Development of Polarizable Force Fields and Hybrid QM/MM Methods for the Study of Reaction Mechanisms

Benjamin M. Webb

A thesis submitted for the degree of

Doctor of Philosophy in the

University of Oxford

Wolfson College

Trinity Term 2003

Development of Polarizable Force Fields and Hybrid QM/MM Methods for the Study of Reaction Mechanisms

Abstract of a thesis submitted for the degree of Doctor of Philosophy in the University of Oxford

Benjamin M. Webb Wolfson College

Trinity Term 2003

Computational chemists have successfully simulated many systems by applying the principles of quantum mechanics, while approximate molecular mechanical models have seen great utility in problems of biochemical interest. In recent years, a number of methods have been developed to combine the advantages of both techniques. In this study the so-called QM/MM method is developed and applied to the determination of the free energy of a simple Menshutkin S_N2 chemical reaction. This is an extremely demanding process, well beyond the computational capacity of an average workstation, and thus a Beowulf-class Linux cluster is constructed to perform the calculations, and tested for a variety of computational chemistry applications.

A number of methods for improving the QM/MM approach are considered in this work. The Fluctuating Charge, or FlucQ, polarizable molecular mechanics force field is implemented in a flexible manner within the CHARMM package and tested for a variety of systems, including the S_N2 test case. Several drawbacks of the original method are addressed and overcome.

Both molecular dynamics and Monte Carlo techniques are used within the QM/MM framework to investigate the S_N2 reaction, and the two methods are compared. Techniques are developed and tested to increase the efficiency of QM/MC calculations to the point where they become competitive with QM/MD.

Extremely expensive QM treatments are shown to be required to obtain accurate energies for the Menshutkin reaction. A method is developed and tested, and compared with the traditional ONIOM technique, for dramatically reducing the computational time required to use these treatments for QM/MC simulations, paving the way for fully *ab initio* high basis set QM/MM simulation.

Acknowledgements

This thesis would not have been possible without the tireless proofreading efforts of my colleagues Maya, Owen and Stewart, when they could far more easily have been relaxing in the Californian (or Nottinghamshire) sun. Cheers! Who would know how much better southern California is than rainy England without Stewart's handy reminders? And I have Owen to thank for teaching me the dangers of nonalcoholic drinks. Maya, of course, deserves thanks for her always calming influence in the lab, and for keeping the hard disks nicely topped up. And for not killing anybody.

Thanks also to Graham, for supervision, funding, and (of course) those all-important trips to the KA. And to Paul, for always demanding the impossible (and for then leaving me to figure out the answer).

Life in the lab wouldn't be the same without the local friendly postdocs. These included Meir, always keeping up the good work on the warp drive, and Aaron, who was exceptionally diligent and single-minded in his task of making sure that I woke up early every morning. And, of course, who could forget the Amazing Singing Postdoc?

Thanks to the every-chirpy Mervyn for making those days at the races a little more bearable, to Biggi and Mike for always foiling my evil plans to get the computers working so that I could go home early (I blame the bananas), and to the inimitable PV (but, *why?*).

Thanks to Yalini for ensuring that no day in Stevenage could ever be described as 'dull', and to Caroline for seeing to it that my head can still fit through the smallest of doors. To friends and family in sunny Suffolk, for their eternal patience in the face of an apparently never-ending thesis.

Many thanks to the authors of the MOLDEN [1] and VMD [2] packages, which were enormously useful for visualisation and presentation of the results of this work. Thanks also to the Oxford Supercomputer Centre for CPU time [3] and to Andrew Leach and GlaxoSmithKline for financial support.

Finally, thanks to all the fun people I've met during my time here, who've made Oxford a laugh a minute — or, at least, a chuckle a week.

Contents

Introduction **1**

1 Theoretical background **5**

1.1 Overview . . . 5

1.2 Quantum mechanical treatments . . . 5

1.2.1 Semi-empirical methods . . . 16

1.2.2 Density Functional Theory . . . 20

1.3 Molecular mechanics force fields . . . 24

1.4 Hybrid QM/MM . . . 27

1.5 Polarization . . . 31

1.6 Optimization of pair potentials . . . 31

1.7 Boundary treatments . . . 32

1.7.1 Periodic boundary conditions . . . 33

1.7.2 Ewald summation . . . 34

1.7.3 Stochastic boundary techniques . . . 36

1.8 Charge determination . . . 37

1.9 Determination of minimum energy system configurations . . . 39

1.9.1 Zero-order methods . . . 39

1.9.2 First-order methods . . . 40

1.9.3 Second-order methods . . . 42

1.10 Determination of free energy . . . 43

1.10.1 Molecular dynamics . . . 43

1.10.2 Temperature control . . . 46

1.10.3 Ensemble averages . . . 48
1.10.4 Monte Carlo simulation . . . 49
1.10.5 Solvent treatments . . . 52
1.10.6 Reaction free energies via umbrella sampling . . . 54
1.11 Summary . . . 57

2 Computational infrastructure **59**
2.1 Overview . . . 59
2.2 Parallelisation of computational problems . . . 60
2.3 Types of parallelism . . . 62
2.3.1 Vector processors . . . 62
2.3.2 Shared memory systems . . . 63
2.3.3 Distributed memory systems . . . 64
2.4 Parallel programming . . . 65
2.5 Design and implementation of a Beowulf-class Linux cluster . . . 66
2.5.1 Local versus network storage . . . 69
2.5.2 Factors affecting network performance . . . 70
2.5.3 Hardware setup . . . 72
2.5.4 Software setup . . . 72
2.6 Tests of the cluster performance . . . 75
2.6.1 Base network performance . . . 75
2.6.2 Real-world application test cases . . . 78
2.7 Summary . . . 81

3 Implementation of fluctuating charge force fields **83**
3.1 Introduction . . . 83
3.1.1 Dipole polarizable methods . . . 84
3.1.2 Electronegativity equalisation approaches . . . 85
3.2 Determination of polarization energy . . . 86
3.3 Molecular dynamics via extended Lagrangian methods . . . 88
3.4 Charge minimization . . . 92

3.5 Thermostatting . . . 94
3.6 Monte Carlo implementation . . . 95
3.7 Incorporation into QM/MM models . . . 96
3.8 Ewald summation . . . 97
3.9 Intramolecular interaction . . . 98
3.9.1 Flexible water models . . . 99
3.10 FlucQ energy terms . . . 102
3.11 Code implementation in CHARMM . . . 102
3.11.1 Energy and force determination . . . 103
3.11.2 Charge dynamics . . . 104
3.11.3 General modifications . . . 104
3.12 Summary . . . 105

4 Fluctuating charge test cases **106**
4.1 Overview . . . 106
4.2 Bulk solvent test cases . . . 106
4.2.1 Use of cutoff approximations . . . 107
4.2.2 Thermostat test cases . . . 110
4.2.3 Simulation using Ewald summation . . . 115
4.2.4 Monte Carlo simulation . . . 116
4.3 The water dimer . . . 122
4.3.1 Dipole determination . . . 123
4.3.2 Minimum-energy properties . . . 124
4.3.3 Determination of the water-water binding energy profile . . . 127
4.3.4 Energy decomposition . . . 129
4.3.5 Charge transfer . . . 129
4.3.6 QM/MM studies . . . 130
4.4 Other bimolecular systems . . . 132
4.4.1 Water/chloride . . . 133
4.4.2 Water/fluoride . . . 135
4.4.3 Water/sodium . . . 137

4.5 Solvated QM/MM water 138
4.6 Parameterization of the FlucQ model 143
4.7 Summary 146

5 Application to the Menshutkin reaction 147
5.1 Introduction 147
5.1.1 S_N2 reactions 147
5.1.2 Type I S_N2 reactions 148
5.1.3 Type II S_N2: The Menshutkin reaction 149
5.2 Adiabatic mapping of the reaction surface 151
5.2.1 The adiabatic mapping method 152
5.2.2 Choice of reaction coordinate 152
5.2.3 Comparison of methods in the gas phase 153
5.2.4 Solution-phase energies from continuum treatments 155
5.3 Hybrid QM/MD studies 157
5.3.1 Overview 157
5.3.2 Parameterization of the QM/MM interaction 158
5.3.3 Preparation of the system 165
5.3.4 Umbrella sampling 168
5.3.5 Discussion and analysis 171
5.4 Summary 176

