surfactant-templated mesostructured cellular foams," *Journal of Chemical Physics*, vol. 123, p. 134907, oct 2005.

[233] J. S. Lettow, Y. J. Han, P. Schmidt-Winkel, P. Yang, D. Zhao, G. D. Stucky, and J. Y. Ying, "Hexagonal to Mesocellular Foam Phase Transition in Polymer-Templated Mesoporous Silicas," *Langmuir*, vol. 16, pp. 8291–8295, 2000.

[234] P. Altevogt, O. Evers, J. Fraaije, N. Maurits, and B. van Vlimmeren, "The MesoDyn project: software for mesoscale chemical engineering," *Journal of Molecular Structure: THEOCHEM*, vol. 463, pp. 139–143, apr 1999.

[235] B. A. C. van Vlimmeren, M. N. M., . A. V. , Zvelindovsky, S. G. J. A., J. G. E. M. Fraaije, B. A. C. van Vlimmeren, M. N. M., . A. V. , Zvelindovsky, S. G. J. A., and J. G. E. M. Fraaije, "Simulation of 3D Mesoscale Structure Formation in Concentrated Aqueous Solution of the Triblock Polymer Surfactants (Ethylene Oxide)13(Propylene Oxide)30(Ethylene Oxide)13 and (Propylene Oxide)19(Ethylene Oxide)33(Propylene Oxide)19. Application of Dynami," *Macromolecules*, vol. 32, no. 3, pp. 646–656, 1999.

[236] S.-L. Yuan, X.-Q. Zhang, and K.-Y. Chan, "Effects of Shear and Charge on the Microphase Formation of P123 Polymer in the SBA-15 Synthesis Investigated by Mesoscale Simulations," *Langmuir*, vol. 25, pp. 2034–2045, 2009.

[237] C. Huiyong, W. Ying, T. Yanzhi, L. Xuehui, Q. Yu, and X. Hongxia, "Mesoscopic simulation of surfactant/silicate self-assembly in the mesophase forma-

tion of SBA-15 under charge matching interactions," *European Polymer Journal*, vol. 48, pp. 1892–1900, 2012.

[238] A. Patti, A. D. Mackie, and F. R. Siperstein, "Monte Carlo simulation of self-assembled ordered hybrid materials," *Langmuir*, vol. 23, no. 12, pp. 6771–6780, 2007.

[239] B. Hatton, K. Landskron, W. Whitnall, D. Perovic, and G. A. Ozin, "Past, Present, and Future of Periodic Mesoporous Organosilicass - The PMOs," *Acc. Chem. Res.*, vol. 38, pp. 305–312, 2005.

[240] A. Patti, F. R. Siperstein, and A. D. Mackie, "Phase behavior of model surfactants in the presence of hybrid particles," *Journal of Physical Chemistry C*, vol. 111, no. 43, pp. 16035–16044, 2007.

[241] M. Jorge, J. R. Gomes, M. N. D. Cordeiro, and N. A. Seaton, "Molecular simulation of silica/surfactant self-assembly in the synthesis of periodic mesoporous silicas," *Journal of the American Chemical Society*, vol. 129, pp. 15414–15415, dec 2007.

[242] W. L. Jorgensen and J. Gao, "Monte Carlo Simulations of the Hydration of Ammonium and Carboxylate Ions," *J. Phys. Chem*, vol. 90, pp. 2174–2182, 1986.

[243] J. R. Gomes, M. N. D. Cordeiro, and M. Jorge, "Gas-phase molecular structure and energetics of anionic silicates," *Geochimica et Cosmochimica Acta*, vol. 72, no. 17, pp. 4421–4439, 2008.

[244] C. T. G. Knight, "Are zeolite secondary building units really red herrings?," *Zeolites*, vol. 10, pp. 140–144, 1990.

[245] G. Pérez-Sánchez, S.-C. Chien, J. R. B. Gomes, M. N. D. S. Cordeiro, S. M. Auerbach, P. A. Monson, and M. Jorge, "Multiscale Model for the Templated Synthesis of Mesoporous Silica: The Essential Role of Silica Oligomers," *Chemistry of Materials*, vol. 28, pp. 2715–2727, apr 2016.

[246] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. De Vries, "The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations," *J. Phys. Chem. B*, vol. 111, pp. 7812–7824, 2007.

[247] R. Wu, M. Deng, B. Kong, and X. Yang, "Coarse-Grained Molecular Dynamics Simulation of Ammonium Surfactant Self-Assemblies: Micelles and Vesicles," *J. Phys. Chem. B*, vol. 113, pp. 15010–15016, 2009.

[248] M. Jorge, "Molecular dynamics simulation of self-assembly of n-decyltrimethylammonium bromide micelles," *Langmuir*, vol. 24, no. 11, pp. 5714–5725, 2008.

[249] S. C. Chien, G. Pérez-Sánchez, J. R. Gomes, M. N. D. Cordeiro, M. Jorge, S. M. Auerbach, and P. A. Monson, "Molecular Simulations of the Synthesis of Periodic Mesoporous Silica Phases at High Surfactant Concentrations," *Journal of Physical Chemistry C*, vol. 121, pp. 4564–4575, mar 2017.