6 Use of hybrid QM/MC methods 177
6.1 Overview 177
6.2 Preferential sampling techniques 179
6.3 Perturbative QM/MC 181
6.3.1 Quantification of error 183
6.3.2 Error control in the case of a changing QM region 184
6.4 Simulation details 185
6.4.1 Move set 186
6.5 Perturbative QM/MC test cases 188

6.6 Preferential sampling test cases . . . 191
6.7 A closer examination of the error added by perturbative QM/MC . . . 192
6.8 Umbrella sampling . . . 196
6.9 Improvement of the energy function . . . 199
6.9.1 The ONIOM method . . . 200
6.9.2 Consideration of the QM/MM interaction energy . . . 201
6.9.3 Approximate correction of the total QM/MM energy . . . 204
6.9.4 Energy test cases . . . 206
6.9.5 Minimization test cases . . . 209
6.9.6 Caveats . . . 212
6.10 Summary . . . 214

7 Conclusions and future research directions **215**
7.1 Conclusions . . . 215
7.2 Future research directions . . . 219

Bibliography **223**

N.B. A CD containing the bulk of the code written for this work, as well as documentation and test cases, is included on the inside back cover.

Introduction

In 1965, only a few short years after the first planar integrated circuit had been designed, Gordon Moore made the observation that "the complexity for minimum component costs has increased at a rate of roughly a factor of two per year" [4] and predicted that this trend could continue for perhaps another decade. Nearly forty years on, this statement is known in very simplified form as "computers double in speed every 12 months" and is dubbed 'Moore's Law'. Remarkably, this 'law' still holds, and the exponential increase in calculation power has revolutionised a large number of fields. Science has of course been strongly affected, and the young field of computational chemistry is driven in large part by this continual increase in computational power.

The principles of quantum mechanics [5] have been known since the early years of the twentieth century, predating the onset of electronic computers, and provide an exact treatment of chemical systems at the level of nuclei and electrons. However, the equations are not soluble for systems of more than two particles, which restricted their application before computing machines provided the means to numerically evaluate the large number of integrals required to obtain approximate solutions of the equations for larger systems, by use of the Hartree-Fock method [6, 7]. The increase in computer power has enabled the use of increasingly better approximations to *ab initio* properties, such as configuration interaction [5], perturbation [8] and Density Functional [9] techniques. At the same time, it has allowed the study of bigger systems and for longer timescales. Still larger systems, such as entire proteins, have been successfully simulated by neglecting the electronic degrees of freedom and approximating the energy

of the system by means of a 'molecular mechanics' (MM) force field parameterized from experimental or *ab initio* data [10, 11].

Computer power can be further increased by linking a collection of computers together to work on a problem, or collection of problems, in concert; the most powerful supercomputers currently available are built in this way. However, many methods do not easily lend themselves to being split over several computers. The more advanced techniques tend to scale poorly with the number of atoms being simulated, and many problems of biochemical interest involve very large numbers of atoms (such as lipid bilayer and ion channel simulations, protein-protein interactions, protein folding studies, and the behaviour of protein assemblies) while very large numbers of configurations need to be considered for free energy determination, vital in the simulation of chemical reactions and determination of equilibrium constants. Thus, the more advanced methods cannot currently be applied to the largest systems, and there is still a large demand for approximate methods and so-called 'linear scaling' techniques [12, 13].

Several methods exist for the approximate evaluation of the free energy [14, 15]. For the determination of the system energy itself, a group of approaches that combine molecular mechanical techniques with *ab initio* treatments, known as Hybrid QM/MM methods [16, 17], have shown great utility. These approaches apply QM techniques to a small part of a system, for example the active site of a protein, and use the more approximate MM techniques to model the surrounding environment. QM/MM methods have been greatly developed in the last decade, for example for the calculation of chemical shifts [18], and have been applied to a wide variety of systems [19–21]. However, the techniques are not easy to use, requiring careful setup and calibration for each system, while they are not nearly as well supported as MM or QM techniques by the common computational chemistry simulation programs. This currently restricts the applicability of such methods to rather specialist applications.

One major drawback of QM/MM methods is the difficulty of interfacing between the MM and QM regions; the former's lack of an explicit treatment of the electron density requires that the interface is chosen in a region where the density is not expected to fluctuate. Many solutions have been attempted [22,23], but the choice of where to place the interface is still not a trivial one, and there are no solid rules to guide the researcher. A related problem is that QM/MM simulations are not fully polarizable; although the QM region is able to polarize in response to changes in the MM environment, the reverse is not possible. This lack of polarizability in MM descriptions has been suggested as the cause of observed discrepancies between QM/MM predictions and experiment [24]; furthermore, even when carrying out conventional MM simulations, polarization can make non-negligible contributions to the energy, particularly for highly polar solvents such as water. Thus, there has been significant interest in the development of polarizable molecular mechanics force fields [25–30]; however, most of these treatments are computationally expensive. In this work, a novel method based on the concept of electronegativity equalisation [31, 32], the so-called Fluctuating Charge or FlucQ method [30, 33–35], is evaluated for use both in conventional MM simulation and for the QM/MM determination of free energy.

While FlucQ aims to improve the accuracy of the MM part of a QM/MM system, the QM region can be similarly improved by simply using a higher level of QM theory. However, this can increase the amount of computer time required to perform the calculations by orders of magnitude. Thus, there is significant interest in methods for 'correcting' energies obtained at a low level to a higher level in some approximate way [36]. In this work, a new method is developed and tested, which aims to achieve such a correction.

In Chapter 1 of this study, the various theories and methods used in modern computational chemistry are outlined. As mentioned above, the application of many of these procedures requires computing power beyond that available on the desktop;

thus in Chapter 2 the various solutions for High Performance and High Throughput Computing are explained, and the computational infrastructure put in place and used for this research is described and profiled. Conventional supercomputer solutions are compared with the more recent innovation of using clusters of inexpensive, commodity machines.

Chapter 3 gives an overview of the FlucQ method, and describes the developments made to it during this study. The implementation of the method in a flexible manner within CHARMM [10], a package for MM and QM/MM simulation, is also detailed. The model is then applied to a variety of MM and QM/MM systems in Chapter 4 to evaluate its performance. In Chapter 5 it is further applied to the determination of the free energy profile of the Menshutkin reaction, a simple S_N2 process, and compared with other non-polarizable QM/MM methods used for the same purpose. Finally, the technique of using Monte Carlo simulation methods in conjunction with a QM/MM energy treatment is investigated in Chapter 6. A number of methods for optimizing the efficiency of these techniques are developed and tested.

Chapter 1

Theoretical background

1.1 Overview

Computational chemistry is applied to a wide range of systems; a correspondingly large number of theories and methodologies have been devised to obtain various properties from these systems at differing levels of accuracy. In this chapter, the basic tenets of quantum mechanics, the underpinning of modern physics and chemistry, are outlined and the approaches used to obtain solutions for chemical problems are elucidated. More approximate techniques are also summarised, and some subtleties that arise when simulating macromolecular systems, such as the treatment of surface effects, are tackled. Finally, the determination of energy minima and free energies, vital for the detailed study of chemical systems and for comparison with experimental properties, is also explained.