[250] R. Futamura, M. Jorge, and J. R. B. Gomes, "Structures and energetics of organosilanes in the gaseous phase: a computational study," *Theoretical Chemistry Accounts*, vol. 132, pp. 1323–1327, 2013.

[251] R. Futamura, M. Jorge, and J. R. B. Gomes, "Role of the organic linker in the early stages of the templated synthesis of PMOs," *Phys. Chem. Chem. Phys.*, vol. 15, p. 6166, 2013.

[252] S. Inagaki, S. Guan, T. Ohsuna, and O. Terasaki, "An ordered mesoporous organosilica hybrid material with a crystal-like wall structure," *Nature*, vol. 416, pp. 304–307, mar 2002.

[253] B. P. Feuston and S. H. Garofalini, "Empirical three-body potential for vitreous silica," *The Journal of Chemical Physics*, vol. 89, pp. 5818–5824, nov 1988.

[254] N. Z. Rao and L. D. Gelb, "Molecular dynamics simulations of the polymerization of aqueous silicic acid and analysis of the effects of concentration on silica

polymorph distributions, growth mechanisms, and reaction kinetics," *Journal of Physical Chemistry B*, vol. 108, no. 33, pp. 12418–12428, 2004.

[255] Z. Jing, L. Xin, and H. Sun, "Replica exchange reactive molecular dynamics simulations of initial reactions in zeolite synthesis," *Phys. Chem. Chem. Phys.*, vol. 17, no. 38, pp. 25421–25428, 2015.

[256] M. G. Wu and M. W. Deem, "Monte Carlo study of the nucleation process during zeolite synthesis," *Journal of Chemical Physics*, vol. 116, no. 101, pp. 2125–2137, 2002.

[257] X.-Q. Zhang, T. T. Trinh, R. A. Van Santen, and A. P. J. Jansen, "Structure-Directing Role of Counterions in the Initial Stage of Zeolite Synthesis," *J. Phys. Chem. C*, vol. 115, pp. 9561–9567, 2011.

[258] X.-Q. Zhang, R. A. van SantenâĂą, and A. P. J. Jansen, "Kinetic Monte Carlo modeling of silicate oligomerization and early gelationâĂă," *Physical Chemistry Chemical Physics*, vol. 14, no. 34, p. 11969, 2012.

[259] Y. Tu and J. Tersoff, "Structure and Energetics of the Si-SiO2 Interface," *Physical Review Letters*, vol. 84, no. 19, pp. 4393–4396, 2000.

[260] V. M. Burlakov, G. A. D. Briggs, A. P. Sutton, and Y. Tsukahara, "Monte Carlo Simulation of Growth of Porous SiOx by Vapor Deposition," *Physical Review Letters*, vol. 86, no. 14, pp. 3052–3055, 2001.

[261] C. Schumacher, J. Gonzalez, P. A. Wright, and N. A. Seaton, "Generation of atomistic models of periodic mesoporous silica by kinetic Monte Carlo simulation of the synthesis of the material," *Journal of Physical Chemistry B*, vol. 110, no. 1, pp. 319–333, 2006.

[262] C. A. Ferreiro-Rangel, M. M. Lozinska, P. A. Wright, N. A. Seaton, and T. Düren, "Kinetic Monte Carlo simulation of the synthesis of periodic mesoporous silicas SBA-2 and STAC-1: Generation of realistic atomistic models," *Journal of Physical Chemistry C*, vol. 116, no. 39, pp. 20966–20974, 2012.

[263] Q. Huo, R. Leon, P. M. Petroff, and G. D. Stucky, "Mesostructure Design with Gemini Surfactants: Supercage Formation in a Three-Dimensional Hexagonal Array," *Science*, vol. 268, no. 5215, pp. 1324–1327, 1995.

[264] W. Zhou, H. M. A. Hunter, P. A. Wright, Q. Ge, and J. M. Thomas, "Imaging the Pore Structure and Polytypic Intergrowths in Mesoporous Silica," *J. Phys. Chem. B*, vol. 102, no. 36, pp. 6933–6936, 1998.

[265] B. Ivan and O. Alla, "Multiple Phases of Liquid Water," *ChemPhysChem*, vol. 9, no. 18, pp. 2660–2675, 2008.

[266] J. R. Errington and P. G. Debenedetti, "Relationship between structural order and the anomalies of liquid water," *Nature*, vol. 409, pp. 318–321, jan 2001.

[267] J. Russo and H. Tanaka, "Understanding water's anomalies with locally favoured structures," *Nat. Commun.*, vol. 5, p. 3556, dec 2014.

[268] C. H. Cho, S. Singh, and G. W. Robinson, "Understanding all of water's anomalies with a nonlocal potential," *J. Chem. Phys.*, vol. 107, p. 7979, aug 1998.

[269] L. W. Dahl and H. C. Andersen, "A theory of the anomalous thermodynamic properties of liquid water," *J. Chem. Phys.*, vol. 78, pp. 1980–1993, feb 1983.

[270] B. Guillot, "A reappraisal of what we have learnt during three decades of computer simulations on water," *J. Mol. Liq.*, vol. 101, pp. 219–260, nov 2002.

[271] O. A. V. and I. Saeed, "Water models for biomolecular simulations," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 8, no. 2, p. e1347, 2017.

[272] C. Vega and J. L. F. Abascal, "Simulating water with rigid non-polarizable models: a general perspective," *Phys. Chem. Chem. Phys.*, vol. 13, no. 44, pp. 19663–19688, 2011.

[273] K. R. Hadley and C. McCabe, "Coarse-grained molecular models of water: A review," *Molecular Simulation*, vol. 38, no. 8-9, pp. 671–681, 2012.

[274] R. E. Skyner, J. L. McDonagh, C. R. Groom, T. van Mourik, and J. B. O. Mitchell, "A review of methods for the calculation of solution free energies and the modelling of systems in solution," *Phys. Chem. Chem. Phys.*, vol. 17, pp. 6174–6191, feb 2015.

[275] G. D. R. Matos, D. Y. Kyu, H. H. Loeffler, J. D. Chodera, M. R. Shirts, and D. L. Mobley, "Approaches for calculating solvation free energies and enthalpies demonstrated with an update of the FreeSolv database.," *J. Chem. Eng. Data*, vol. 62, pp. 1559–1569, may 2017.

[276] G. Duarte Ramos Matos and D. L. Mobley, "Challenges in the use of atomistic simulations to predict solubilities of drug-like molecules.," *F1000Research*, vol. 7, p. 686, 2018.

[277] H. W. Horn, W. C. Swope, and J. W. Pitera, "Characterization of the TIP4P-Ew water model: Vapor pressure and boiling point," *Journal of Chemical Physics*, vol. 123, no. 19, pp. 234505–154707, 2005.

[278] K. Laasonen, M. Sprik, M. Parrinello, and R. Car, "Ab initio liquid water," *J. Chem. Phys.*, vol. 99, pp. 9080–9089, dec 1993.

[279] P. Silvestrelli and M. Parrinello, "Water Molecule Dipole in the Gas and in the Liquid Phase," *Physical Review Letters*, vol. 82, no. 16, pp. 3308–3311, 1999.

[280] P. L. Silvestrelli and M. Parrinello, "Structural, electronic, and bonding properties of liquid water from first principles," *Journal of Chemical Physics*, vol. 111, no. 117, pp. 130901–926, 1999.

[281] T. Todorova, A. P. Seitsonen, J. Hutter, I.-F. W. Kuo, and C. J. Mundy, "Molecular Dynamics Simulation of Liquid Water: Hybrid Density Functionals," *J. Phys. Chem. B*, vol. 110, no. 8, pp. 3685–3691, 2005.

[282] M. J. McGrath, J. I. Siepmann, I.-F. W. Kuo, and C. J. Mundy, "Spatial correlation of dipole fluctuations in liquid water," *Molecular Physics*, vol. 105, pp. 1411–1417, may 2007.

[283] C. Zhang, J. Wu, G. Galli, and F. Gygi, "Structural and Vibrational Properties of Liquid Water from van der Waals Density Functionals," *Journal of Chemical Theory and Computation*, vol. 7, pp. 3054–3061, oct 2011.

[284] I.-C. Lin, A. P. Seitsonen, I. Tavernelli, and U. Rothlisberger, "Structure and Dynamics of Liquid Water from ab Initio Molecular DynamicsâĂŤComparison of BLYP, PBE, and revPBE Density Functionals with and without van der Waals Corrections," *Journal of Chemical Theory and Computation*, vol. 8, pp. 3902–3910, oct 2012.

[285] R. Z. Khaliullin and T. D. Kühne, "Microscopic properties of liquid water from combined ab initio molecular dynamics and energy decomposition studies," *Phys. Chem. Chem. Phys.*, vol. 15, p. 15746, sep 2013.

[286] A. Bankura, A. Karmakar, V. Carnevale, A. Chandra, and M. L. Klein, "Structure, Dynamics, and Spectral Diffusion of Water from First-Principles Molecular Dynamics," *The Journal of Physical Chemistry C*, vol. 118, pp. 29401–29411, dec 2014.

[287] M. J. Gillan, D. Alfè, and A. Michaelides, "Perspective: How good is DFT for water?," *The Journal of Chemical Physics*, vol. 144, p. 130901, apr 2016.

[288] H. A. Scheraga, M. Khalili, and A. Liwo, "Protein-Folding Dynamics: Overview of Molecular Simulation Techniques," *Annual Review of Physical Chemistry*, vol. 58, pp. 57–83, may 2007.

[289] Y. M. Rhee, E. J. Sorin, G. Jayachandran, E. Lindahl, and V. S. Pande, "Simulations of the role of water in the protein-folding mechanism.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, pp. 6456–6461, apr 2004.

[290] I. V. Leontyev and A. A. Stuchebrukhov, "Polarizable Mean-Field Model of Water for Biological Simulations with AMBER and CHARMM Force Fields," *Journal of Chemical Theory and Computation*, vol. 8, pp. 3207–3216, sep 2012.

[291] K. Heinzinger, "Computer simulations of aqueous electrolyte solutions," *Phys. B+C*, vol. 131, pp. 196–216, aug 1985.