1.2 Quantum mechanical treatments

Quantum mechanics generally involves solutions to the time-independent and non-relativistic Schrödinger equation [5], which links the energy of the system, E, to its

wavefunction, Ψ, via the Hamiltonian operator $\hat{H}$:

$$\hat{H}\Psi = E\Psi \tag{1.1}$$

This is an *ab initio* approach (literally 'from first principles') and requires no parameterization, requiring only fundamental physical constants such as the mass and charge of the electron. The wavefunction Ψ describes the system completely, and in theory can be used to give exact values for any measurable property or *observable* (such as, in this case, the energy). However, the Schrödinger equation cannot be solved exactly for systems involving more than two particles; such systems are non-integrable. A number of approximations must therefore be made in order to be able to use it to derive energies for almost all systems of interest. The first of these is the *Born-Oppenheimer Approximation* [37]. Nuclei are very much heavier than electrons, and so can be assumed to be stationary on the timescale of motion of the electrons. Thus, it is only necessary to solve the *electronic* Schrödinger equation,

$$\hat{H}_{elec}\Psi_{elec} = E_{elec}\Psi_{elec} \tag{1.2}$$

where, for a system of N electrons and M nuclei, in atomic units,

$$\hat{H}_{elec} = -\sum_{i=1}^{N}\frac{1}{2}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} \tag{1.3}$$

The first term in this expression accounts for the kinetic energy of the electrons (∇^2 being shorthand for $\partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$). The second covers the attractive interaction between each electron i and the nuclear charges Z_A, separated by a distance r_{iA}, and the third accounts for the repulsion between two electrons i and j.

After solving the electronic Schrödinger equation, a contribution is made to the total energy from the repulsions between the fixed nuclei, E_{nuc}:

$$E = E_{elec} + E_{nuc} = E_{elec} + \sum_{A=1}^{M} \sum_{B>A}^{M} \frac{Z_A Z_B}{R_{AB}} \tag{1.4}$$

This approximation is so general that from here on, energies and operators without a subscript should be assumed to be referring to the electronic system.

Hartree-Fock approximation

The electronic wavefunction of an N-electron system, Ψ, can be written as a *Slater determinant* [38, 39] of N spin orbitals, χ:

$$\begin{aligned} \Psi(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n) &= |\chi_i \chi_j \cdots \chi_k\rangle \\ &= (N!)^{-1/2} \begin{vmatrix} \chi_i(\boldsymbol{x}_1) & \chi_j(\boldsymbol{x}_1) & \ldots & \chi_k(\boldsymbol{x}_1) \\ \chi_i(\boldsymbol{x}_2) & \chi_j(\boldsymbol{x}_2) & \ldots & \chi_k(\boldsymbol{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_i(\boldsymbol{x}_N) & \chi_j(\boldsymbol{x}_N) & \ldots & \chi_k(\boldsymbol{x}_N) \end{vmatrix} \end{aligned} \tag{1.5}$$

The spin orbital $\chi_i(\boldsymbol{x}_1)$ is a wavefunction for a single particle, and is the product of a spatial orbital $\psi_i(\boldsymbol{r}_1)$ and a spin function (α or β) such that every spatial orbital can contain two, spin-paired, electrons. The Slater determinant form of the wavefunction ensures that the wavefunction is antisymmetric with respect to exchange of identical electrons, in line with Pauli's principle [40]. The spatial orbitals used are generally *orthonormal*, i.e.

$$\int \mathrm{d}\boldsymbol{r} \psi_i^*(\boldsymbol{r}) \psi_j(\boldsymbol{r}) = \delta_{ij} \quad \text{or} \quad \langle i|j\rangle = \delta_{ij} \tag{1.6}$$

where δ_{ij} is the Kronecker delta,

$$\delta_{ij} = \begin{cases} 0 & : \quad i \neq j \\ 1 & : \quad i = j \end{cases} \tag{1.7}$$

The *variational principle* states that the energy of any trial wavefunction will always be greater than that of the 'true' wavefunction. Thus, the best guess at the true wavefunction will be that which corresponds to the minimum energy. By varying the functional forms of the spin orbitals — subject to the constraint that they remain orthonormal — such that the energy is minimized, the Hartree-Fock [6, 7] equation is derived, which is an eigenvalue equation of the form

$$f\chi_i = \epsilon_i \chi_i \tag{1.8}$$

where the Fock operator, $f(1)$, is an effective one-electron operator, acting in this case on electron 1:

$$f(1) = h(1) + \sum_{j=1}^{N} [J_j(1) - K_j(1)] \tag{1.9}$$

Here $h(1)$ accounts for the kinetic and nuclear potential energy of electron 1:

$$h(1) = -\frac{1}{2}\nabla_1^2 - \sum_{A=1}^{M} \frac{Z_A}{r_{1A}} \tag{1.10}$$

$J_j(1)$ is the *Coulomb operator*, which quantifies the average electrostatic interaction between the electron and all other electrons:

$$J_j(1)\chi_i(1) = \left[\int \chi_j^*(2) \frac{1}{r_{12}} \chi_j(2) \mathrm{d}r_2 \right] \chi_i(1) \tag{1.11}$$

$K_j(1)$ is the *Exchange operator*, which quantifies the purely quantum mechanical exchange interaction between particles of like spin:

$$K_j(1)\chi_i(1) = \left[\int \chi_j^*(2)\frac{1}{r_{12}}\chi_i(2)\mathrm{d}r_2\right]\chi_j(1) \tag{1.12}$$

Basis sets

The exact form of the set of spatial orbitals $\{\psi_i\}$ is not known. Thus, each orbital is defined as a linear combination of known functions, $\{\phi_\mu\}$, termed a *basis set*. The calculation of energy and other properties then becomes a matter of determining the optimum values of the coefficients $C_{\mu i}$:

$$\psi_i = \sum_{\mu=1}^{K} C_{\mu i}\phi_\mu \tag{1.13}$$

It is preferable to represent the individual orbitals that go to make up the wavefunction by Slater-type functions, of the form $A|\boldsymbol{r} - \boldsymbol{R}_A|^n \mathrm{e}^{-\zeta|\boldsymbol{r}-\boldsymbol{R}_A|}$, where A is a normalization constant, $\boldsymbol{R}_A$ the centre of the orbital, and n and ζ function parameters; the exact solution of the Schrödinger equation for the hydrogen atom, for example, fits this form, and it has been shown [5] that molecular orbitals are of the form $\psi_i \sim \mathrm{e}^{-a_i r}$ for large r.

Unfortunately, multicentre integrals of Slater-type functions are very hard or very slow to calculate. If, however, the functions are replaced by Gaussian-type functions, of the form $A\mathrm{e}^{-\alpha|\boldsymbol{r}-\boldsymbol{R}_A|^2}$, then these integrals become straightforward; the product of two Gaussians is simply a third Gaussian, centred on the line between the original two centres, and this reduces a complex four-centre integral to a much simpler two-centre problem. However, Gaussian functions do not fit the behaviour of molecular orbitals in the same way that Slater functions do; they do not have a cusp at $r = 0$ (the gradient of a Gaussian here is in fact zero) and they decay to zero too quickly with increasing r. These problems can be countered somewhat by approximating a

given Slater-type function, centred at $\boldsymbol{R}_A$, as a linear combination of one or more Gaussian-type functions, g_p [41]. The Gaussian functions are centred at $\boldsymbol{R}_P$; this is almost always, however, set to be equal to $\boldsymbol{R}_A$. Thus, a given Gaussian function in the sum is characterised by its exponent $\alpha_{p\mu}$ and its contraction coefficient $d_{p\mu}$, and is also known as a 'primitive'. The resulting linear combination is known as a *contracted Gaussian function*:

$$\phi_\mu^{CGF}(\boldsymbol{r} - \boldsymbol{R}_A) = \sum_{p=1}^{L} d_{p\mu} g_p(\alpha_{p\mu}, \boldsymbol{r} - \boldsymbol{R}_P) \tag{1.14}$$

In principle, this equation is the same as that in equation 1.13, and the coefficients $d_{p\mu}$ could be determined during an energy calculation in the same way that $C_{\mu i}$ are. However, in practice it is more efficient to fix the values of $d_{p\mu}$ beforehand and not allow them to vary. Optimum contraction coefficients are chosen by a least squares fit between the desired Slater function and the combination of Gaussians.