[292] P. B. Balbuena, K. P. Johnston, and P. J. Rossky, "Molecular Dynamics Simulation of Electrolyte Solutions in Ambient and Supercritical Water. 1. Ion Solvation," *The Journal of Physical Chemistry*, vol. 100, pp. 2706–2715, jan 1996.

[293] S. Chowdhuri and A. Chandra, "Molecular dynamics simulations of aqueous NaCl and KCl solutions: Effects of ion concentration on the single-particle, pair, and collective dynamical properties of ions and water molecules," *J. Chem. Phys.*, vol. 115, pp. 3732–3741, aug 2001.

[294] W. L. Jorgensen and C. Ravimohan, "Monte Carlo simulation of differences in free energies of hydration," *J. Chem. Phys.*, vol. 83, pp. 3050–3054, sep 1985.

[295] D. L. Mobley, C. I. Bayly, M. D. Cooper, and K. A. Dill, "Predictions of Hydration Free Energies from All-Atom Molecular Dynamics Simulations," *The Journal of Physical Chemistry B*, vol. 113, pp. 4533–4537, apr 2009.

[296] M. Misin, M. V. Fedorov, and D. S. Palmer, "Hydration Free Energies of Molecular Ions from Theory and Simulation," *The Journal of Physical Chemistry B*, vol. 120, pp. 975–983, feb 2016.

[297] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, "Interaction Models for Water in Relation to Protein Hydration," in *Intermolecular Forces*, pp. 331–342, 1981.

[298] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *J. Chem. Phys.*, vol. 79, pp. 926–935, jul 1983.

[299] W. L. Jorgensen and J. D. Madura, "Temperature and size dependence for Monte Carlo simulations of TIP4P water," *Molecular Physics*, vol. 56, pp. 1381–1392, dec 1985.

[300] M. W. Mahoney and W. L. Jorgensen, "A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions," *J. Chem. Phys.*, vol. 112, p. 8910, may 2000.

[301] C. Vega, J. L. F. Abascal, E. Sanz, L. G. MacDowell, and C. McBride, "Can simple models describe the phase diagram of water?," *J. Phys. Condens. Matter*, vol. 17, pp. S3283—S3288, nov 2005.

[302] S. Izadi, R. Anandakrishnan, and A. V. Onufriev, "Building Water Models: A Different Approach," *Physical Chemistry Letters*, vol. 5, pp. 3863–3871, 2014.

[303] F. Figueirido, G. S. Del Buono, and R. M. Levy, "On finite size effects in computer simulations using the Ewald potential," *J. Chem. Phys.*, vol. 103, pp. 6133–6142, oct 1995.

[304] P. E. Smith and B. M. Pettitt, "Ewald artifacts in liquid state molecular dynamics simulations," *J. Chem. Phys.*, vol. 105, p. 4289, aug 1998.

[305] J. S. Hub, B. L. de Groot, H. Grubmüller, and G. Groenhof, "Quantifying Artifacts in Ewald Simulations of Inhomogeneous Systems with a Net Charge," *Journal of Chemical Theory and Computation*, vol. 10, pp. 381–390, jan 2014.

[306] H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura, and T. Head-Gordon, "Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew," *J. Chem. Phys.*, vol. 120, pp. 9665–9678, may 2004.

[307] J. L. F. Abascal and C. Vega, "A general purpose model for the condensed phases of water: TIP4P/2005," *J. Chem. Phys.*, vol. 123, p. 234505, dec 2005.

[308] C. J. Fennell, L. Li, and K. A. Dill, "Simple Liquid Models with Corrected Dielectric Constants," *J. Phys. Chem. B*, vol. 116, pp. 6936–6944, jun 2012.

[309] R. Fuentes-Azcatl and J. Alejandre, "Non-Polarizable Force Field of Water Based on the Dielectric Constant: TIP4P/ϵ," *The Journal of Physical Chemistry B*, vol. 118, pp. 1263–1272, feb 2014.

[310] L.-P. Wang, T. J. Martinez, and V. S. Pande, "Building Force Fields: An Automatic, Systematic, and Reproducible Approach," *The Journal of Physical Chemistry Letters*, vol. 5, pp. 1885–1891, jun 2014.

[311] J. L. F. Abascal, E. Sanz, R. García Fernández, and C. Vega, "A potential model for the study of ices and amorphous water: TIP4P/Ice," *J. Chem. Phys.*, vol. 122, p. 234511, jun 2005.

[312] D. D. Kemp and M. S. Gordon, "An Interpretation of the Enhancement of the Water Dipole Moment Due to the Presence of Other Water Molecules," *The Journal of Physical Chemistry A*, vol. 112, pp. 4885–4894, jun 2008.

[313] P. Jedlovszky and J. Richardi, "Comparison of different water models from ambient to supercritical conditions: A Monte Carlo simulation and molecular OrnsteinâĂŞZernike study," *The Journal of Chemical Physics*, vol. 110, pp. 8019–8031, apr 1999.

[314] G. Lamoureux, A. D. MacKerell, and B. Roux, "A simple polarizable model of water based on classical Drude oscillators," *The Journal of Chemical Physics*, vol. 119, pp. 5185–5197, sep 2003.