The simplest application of this method is used in a *minimal basis*. The wavefunction is built up from a combination of orbitals that cover the conventional atomic orbitals up to the valence shell; for example, Be is modelled with a single *1s* orbital and four *2sp* orbitals. Each of these orbitals is approximated by a linear combination of Gaussians; for example, the STO-3G basis set [42] uses 3 Gaussian functions for each Slater-type orbital. Such basis sets require the minimum of integral evaluations, and are useful for quick results. They do not generally give accurate molecular properties, however [5].

To describe the system more exactly, the size of the atomic orbitals should be optimized. This could be achieved by adjusting the orbital exponents during the energy calculation, to give optimum properties. However, this is rather time-consuming; the usual approach is to replace the one function per orbital approach used for a minimal basis with two functions. The coefficients of these functions in equation 1.13 are then optimized during the energy calculation. Generally, one of these functions is slightly

more dense than the minimal basis solution, and the other is more diffuse. By varying the two coefficients, the size of the resulting orbital can be adjusted. A basis set employing double functions in this way is known as a *double zeta* basis. (Triple zeta basis sets improve the fit still further by using three functions, and so on.) This approach increases the amount of computer time required tremendously, of course, and usually only the valence orbitals are split in this way, resulting in a so-called 'split-valence' basis set [43]. An example is the 6-31G basis, which approximates each core orbital by a contraction of 6 Gaussian functions, and each valence orbital by two functions, one of which is a contraction of 3 Gaussians, and the other a contraction of a single Gaussian.

Although double zeta basis sets can deal with size fluctuations, they are unable to cope with anisotropy of the system. Such anisotropy can be introduced by polarization of the electron density; although this density is spherically symmetric in an isolated atom, an applied electric field will perturb this. Such a field will be created by the mere presence of the other atoms in a molecular system. This anisotropy is accounted for by the introduction of polarization functions; these represent orbitals of higher angular momentum than the valence orbitals [44–46]. For example, 3 p-type primitive Gaussians are added to hydrogen, and a set of d-type functions are added to heavier elements (in fact, it is easier to add 6 functions, which are linear combinations of the 5 familiar d orbitals and one s-type orbital). A basis set incorporating polarization functions can be written with their types explicitly stated — e.g. 6-31G(d,p) [44] uses d functions on heavy atoms, and p functions on hydrogen and helium — or with a convenient asterisk shorthand, in which a single asterisk is taken to mean 'polarization functions only on heavy atoms' and two as 'polarization functions on all atoms'. The previous basis can thus be written as 6-31G**.

Finally, all the orbitals considered so far have been centred on nuclei. This causes problems when dealing with electron density far from the nuclei — e.g. for lone pairs or

anions — as the functions are rather close to zero in that region. To try to counter this, additional diffuse Gaussian functions of s and p type can be added to the basis [47,48]. The presence of diffuse functions is denoted in a similar way to polarization functions, except that a plus sign is used; for example, the 6-31+G** basis is a 6-31G basis modified by adding polarization functions on all atoms, but diffuse functions only on the heavy atoms.

The Roothaan-Hall equations

The Roothaan-Hall equations, developed independently by Roothaan [49] and Hall [50], recast the Hartree-Fock system of integro-differential linear equations into matrix form, and solve them for a finite basis set and a closed-shell system. This enables solutions for *ab initio* systems to be determined by computer programs. The Roothaan-Hall equations can be written in matrix form as

$$\mathsf{FC} = \mathsf{SCE} \tag{1.15}$$

where F is a K-by-K matrix (where K is the number of basis functions). The individual elements of the Fock matrix are given as

$$F_{\mu\nu} = H_{\mu\nu}^{core} + \sum_{\lambda=1}^{K}\sum_{\sigma=1}^{K} P_{\lambda\sigma}\left[(\mu\nu|\lambda\sigma) - \tfrac{1}{2}(\mu\lambda|\nu\sigma)\right] \tag{1.16}$$

$H_{\mu\nu}^{core}$ is simply the result of applying the $h(1)$ operator defined in equation 1.10 to the basis:

$$H_{\mu\nu}^{core} = \int \phi_{\mu}^{*} h(1) \phi_{\nu} \mathrm{d}r_1 \tag{1.17}$$

C is a square matrix of the coefficients $C_{\mu i}$, and E is a diagonal matrix of the orbital energies. For convenience, the C matrix is rarely used directly; instead, the *density*

matrix, P, is employed, with elements

$$P_{\mu\nu} = 2\sum_{i=1}^{N/2} C_{\mu i}^{*} C_{\nu i} \tag{1.18}$$

The exchange and Coulomb integrals of the Hartree-Fock equations are replaced by the following 2-electron integrals over the basis set:

$$(\mu\nu|\lambda\sigma) = \int \phi_{\mu}^{*}(1)\phi_{\nu}(1)\frac{1}{r_{12}}\phi_{\lambda}^{*}(2)\phi_{\sigma}(2)\mathrm{d}r_1\mathrm{d}r_2 \tag{1.19}$$

S is the overlap matrix; note that the basis functions, unlike the orbitals, are usually normalised but are not necessarily orthogonal. S has elements

$$S_{\mu\nu} = \int \phi_{\mu}^{*}(1)\phi_{\nu}(1)\mathrm{d}r_1 \tag{1.20}$$

The Fock operator itself is dependent on the spin orbitals χ_i, and thus the Fock matrix is dependent on the coefficient matrix C. Thus, the Roothaan-Hall equations cannot be solved exactly; an iterative procedure must be employed instead. The procedure employed is known as *Self Consistent Field* or SCF. Physically, this corresponds to obtaining a field which produces a particular charge density, which is equal to the field calculated using that same charge density, i.e. the field is self-consistent.

In order to solve the Roothaan-Hall matrix equation, it must first be transformed into a conventional eigenvalue equation, by eliminating the overlap matrix S. This is done by determining a transformation matrix X such that

$$\mathsf{X}^{\dagger}\mathsf{S}\mathsf{X} = \mathsf{I} \tag{1.21}$$

$\mathbf{X}$ is often set to the inverse square root[1] of $\mathbf{S}$. By defining a new coefficient matrix $\mathbf{C'}$ as

$$\mathbf{C'} = \mathbf{X}^{-1}\mathbf{C} \tag{1.22}$$

$$\mathbf{C} = \mathbf{X}\mathbf{C'} \tag{1.23}$$

equation 1.15 can be rewritten as

$$\mathbf{FXC'} = \mathbf{SXC'E} \tag{1.24}$$

By premultiplying by $\mathbf{X}^{\dagger}$, and defining a new transformed Fock matrix

$$\mathbf{F'} = \mathbf{X}^{\dagger}\mathbf{FX} \tag{1.25}$$

the transformed Roothaan equations are produced:

$$\mathbf{F'C'} = \mathbf{C'E} \tag{1.26}$$

The Self Consistent Field procedure is then carried out as follows:

1. Read in the atoms to be modelled by QM. Decide upon a suitable basis set, $\{\phi_\mu\}$.

2. Calculate the overlap integrals, $S_{\mu\nu}$, the transformation matrix $\mathbf{X}$, the core terms, $H_{\mu\nu}^{core}$, and the two electron integrals, $(\mu\nu|\lambda\sigma)$. These do not change during the SCF procedure.

3. Obtain a first guess for the density matrix $\mathbf{P}$, usually from a calculation at a lower level of theory.

[1]This is known as *symmetric orthogonalisation*. In practice, *canonical orthogonalisation* deals rather better with basis sets which are close to linear dependence.

4. Calculate the Fock matrix $\mathbf{F}$ using this density matrix and the precalculated integrals, using equation 1.16. Transform it to $\mathbf{F}'$ using equation 1.25.

5. Diagonalise $\mathbf{F}'$ to obtain $\mathbf{C}'$ and $\mathbf{E}$ with equation 1.26.

6. Calculate the new density matrix $\mathbf{P}$.

7. Check to see if the calculation has converged. If not, use the new density matrix and return to step 4.

The convergence of an SCF calculation is usually determined by comparing the old and new properties. Convergence is achieved if the difference between the old and new energy, or between the old and new density, is below a certain threshold. Under some circumstances, the energy can fluctuate between SCF cycles, rather than smoothly approaching a minimum. This problem can often be countered by using a different initial guess at the density, or by altering the determined coefficients before feeding them back in to the next SCF iteration; one such method is known as Direct Inversion of Iterative Subspace, or DIIS [51].