[315] P. Paricaud, M. PÅŹedota, A. A. Chialvo, and P. T. Cummings, "From dimer to condensed phases at extreme conditions: Accurate predictions of the properties of water by a Gaussian charge polarizable model," *The Journal of Chemical Physics*, vol. 122, p. 244511, jun 2005.

[316] L.-P. Wang, T. Head-Gordon, J. W. Ponder, P. Ren, J. D. Chodera, P. K. Eastman, T. J. Martinez, and V. S. Pande, "Systematic Improvement of a Classical Molecular Model of Water," *J. Phys. Chem. B*, vol. 117, no. 34, pp. 9956–9972, 2013.

[317] R. Qi, L.-P. Wang, Q. Wang, V. S. Pande, and P. Ren, "United polarizable multipole water model for molecular mechanics simulation," *Journal of Chemical Physics*, vol. 143, no. 143, pp. 14504–234505, 2015.

[318] V. P. Sokhan, A. P. Jones, F. S. Cipcigan, J. Crain, and G. J. Martyna, "Signature properties of water: Their molecular electronic origins.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, pp. 6341–6, may 2015.

[319] C. Chipot, "Rational determination of charge distributions for free energy calculations," *J. Comput. Chem.*, vol. 24, pp. 409–415, mar 2003.

[320] D. S. Cerutti, J. E. Rice, W. C. Swope, and D. A. Case, "Derivation of Fixed Partial Charges for Amino Acids Accommodating a Specific Water Model and Implicit Polarization," *The Journal of Physical Chemistry B*, vol. 117, pp. 2328–2338, feb 2013.

[321] P. Zhang, P. Bao, and J. Gao, "Dipole Preserving and Polarization Consistent Charges," *Journal of Chemical Theory and Computation*, vol. 32, pp. 2127–2139, 2011.

[322] W. C. Swope, H. W. Horn, and J. E. Rice, "Accounting for polarization cost when using fixed charge force fields. I. Method for computing energy," *Journal of Physical Chemistry B*, vol. 114, no. 26, pp. 8621–8630, 2010.

[323] W. C. Swope, H. W. Horn, and J. E. Rice, "Accounting for Polarization Cost When Using Fixed Charge Force Fields. II. Method and Application for Computing Effect of Polarization Cost on Free Energy of Hydration," *The Journal of Physical Chemistry B*, vol. 114, pp. 8631–8645, jul 2010.

[324] I. Leontyev and A. Stuchebrukhov, "Accounting for electronic polarization in non-polarizable force fields," *Physical Chemistry Chemical Physics*, vol. 13, no. 7, p. 2613, 2011.

[325] P. F. B. Gonçalves and H. Stassen, "Calculation of the free energy of solvation from molecular dynamics simulations," *Pure and Applied Chemistry*, vol. 76, pp. 231–240, jan 2004.

[326] S. S. Park, S. Lee, J. Y. Bae, and F. Hagelberg, "Refractive indices of liquid-forming organic compounds by density functional theory," *Chemical Physics Letters*, vol. 511, pp. 466–470, aug 2011.

[327] M. Jorge, N. M. Garrido, C. J. V. Simões, C. G. Silva, and R. M. M. Brito, "Predicting hydrophobic solvation by molecular simulation: 1. Testing united-atom alkane models," *Journal of Computational Chemistry*, vol. 38, pp. 346–358, mar 2017.

[328] M. Jorge, "Predicting hydrophobic solvation by molecular simulation: 2. New united-atom model for alkanes, alkenes, and alkynes," *Journal of Computational Chemistry*, vol. 38, pp. 359–369, mar 2017.

[329] E. Lindahl, B. Hess, and D. van der Spoel, "GROMACS 3.0: a package for molecular simulation and trajectory analysis," *J. Mol. Model.*, vol. 7, pp. 306–317, 2001.

[330] W. F. Van Gunsteren and H. J. C. Berendsen, "A Leap-frog Algorithm for Stochastic Dynamics," *Molecular Simulation*, vol. 1, pp. 173–185, mar 1988.

[331] W. van Gunsteren and H. Berendsen, "Algorithms for brownian dynamics," *Molecular Physics*, vol. 45, pp. 637–647, feb 1982.

[332] S. Miyamoto and P. A. Kollman, "Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models," *Journal of Computational Chemistry*, vol. 13, pp. 952–962, oct 1992.

[333] M. Jorge, N. M. Garrido, A. J. Queimada, I. G. Economou, and E. A. Macedo, "Effect of the Integration Method on the Accuracy and Computational Efficiency of Free Energy Calculations Using Thermodynamic Integration," *Journal of Chemical Theory and Computation*, vol. 6, pp. 1018–1027, apr 2010.

[334] C. Shyu and F. M. Ytreberg, "Reducing the bias and uncertainty of free energy estimates by using regression to fit thermodynamic integration data," *Journal of Computational Chemistry*, vol. 30, pp. NA–NA, nov 2009.

[335] M. Born, "Volumen und Hydratationswärme der Ionen," *Zeitschrift für Physik*, vol. 1, pp. 45–48, feb 1920.

[336] J. G. Kirkwood, "Theory of Solutions of Molecules Containing Widely Separated Charges with Special Application to Zwitterions," *The Journal of Chemical Physics*, vol. 2, pp. 351–361, jul 1934.

[337] J. G. Kirkwood, "The Dielectric Polarization of Polar Liquids," *The Journal of Chemical Physics*, vol. 7, pp. 911–919, oct 1939.

[338] K. V. M. Y. Luo, H. Agren, "Unique Determination of the Cavity in Onsager Reaction Field Theory," *Chemical Physics Letters*, vol. 275, pp. 145–150, 1997.

[339] L. Onsager, "Electric Moments of Molecules in Liquids," *Journal of the American Chemical Society*, vol. 58, pp. 1486–1493, aug 1936.

[340] J. Barker and R. Watts, "Monte Carlo studies of the dielectric properties of water-like models," *Molecular Physics*, vol. 26, pp. 789–792, sep 1973.

[341] S. Niu, M.-L. Tan, and T. Ichiye, "The large quadrupole of water molecules," *J. Phys. Chem.*, vol. 134, p. 134501, 2011.

[342] L. D. Site, A. Alavi, and R. M. Lynden-Bell, "The electrostatic properties of water molecules in condensed phases: an ab initio study," *Molecular Physics*, vol. 96, pp. 1683–1693, jun 1999.

[343] K. Coutinho, R. C. Guedes, B. J. Costa Cabral, and S. Canuto, "Electronic polarization of liquid water: converged Monte Carlo-quantum mechanics results for the multipole moments," *Chemical Physics Letters*, vol. 369, pp. 345–353, feb 2003.

[344] A. V. Gubskaya and P. G. Kusalik, "The total molecular dipole moment for liquid water," *The Journal of Chemical Physics*, vol. 117, pp. 5290–5302, sep 2002.

[345] H. Liu, Y. Wang, and J. M. Bowman, "Transferable ab Initio Dipole Moment for Water: Three Applications to Bulk Water," *The Journal of Physical Chemistry B*, vol. 120, pp. 1735–1742, mar 2016.

[346] D. M. Elking, L. Perera, R. Duke, T. Darden, and L. G. Pedersen, "A finite field method for calculating molecular polarizability tensors for arbitrary multipole rank," *Journal of Computational Chemistry*, vol. 32, pp. 3283–3295, nov 2011.

[347] T. Pluta, J. Noga, and R. J. Bartlett, "Determination of higher electric polarizability tensors from unrelaxed coupled cluster density matrix calculations of electric multipole moments," *International Journal of Quantum Chemistry*, vol. 52, pp. 379–393, feb 1994.

[348] O. Loboda, F. Ingrosso, M. F. Ruiz-López, H. Reis, and C. Millot, "Dipole and quadrupole polarizabilities of the water molecule as a function of geometry," *Journal of Computational Chemistry*, vol. 37, pp. 2125–2132, sep 2016.

[349] J.-F. Truchon, A. Nicholls, R. I. Iftimie, B. Roux, and C. I. Bayly, "Accurate Molecular Polarizabilities Based on Continuum Electrostatics," *Journal of Chemical Theory and Computation*, vol. 4, pp. 1480–1493, sep 2008.

[350] J. Tomasi, B. Mennucci, and R. Cammi, "Quantum Mechanical Continuum Solvation Models," *Chemical Reviews*, vol. 105, pp. 2999–3094, aug 2005.

[351] Ganesh Kamath, G. Georgiev, and Jeffrey J. Potoff, "Molecular Modeling of Phase Behavior and Microstructure of Acetone-Chloroform-Methanol Binary Mixtures," *J. Phys. Chem. B*, vol. 109, no. 41, pp. 19463–19473, 2005.

[352] T. J. Zielinski, E. Harvey, R. Sweeney, and D. M. Hanson, "Quantum States of Atoms and Molecules," *Journal of Chemical Education*, vol. 82, p. 1880, dec 2005.

[353] P. Popelier and P. Bultinck, "Atoms in Molecules and Population Analysis," in *Chemical Reactivity Theory*, CRC Press, feb 2009.

[354] F. L. Hirshfeld, "Bonded-atom fragments for describing molecular charge densities," *Theoretica Chimica Acta*, vol. 44, no. 2, pp. 129–138, 1977.

[355] E. R. Davidson and S. Chakravorty, "A test of the Hirshfeld definition of atomic charges and moments," *Theoretica Chimica Acta*, vol. 83, no. 5-6, pp. 319–330, 1992.

[356] T. A. Manz and N. G. Limas, "Introducing DDEC6 atomic population analysis: part 1. Charge partitioning theory and methodology," *RSC Advances*, vol. 6, pp. 47771–47801, may 2016.

[357] F. A. Momany, "Determination of partial atomic charges from ab initio molecular electrostatic potentials. Application to formamide, methanol, and formic acid," *Journal of Physical Chemistry*, vol. 82, pp. 592–601, mar 1978.

[358] P. Matczak, Matczak, and Piotr, "A Test of Various Partial Atomic Charge Models for Computations on Diheteroaryl Ketones and Thioketones," *Computation*, vol. 4, p. 3, jan 2016.

[359] U. C. Singh and P. A. Kollman, "An approach to computing electrostatic charges for molecules," *Journal of Computational Chemistry*, vol. 5, pp. 129–145, apr 1984.