At completion of the SCF, a set of K spatial orbitals are determined, with associated energies. The $N/2$ orbitals with the lowest energy are assumed to each be filled by two electrons with paired spins.

It should be pointed out that a typical QM calculation requires an enormous number of two-electron integrals $(\mu\nu|\lambda\sigma)$ to be calculated. Their values do not change during a calculation; however, the amount of storage required for them all may be prohibitive, such that they must be stored on a computer's hard disk rather than directly in memory. Hard disks are comparatively slow, and thus it may actually be faster to recalculate some or all of these integrals if they cannot be held in memory. Such an approach is known as *direct* SCF. An *in-core* SCF calculation avoids the disk overhead by storing all integrals directly in memory, and thus is only suitable for small systems.

Electron correlation

A consequence of the Hartree-Fock approximation, that each electron moves in the average field of the other electrons, is that the motions of electrons are not as correlated as they should be. The difference between the Hartree-Fock energy and the true value is thus termed 'correlation energy' [52]. A complete treatment of correlation is important for accurate energetics, particularly in systems with many electrons. Within the framework of conventional Hartree-Fock theory, two main methods are used to account for correlation energy. The first of these is configuration interaction [5] (CI) in which the wavefunction is actually written as a linear combination of the standard Hartree-Fock solution for the ground state, plus additional excited state configurations. An alternative method is Møller-Plesset theory [8], in which the wavefunction is improved by the addition of higher-order perturbations. Both methods are orders of magnitude more expensive in terms of computer time, however, and so are generally only ever used to give more accurate energies for geometries optimized at a lower level of theory.

1.2.1 Semi-empirical methods

The majority of the time in a typical QM calculation is spent calculating the 2-electron integrals depicted in equation 1.19. Thus, significant time savings could be achieved by ignoring or approximating some of these integrals; this is exactly what is done by the 'semi-empirical' methods, so called because the approximations to certain integrals are parameterized from experiment. Such methods are thus a combination of the conventional SCF procedure used in *ab initio* methods and empirical data. The approximations result in large time savings, and the parameterization from experiment means that better results can often be obtained from semi-empirical methods than from *ab initio* calculations with small basis sets, partly because electron correlation is implicitly included in the parameterization.

The number of one- and two-electron integrals can be further reduced by considering only the valence electrons in a calculation. The remaining 'core' electrons are subsumed into the nucleus, and their charges added to the effective nuclear charge. This is in general a reasonable approximation, as the valence electrons play the largest part in bonding. The valence electrons are then treated with a basis set of Slater-type atomic orbitals — for example, Li to Ne require four such orbitals: $2s$, $2p_x$, $2p_y$ and $2p_z$. This can be done (rather than using Gaussian contractions) since the approximations employed greatly simplify the functional form of the various integrals.

Neglect of Diatomic Differential Overlap

The 'neglect of diatomic differential overlap' (NDDO) model [53] ignores only the overlap between orbitals on different atoms. (Some other semi-empirical methods of the same era, such as the similar 'complete neglect' (CNDO) method [54], ignore other overlap terms as well, in more severe approximations.) This results in the following Fock matrix elements, modified from the *ab initio* matrix presented in equation 1.16, where the assumption is made that the atomic orbitals ϕ_μ and ϕ_ν are located on atom A, while ϕ_λ and ϕ_σ are located on atom B:

$$F_{\mu\mu} = H_{\mu\mu}^{core} + \sum_{\nu}^{A} P_{\nu\nu} \left[(\mu\mu|\nu\nu) - \tfrac{1}{2}(\mu\nu|\mu\nu)\right] + \sum_{B\neq A} \sum_{\lambda,\sigma}^{B} P_{\lambda\sigma}(\mu\mu|\lambda\sigma) \tag{1.27}$$

$$F_{\mu\nu} = H_{\mu\nu}^{core} + \tfrac{1}{2} P_{\mu\nu} \left[3(\mu\nu|\mu\nu) - (\mu\mu|\nu\nu)\right] + \sum_{B\neq A} \sum_{\lambda,\sigma}^{B} P_{\lambda\sigma}(\mu\nu|\lambda\sigma) \tag{1.28}$$

$$F_{\mu\lambda} = H_{\mu\lambda}^{core} - \tfrac{1}{2} \sum_{\nu}^{A} \sum_{\sigma}^{B} P_{\nu\sigma}(\mu\nu|\lambda\sigma) \tag{1.29}$$

The core Hamiltonian matrix is also altered from the *ab initio* expression (equation 1.17) as below:

$$H_{\mu\mu}^{core} = U_{\mu\mu} + \sum_{B \neq A} V_{\mu\mu,B} \tag{1.30}$$

$$H_{\mu\nu}^{core} = \sum_{B \neq A} V_{\mu\nu,B} \tag{1.31}$$

$$H_{\mu\lambda}^{core} = \beta_{\mu\lambda} \tag{1.32}$$

$U_{\mu\mu}$ is the kinetic energy of an electron in orbital ϕ_μ, plus its interaction with the core of atom A. Note that Z_A is the core charge of atom A, comprising the nucleus plus core electrons, and not the nuclear charge as in the similar *ab initio* expression (equation 1.17). $U_{\mu\mu}$ can be calculated explicitly, but is more usually parameterized to match experimental data such as ionisation energies.

$$U_{\mu\mu} = \langle \mu | -\tfrac{1}{2}\nabla_1^2 - \frac{Z_A}{r_{1A}} | \mu \rangle \tag{1.33}$$

$V_{\mu\nu,B}$ is the interaction between an electron in the distribution $\phi_\mu\phi_\nu$ (where μ and ν may be the same) and the core of atom B:

$$V_{\mu\nu,B} = \langle \mu | -\frac{Z_B}{r_{1B}} | \nu \rangle \tag{1.34}$$

Finally, $\beta_{\mu\lambda}$ is the *resonance integral*, which is the main contribution to bonding, and is assumed to be proportional to the overlap between the orbitals, $S_{\mu\lambda}$, and to a parameter dependent purely on the types of atoms A and B, β_{AB}^0. $S_{\mu\lambda}$ itself is calculated explicitly from the overlap of the necessary Slater orbitals, and thus the orbital exponents ζ are introduced as parameters into the method.

$$\beta_{\mu\lambda} = \beta_{AB}^0 S_{\mu\lambda} \tag{1.35}$$

Modified Neglect of Differential Overlap

The 'modified neglect of differential overlap' (MNDO) model [55] improves NDDO, largely by refining the H^{core} term, to overcome observed problems caused by the neglect of penetration. Firstly, the electron-core interaction is written as an interaction between the electron density $\phi_\mu\phi_\nu$ and a representative s orbital on atom B:

$$V_{\mu\nu,B} = -Z_B(\mu_A\nu_A|s_Bs_B) \tag{1.36}$$

The core-core repulsion terms are also altered from the *ab initio* functional form given in equation 1.4, such that they agree rather better with the electron-core terms above. This introduces the extra parameter α_A into the model. Note that a different functional form is used for N–H and O–H bonds, to improve the fit to experimental data:

$$E_{AB} = Z_AZ_B(s_As_A|s_Bs_B)\{1 + \mathrm{e}^{-\alpha_A R_{AB}} + \mathrm{e}^{-\alpha_B R_{AB}}\} \tag{1.37}$$

$$E_{XH} = Z_XZ_H(s_Xs_X|s_Hs_H)\{1 + R_{XH}\mathrm{e}^{-\alpha_X R_{XH}} + \mathrm{e}^{-\alpha_H R_{XH}}\} \tag{1.38}$$

Finally, a slightly different form is used in MNDO for the resonance integral, adding parameters β_μ^A that are dependent both on the type of orbital and the atom:

$$\beta_{\mu\lambda} = \tfrac{1}{2}(\beta_\mu^A + \beta_\lambda^B)S_{\mu\lambda} \tag{1.39}$$

More recent methods

The AM1 (Austin Model 1) model [56] further improves upon MNDO, which has difficulty in reproducing the energies of many systems such as those containing hydrogen bonds, crowded molecules, or those including four-membered rings. Most of MNDO's problems are believed to stem from its tendency to overestimate repulsions between atoms when separated by roughly their van der Waals distance, and thus AM1

aims to improve on this by adding both attractive and repulsive Gaussian terms to the core-core term. The AM1 core-core term, $E_{AB,AM1}$, is thus related to the corresponding MNDO term by the following:

$$E_{AB,AM1} = E_{AB,MNDO} + \frac{Z_A Z_B}{R_{AB}} \left\{ \sum_i K_{A_i} \mathrm{e}^{-L_{A_i}(R_{AB}-M_{A_i})^2} + \sum_j K_{B_j} \mathrm{e}^{-L_{B_j}(R_{AB}-M_{B_j})^2} \right\} \tag{1.40}$$

Thus, AM1 adds a number of extra K, L and M parameters to each atom type, on top of those already used by MNDO.

The PM3 model [57, 58] uses the same terms as AM1, and differs only in the parameters used. AM1 parameters were manually chosen based on chemical knowledge, while PM3 parameters were automatically fitted to experimental properties (and, where experimental data were unavailable, *ab initio* results).

AM1, PM3 and MNDO are perhaps the most frequently used semi-empirical methods in modern computational chemistry. (Simpler methods are generally only used to obtain a starting guess for the density for use in *ab initio* methods.) They are unable to cover many important elements, however, such as most transition metals, due to the lack of consideration of d orbitals in their parameterization. This problem has been remedied somewhat in recent years, however, with the development of methods that do include d orbitals, such as AM1/d [59] and MNDO/d [60–62].

1.2.2 Density Functional Theory

As has been previously mentioned, *ab initio* methods do not account for electron correlation explicitly. Although there are ways of adding electron correlation, they are extremely expensive. An alternative approach is that provided by Density Functional Theory (DFT) which uses the electron density $\rho(\boldsymbol{r})$ rather than an N-electron wave function [9]. DFT was initially proposed as a statistical technique [63], and became

more widely applicable after the work of Hohenberg and Kohn [64], who presented two fundamental theorems.

The first Hohenberg-Kohn theorem states that the external electrostatic potential is uniquely determined by the electron density $\rho(\boldsymbol{r})$, i.e. the density determines a unique Hamiltonian and wavefunction. The energy can thus be written as a *functional* of the density:

$$\begin{aligned} E[\rho] &= V_{ne}[\rho] + T[\rho] + V_{ee}[\rho] \\ &= \int \rho(\boldsymbol{r})\nu(\boldsymbol{r})\mathrm{d}\boldsymbol{r} + T[\rho] + V_{ee}[\rho] \end{aligned} \tag{1.41}$$

where the functionals $V_{ne}[\rho]$, $T[\rho]$ and $V_{ee}[\rho]$ determine the nuclear-electron interaction energy, the electron kinetic energy, and the electron-electron interaction energy respectively from the density, and $\nu(\boldsymbol{r})$ is the potential in which the electrons move.

The second theorem allows the use of the variational principle for determining a DFT wavefunction, i.e. a guess at the density ρ' with an associated wavefunction Ψ' is related to the true density ρ as follows:

$$\langle \Psi' | H | \Psi' \rangle = E[\rho'] \geq E[\rho] \tag{1.42}$$

Unfortunately, the form of the $T[\rho]$ and $V_{ee}[\rho]$ functionals is not known. This problem was addressed by Kohn and Sham [65] who considered a system with the same electron density but non-interacting electrons. (Such a system can be solved exactly via a traditional SCF technique.) The energy functional can then be written as

$$E[\rho] = V_{ne}[\rho] + T_s[\rho] + J_s[\rho] + E_{xc}[\rho] \tag{1.43}$$

where the $T_s[\rho]$ and $J_s[\rho]$ functionals are of known form and give the kinetic and Coulombic energy of the non-interacting system, while the $E_{xc}[\rho]$ functional accounts for the exchange and correlation corrections between the real and non-interacting

systems. This leads to the Kohn-Sham equations, which allow Kohn-Sham orbitals ψ_i to be determined by means of a conventional SCF approach:

$$\left[-\frac{1}{2}\nabla^2 + \nu(\boldsymbol{r}) + \int \frac{\rho(\boldsymbol{r'})}{|\boldsymbol{r} - \boldsymbol{r'}|}\mathrm{d}\boldsymbol{r'} + \nu_{xc}(\boldsymbol{r})\right]\psi_i(\boldsymbol{r}) = E_i\psi_i(\boldsymbol{r}) \tag{1.44}$$

Note that the exchange term from Hartree-Fock theory is replaced by the exchange-correlation potential $\nu_{xc}(\boldsymbol{r})$, which is related to the corresponding functional by

$$\nu_{xc}(\boldsymbol{r}) = \frac{\delta E_{xc}[\rho]}{\delta\rho(\boldsymbol{r})} \tag{1.45}$$

There is currently no way in which this term can be exactly calculated for molecular systems; in fact, an approximate guess cannot even be systematically improved by using a larger basis set and/or more SCF cycles, as in Hartree-Fock theory. It can, however, be calculated exactly for certain model systems, and then corrected for use in real systems. (It is for this reason that DFT is often considered a semi-empirical theory.)

One such model system is the uniform electron gas — a number of electrons in a fixed volume, throughout which positive charge is homogeneously distributed so as to maintain overall neutrality. If the electron density in a molecular system is also assumed to be largely uniform, or at least a slowly varying function of position, then this exact solution can be applied. This is known as the Local Spin Density Approximation (LSDA) or Local Density Approximation (LDA). The most common 'LDA' functional is known as S-VWN5, which uses Slater's [66] expression for the exchange and Vosko, Wilk and Nusair's [67] expression for the correlation.

LDA is a rather poor approximation in practice, as electron density tends to actually be rather inhomogeneous due to the presence of the nuclei. Thus, more recent exchange and correlation functionals incorporate the gradient of the electron density, a non-local correction, and are said to employ the 'generalised gradient approximation' (GGA) [68]. For example, the BLYP functional incorporates Becke's B88 [69] gradient-corrected

exchange functional and Lee, Yang and Parr's LYP [70] gradient-corrected correlation term.

Recent work has suggested that actually incorporating exact Hartree-Fock exchange into the functional gives improved results. This leads to so-called 'hybrid GGAs' such as B3PW91 [71], where the standard LSDA exchange energy E_x^{LSDA} is mixed with the exact HF exchange energy, with Becke's gradient correction ΔE_x^{B88}, while the standard LSDA correlation energy E_c^{LSDA} is gradient-corrected by the PW91 [72] functional:

$$E_{xc}^{B3PW91} = (1 - a_0)E_x^{LSDA} + a_0 E_x^{HF} + a_x \Delta E_x^{B88} + E_c^{LSDA} + a_c \Delta E_c^{PW91} \qquad (1.46)$$

The B3LYP [73] functional is more widely used, in which the LYP functional is used instead of PW91 (as LYP is not easily separable into a local and non-local part, local VWN correlation is included to approximate this):

$$E_{xc}^{B3LYP} = (1 - a_0)E_x^{LSDA} + a_0 E_x^{HF} + a_x E_x^{B88} + a_c E_c^{LYP} + (1 - a_c)E_c^{VWN} \qquad (1.47)$$

In both of these equations, a_0, a_x and a_c are adjustable parameters. Finally, the 'half-and-half' (BH&HLYP) functional is occasionally used, in which the exchange contribution is made up from equal parts of exact Hartree-Fock and DFT gradient-corrected exchange [74].

The most modern functionals such as B98 [75] fall into the 'meta GGA' category, which improve on hybrid GGAs by adding additional terms such as the Laplacian of the density, and have been shown [68] to give the most reliable DFT molecular properties to date.