[360] C. M. Breneman and K. B. Wiberg, "Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis," *Journal of Computational Chemistry*, vol. 11, pp. 361–373, apr 1990.

[361] C. I. Bayly, P. Cieplak, W. Cornell, and P. A. Kollman, "A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model," *The Journal of Physical Chemistry*, vol. 97, pp. 10269–10280, oct 1993.

[362] C. Campañá, B. Mussard, and T. K. Woo, "Electrostatic Potential Derived Atomic Charges for Periodic Systems Using a Modified Error Functional," *Journal of Chemical Theory and Computation*, vol. 5, pp. 2866–2878, oct 2009.

[363] T. A. Manz and D. S. Sholl, "Chemically meaningful atomic charges that reproduce the electrostatic potential in periodic and nonperiodic materials," *Journal of Chemical Theory and Computation*, vol. 6, no. 8, pp. 2455–2468, 2010.

[364] O. Andreussi, I. Dabo, and N. Marzari, "Revised self-consistent continuum solvation in electronic-structure calculations," *Journal of Chemical Physics*, vol. 136, p. 064102, feb 2012.

[365] S. Miertuš, E. Scrocco, and J. Tomasi, "Electrostatic interaction of a solute with a continuum. A direct utilizaion of AB initio molecular potentials for the prevision of solvent effects," *Chemical Physics*, vol. 55, pp. 117–129, feb 1981.

[366] A. Klamt and G. Schüürmann, "COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient," *J. Chem. Soc., Perkin Trans. 2*, vol. 0, pp. 799–805, jan 1993.

[367] E. Cancès, B. Mennucci, and J. Tomasi, "A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anisotropic dielectrics," *The Journal of Chemical Physics*, vol. 107, pp. 3032–3041, aug 1997.

[368] E. Cancès and B. Mennucci, "New applications of integral equations methods for solvation continuum models: ionic solutions and liquid crystals," *Journal of Mathematical Chemistry*, vol. 23, no. 3/4, pp. 309–326, 1998.

[369] A. V. Marenich, C. J. Cramer, and D. G. Truhlar, "Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions," *The Journal of Physical Chemistry B*, vol. 113, pp. 6378–6396, may 2009.

[370] C. J. C. And and D. G. Truhlar*, "Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics," *Chem. Rev.*, vol. 99, no. 8, pp. 2161–2200, 1999.

[371] J.-L. Fattebert and F. Gygi, "Density functional theory for efficientab initio molecular dynamics simulations in solution," *Journal of Computational Chemistry*, vol. 23, pp. 662–666, apr 2002.

[372] C. Dupont, O. Andreussi, and N. Marzari, "Self-consistent continuum solvation (SCCS): The case of charged systems," *The Journal of Chemical Physics*, vol. 139, p. 214110, dec 2013.

[373] "The CP2K Developers Group."

[374] B. Chen, J. J. Potoff, and J. I. Siepmann, "Monte Carlo calculations for alcohols and their mixtures with alkanes. Transferable potentials for phase equilibria. 5. United-atom description of primary, secondary, and tertiary alcohols," *Journal of Physical Chemistry B*, vol. 105, no. 15, pp. 3093–3104, 2002.

[375] J. Pérez-Pellitero, E. Bourasseau, I. Demachy, J. Ridard, P. Ungerer, and A. D. Mackie, "Anisotropic United-Atoms (AUA) potential for alcohols," *Journal of Physical Chemistry B*, vol. 112, no. 32, pp. 9853–9863, 2008.

[376] N. Sieffert, M. Bühl, M. P. Gaigeot, and C. A. Morrison, "Liquid methanol from DFT and DFT/MM molecular dynamics simulations," *Journal of Chemical Theory and Computation*, vol. 9, no. 1, pp. 106–118, 2013.

[377] B. Chen and J. I. Siepmann, "Transferable Potentials for Phase Equilibria. 3. Explicit-Hydrogen Description of Normal Alkanes," *The Journal of Physical Chemistry B*, vol. 103, no. 25, pp. 5370–5379, 1999.

[378] E. R. Cruz-Chu, A. Aksimentiev, and K. Schulten, "WaterâĹŠSilica Force Field for Simulating Nanodevices," *J. Phys. Chem. B*, vol. 110, no. 43, pp. 21497–21508, 2006.

[379] H. Heinz, T.-J. Lin, R. Kishore Mishra, and F. S. Emami, "Thermodynamically Consistent Force Fields for the Assembly of Inorganic, Organic, and Biological Nanostructures: The INTERFACE Force Field," *Langmuir*, vol. 29, pp. 1754–1765, feb 2013.

[380] E. G. Rochow, *The Chemistry of Silicon.* Pergamon International Library of Science, Technology, Engineering and Social Studies, 1973.

[381] C. Yokoyama, T. Takagi, and S. Takahashi, "Densities of tetramethylsilane, tetraethylsilane, and tetraethoxysilane under high pressures," *International Journal of Thermophysics*, vol. 11, pp. 477–486, may 1990.