1.3 Molecular mechanics force fields

A complete quantum mechanical treatment of a typical system of biochemical interest is extremely computationally expensive. The evaluation and storage of all necessary integrals, and the solution of the matrix equations, limits most QM simulations to systems of under 100 atoms in size. In comparison, typical systems (e.g. solvated proteins or membranes) can contain hundreds of thousands of atoms. Thus, a more lightweight and approximate technique is required.

This role is fulfilled by 'classical' or molecular mechanical (MM) force fields. The MM model treats the nuclei as point charges in exactly the same way as QM treatments, but does not explicitly model electron density; instead, the electron density is assumed to be localized. Most usually this density is centred on the nuclear positions, such that only one charge site per atom need be considered, with a total charge equal to the sum of the electronic and nuclear charges. Electronic density can also be placed away from nuclear sites, to simulate lone pairs. Obviously, if every atom in a system were treated independently in this way, the charge on each nucleus would be zero; thus in order to reproduce the electrostatic environment of a molecule, as determined from experiment or QM calculations, each atom is assigned a fixed *partial charge*, consisting of the nuclear charge plus a fraction of the electron density, chosen to represent the average amount of charge that is lost or gained by the atom in a molecular environment.

Such an approximation can reproduce the molecular electrostatic potential relatively accurately, with suitable parameterization. However, the static and implicit electron density is unable to reproduce more subtle features observed in a QM treatment. Firstly, the electron density is not able to concentrate between atoms and form bonds. Thus, a purely electrostatic interaction between bonded atoms will cause the atoms to be violently pushed apart or brought together, dependent on the sign of the product of the partial charges. Secondly, van der Waals interactions are not reproduced, as neither dispersive (London) nor exchange (Pauli) repulsion forces

are predicted by this model. (The fixed nature of the electron density forbids the fluctuations that give rise to the instantaneous dipole-dipole dispersive interactions, while exchange forces are caused by the interaction between electrons of the same spin — spin which is neglected in this treatment.) To overcome these deficiencies, a 'balls and springs' model is used. The 'balls', generally centred on the nuclei, interact by means of a parameterized van der Waals and electrostatic pair potential, while bonds are approximated by parameterized 'springs'. Electrostatic interactions are then not calculated for atoms separated by less than a certain number of bonds (three, in the default CHARMM force field).

The interaction energy of a bond is modelled particularly well by the Morse potential [76]. However, generally a bond is modelled near equilibrium, where it can be treated as a simple spring that obeys Hooke's Law, giving the potential as a simple harmonic function of the bond length l, parameterized by a reference bond length l_0 and a force constant k:

$$V_{bond} = \frac{k}{2}(l - l_0)^2 \tag{1.48}$$

Generally speaking, bond terms alone are not sufficient to maintain realistic molecular geometries; explicit electron density is also responsible for changes in energy as angles between 3 atoms are changed, for hindered rotation of bonds, and for out-of-plane motions. These must also be modelled in an MM force field with explicit energy functions.

Bond angle energies are typically treated as harmonic functions of a reference angle, θ_0, and a force constant, k:

$$V_{angle} = \frac{k}{2}(\theta - \theta_0)^2 \tag{1.49}$$

To account for hindered rotation about bonds, an additional energy term is given as a function of the dihedral angle, χ. (For a set of 4 atoms, [A B C D], χ is the angle between the two planes defined by atoms [A B C] and [B C D] respectively.) This function is parameterized by the *multiplicity* of the bond, n, the *phase*, δ, and a force

constant, K_χ:

$$V_{dihed} = K_\chi[1 + \cos(n\chi - \delta)] \tag{1.50}$$

Finally, out-of-plane motions are often modelled as so-called 'improper' dihedral torsions (where the atoms are not linked in the order A-B-C-D, but rather the three atoms A, C and D are linked to a central atom B) again by a simple harmonic term parameterized by the dihedral angle χ and force constant K_{impr}:

$$V_{impr} = K_{impr}(\chi - \chi_0)^2 \tag{1.51}$$

The van der Waals terms are typically evaluated by means of a Lennard-Jones [77] or '6-12' potential, between pairs of atoms i and j, which combines a steep repulsive r^{-12} term with a shallower attractive r^{-6} term, parameterized by the well depth ϵ_{ij} and the collision diameter σ_{ij}. These two parameters are often determined from corresponding atomic properties by the simple combination rules $\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$ and $\sigma_{ij} = \frac{1}{2}(\sigma_i + \sigma_j)$, i.e. as the geometric and arithmetic mean, respectively.

$$V_{vdW}(r_{ij}) = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{1.52}$$

Electrostatic terms are again evaluated between pairs of atoms using, in the simplest case, Coulomb's law, dependent solely on the dielectric constant ϵ_0, the interatomic distance r_{ij} and the atomic partial charges q_i and q_j:

$$V_{elec}(r_{ij}) = \frac{1}{4\pi\epsilon_0}\frac{q_i q_j}{r_{ij}} \tag{1.53}$$

The total molecular mechanics force field is then simply given as the sum of all the individual energy terms for the whole system:

$$V_{total} = \sum_{\text{bonds}} V_{bond} + \sum_{\text{angles}} V_{angle} + \sum_{\text{dihed.}} V_{dihed} + \sum_{\text{impr.}} V_{impr} + \sum_{i \neq j}(V_{vdW} + V_{elec}) \tag{1.54}$$

1.4 Hybrid QM/MM

It has been seen that quantum mechanical methods are widely applicable, require little or no parameterization, and can give very accurate molecular properties, but at large computational cost (section 1.2). On the other hand, molecular mechanical methods are extremely fast in comparison, and can be used to study very large systems (section 1.3). However, systems of chemical interest are often large (making them difficult for conventional QM treatments) and involve bond breaking or forming processes, highly polar regions, or atom types for which parameterization is difficult (e.g. transition metal ions).

Fortunately, in many systems the 'difficult' region requiring QM is small, and the majority of atoms in the system are amenable to treatment with MM methods. For example, in many proteins the active site involves only a few residues, and the bulk of the protein, and the surrounding solvent, are not *directly* involved in chemical reactions or binding (although they still have steric, electronic, and structural effects of course, which must be accounted for). Such systems can be looked at with hybrid QM/MM methods, where the active site is modelled with a QM Hamiltonian, and the surrounding region with a molecular mechanics potential [16, 17, 78]. The two regions interact and influence each other by means of a coupled QM/MM Hamiltonian, as shown in figure 1.1. (The boundary around the MM region must also be treated; methods for this are discussed in section 1.7.)

In a QM/MM treatment, the total Hamiltonian is given by

$$\hat{H}_{total} = \hat{H}_{QM} + \hat{H}_{MM} + \hat{H}_{QM/MM} \tag{1.55}$$

Here, $\hat{H}_{QM}$ is the Hamiltonian of the QM region, forming part of its Schrödinger equation. $\hat{H}_{MM}$ is the energy of the MM region, and $\hat{H}_{QM/MM}$ covers the interaction

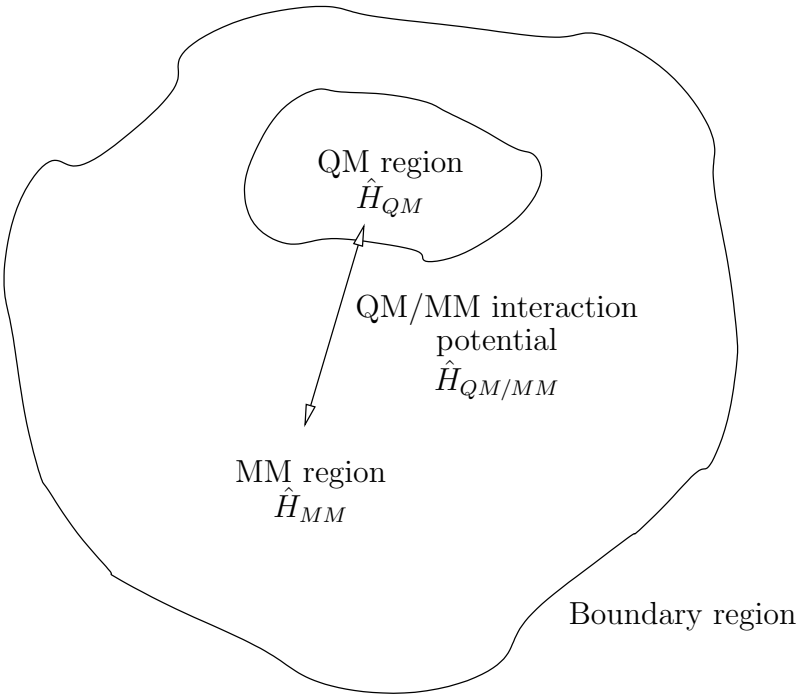

Figure 1.1: The hybrid QM/MM treatment. The system is broken up into a large MM region and a much smaller QM system.