[382] M. Voronkov, V. Baryshok, V. Klyuchnikov, T. Danilova, V. Pepekin, A. Korchagina, and Y. Khudobin, "Thermochemistry of organosilicon compounds: I. Triorganyl-, tetraorganyl-, organylorganoxy- and tetraorganoxy-silanes," *Journal of Organometallic Chemistry*, vol. 345, pp. 27–38, may 1988.

[383] J. Gaube, *The vapour pressures of pure substances-Selected values of the temperature dependence of the vapour pressures of some pure substances in the normal and low pressure region*, vol. 89. John Wiley & Sons, Ltd, mar 1985.

[384] D. R. Stull, "Vapor Pressure of Pure Substances. Organic and Inorganic Compounds," *Industrial & Engineering Chemistry*, vol. 39, pp. 517–540, apr 1947.

[385] W. T. Grubb and R. C. Osthoff, "Physical Properties of Organosilicon Compounds. II. Trimethylsilanol and Triethylsilanol," *Journal of the American Chemical Society*, vol. 75, pp. 2230–2232, may 1953.

[386] O. L. Flaningam, "Vapor pressures of poly(dimethylsiloxane) oligomers," *Journal of Chemical & Engineering Data*, vol. 31, pp. 266–272, jul 1986.

[387] Centers for Disease Control and Prevention (CDC), "NIOSH Pocket Guide to Chemical Hazards - Methyl silicate."

[388] M. van der Vis and E. Cordfunke, "Tetraethoxysilane, Si(OC2H5)4: vapour pressure measurements at temperatures from 323 to 442 K by means of a Bourdon spoon gauge," *Thermochimica Acta*, vol. 265, pp. 129–134, nov 1995.

[389] A. V. Marenich, C. P. Kelly, J. D. Thompson, G. D. Hawkins, C. C. Chambers, D. J. Giesen, P. Winget, C. J. Cramer, and D. G. Truhlar, "Minnesota Solvation Database," tech. rep., University of Minnesota, 2012.

[390] J. D. Thompson, C. J. Cramer, and D. G. Truhlar, "Predicting aqueous solubilities from aqueous free energies of solvation and experimental or calculated

vapor pressures of pure substances," *The Journal of Chemical Physics*, vol. 119, pp. 1661–1670, jul 2003.

[391] W. L. Jorgensen and J. Tiradorives, "The OPLS potential functions for proteins: Crystals of cyclic-peptides and Crambin," *Journal of the American Chemical Society*, vol. 110, no. 6, pp. 1657–1666, 1988.

[392] A. D. Mackerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-Mccarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus, "All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins," *J. Phys. Chem. B*, vol. 102, pp. 3586–3616, 1998.

[393] M. G. Martin and J. I. Siepmann, "Transferable Potentials for Phase Equilibria. 1. United-Atom Description of n -Alkanes," *The Journal of Physical Chemistry B*, vol. 102, no. 97, pp. 2569–2577, 1998.

[394] M. G. Martin and J. I. Siepmann, "Novel Configurational-Bias Monte Carlo Method for Branched Molecules. Transferable Potentials for Phase Equilibria. 2. United-Atom Description of Branched Alkanes," *The Journal of Physical Chemistry B*, vol. 103, no. 21, pp. 4508–4517, 1999.

[395] S. Grigoras and T. H. Lane, "Molecular mechanics parameters for organosilicon compounds calculated from ab initio computations," *Journal of Computational Chemistry*, vol. 9, pp. 25–39, jan 1988.

[396] R. J. Abraham and G. H. Grant, "A Molecular Mechanics Study of the Si-O bond and Alkyl-silanes," *Journal of Computational Chemistry*, vol. 9, pp. 709–718, oct 1988.

[397] B. Beagley, J. Monaghan, and T. Hewitt, "Electron-diffraction studies of tetramethylsilane and hexamethyldisilane, and discussion of the lengths of Si-C bonds," *Journal of Molecular Structure*, vol. 8, pp. 401–411, jun 1971.

[398] J. Koput, "The equilibrium structure and torsional potential energy function of methanol and silanol," *Journal of Physical Chemistry A*, vol. 104, no. 44, pp. 10017–10022, 2000.

[399] A. Almenningen, O. Bastiansen, V. Ewing, K. Hedberg, and M. Traetteberg, "The Molecular Structure of Disiloxane," *Acta Chem. Scand.*, vol. 17, pp. 2455–2460, 1963.

[400] J. Koput and A. Wierzbicki, "The large-amplitude motions in quasi-symmetric top molecules with internal C3v rotors: Interpretation of the low frequency Raman spectrum of disiloxane," *Journal of Molecular Spectroscopy*, vol. 99, pp. 116–132, may 1983.

[401] R. J. Gillespie and S. A. Johnson, "Study of Bond Angles and Bond Lengths in Disiloxane and Related Molecules in Terms of the Topology of the Electron Density and Its Laplacian," *Inorg. Chem.*, vol. 36, no. 14, pp. 3031–3039, 1997.

[402] F. Weinhold and R. West, "The Nature of the Silicon-Oxygen Bond," *Organometallics*, vol. 30, pp. 5815–5824, nov 2011.

[403] ACD/Labs, "PhysChem, ADME/Tox Calculations: Percepta Software."