between the two regions, and is defined as:

$$\hat{H}_{QM/MM} = -\sum_{e,i} \frac{Q_i}{r_{e,i}} + \sum_{m,i} \frac{Z_m Q_i}{r_{m,i}} + \hat{H}_{vdW} \tag{1.56}$$

where i is summed over all MM partial charges, m over all QM nuclei, and e over all QM electrons. The first term gives the one-electron interaction between the QM electron density and the MM partial charges, the second is a standard Coulomb interaction between the QM nuclei and the MM charges, and the final term is a van der Waals interaction term, generally between the QM and MM nuclei. This final term is required since although electron density is explicitly treated in the QM region, it is not in the MM region.

Solution of the coupled Hamiltonian gives the total energy as the sum of the energies of the QM and MM systems, and the interaction energy between them:

$$E_{total} = E_{QM} + E_{MM} + E_{QM/MM} \tag{1.57}$$

The energy of interaction between the QM system and a single MM partial charge, Q_i, is then given by:

$$\begin{aligned} E^i_{QM/MM} &= \left\langle \Psi \middle| \hat{H}^i_{QM/MM} \middle| \Psi \right\rangle \\ &= \sum_\mu \sum_\nu P_{\mu\nu} I^i_{\mu\nu} + \sum_m \frac{Z_m Q_i}{r_{m,i}} + E^i_{vdW} \end{aligned} \tag{1.58}$$

where $P_{\mu\nu}$ is a density matrix element (equation 1.18) and $I^i_{\mu\nu}$ a one-electron integral:

$$I^i_{\mu\nu} = -Q_i \left\langle \mu \right| \frac{1}{r_{e,i}} \left| \nu \right\rangle \tag{1.59}$$

The QM/MM Hamiltonian must be modified when a semi-empirical Hamiltonian is used for the QM region, as methods such as MNDO alter the core-core terms (as discussed in section 1.2.1); simply using the *ab initio* expression tends to overestimate the interaction energies and forces [17]. Instead, the following one-electron integrals are used:

$$I^i_{\mu\nu} = -Q_i (\mu^Q \nu^Q | s^M s^M) \tag{1.60}$$

where s^M is a notional s orbital on the MM atom i, and Q identifies the QM atom. In line with the MNDO modified core-core terms (equation 1.37) the QM core-MM atom interactions are also modified, leading to the following QM/MM energy:

$$\begin{aligned} E^i_{QM/MM,MNDO} = \sum_Q \Big\{ & Z_Q Q_i (s^Q s^Q | s^M s^M) \\ & + |Z_Q Q_i| (s^Q s^Q | s^M s^M) [\mathrm{e}^{-\alpha_Q R_{Qi}} + \mathrm{e}^{-\alpha_i R_{Qi}}] \Big\} \\ & + \sum_\mu \sum_\nu P_{\mu\nu} I^i_{\mu\nu} \end{aligned} \tag{1.61}$$

For AM1 and PM3 the additional terms that contribute to the core-core interaction (equation 1.40) are also carried over into the QM/MM term. The QM/MM interaction is optimized by adjusting the semi-empirical parameters for the MM atoms; generally

speaking the equivalent parameters for the QM atoms are left unchanged from the original QM-only parameterization.

From an implementation point of view, QM/MM is relatively straightforward. The energy is generally evaluated by linking a QM package with an MM program; a commonly-used combination is GAMESS-US [79] for the QM region and CHARMM [10] for the MM region. (In the work described in this thesis, use is also made of GAMESS-UK [80], CADPAC [24, 81], and the version of MOPAC [17, 82] included in CHARMM.) The first term in equation 1.58 requires the most modification to existing code; however, it is a simple one-electron interaction term, and QM codes contain very similar code already (for calculating the interaction between electrons and nuclei). The second term is a simple Coulombic interaction, and can be trivially calculated by either the QM or MM code. The final term is generally calculated by the MM package, in an analogous way to pure MM-MM van der Waals interactions.

The interaction between the QM and MM regions is made more complicated if there are bonds that cross the QM/MM boundary. The electron density of the bond is not properly described, as the QM region lacks the electrons and atomic orbitals of one of the atoms. A number of techniques [22, 23] have been used to treat these bonds. The simplest are perhaps the 'link atom' methods, where the QM region is capped with an additional atom (usually hydrogen) which lies on the bond that crosses the boundary. This atom is used to satisfy the valence of the QM region only, and does not usually interact with the MM region. More complex methods such as Generalized Hybrid Orbital [83] (GHO) attempt to optimize the electron density of such frontier bonds for a better join with the MM system. Link atoms and similar methods are vitally important for the simulation of large systems such as proteins, where it is unfeasible to treat the whole molecule with QM. However, for the systems considered in this work, small molecules are used in the QM region, such that no bonds are cut by the QM/MM boundary, and so the complications of link atoms are avoided.

1.5 Polarization

Conventional molecular mechanics force fields do not explicitly treat electron density. However, it is possible to augment such force fields by adding point polarizable dipoles to the treatment, which approximate the fluctuations in electron density that QM techniques model. An induction energy term [84] is then added to the standard non-polarizable force field energy, which for a set of particles, each with polarizability α, is given by

$$U_{pol} = -\sum_i \boldsymbol{E}_i \cdot \boldsymbol{\mu}_i + \frac{1}{2\alpha}\sum_i |\boldsymbol{\mu}_i|^2 \tag{1.62}$$

where $\boldsymbol{E}_i$ is the electric field acting on particle i and $\boldsymbol{\mu}_i$ the particle's induced dipole moment. Obviously, the electric field is determined by the charge distribution of all other molecules, and the induced dipole is influenced by the field (under most circumstances, $\boldsymbol{\mu}_i = \alpha \boldsymbol{E}_i$). Thus, the induction energy is not pairwise-additive (the energy must be evaluated for the whole system, and not as a sum over subsystems), and it is necessary to solve $3N$ linear equations to determine the energy of an N-particle system. This is computationally expensive; approximate methods do exist, however, as will be discussed in Chapter 3.

1.6 Optimization of pair potentials

The time required to calculate the electrostatic and van der Waals energy of a system of N particles scales as $\mathcal{O}(N^2)$ (i.e. the time is proportional to N^2), as every particle interacts with every other. This scaling is prohibitively expensive in systems comprising large numbers of atoms, and so various approximations are made in order to save on computer time. The simplest such approximation is to truncate each of the pair potentials — i.e. if the interatomic distance r is greater than the cutoff distance r_{off}, then the interaction energy is simply not calculated, and is assumed to be zero. This is a reasonable approximation for a Lennard-Jones interaction, as the r^{-6} term drops

off quickly with distance, but quite poor for a r^{-1} Coulomb potential, which drops off much more slowly. Thus, much better results are achieved if a shifted or switched potential is used. The former scales the original function such that it drops off to zero at r_{off} by multiplying by a shifting function, F_{shift}, of the form:

$$F_{\text{shift}} = \begin{cases} (1 - (r_{ij}/r_{\text{off}})^2)^2 & : \quad r_{ij} < r_{\text{off}} \\ 0 & : \quad r_{ij} \geq r_{\text{off}} \end{cases} \tag{1.63}$$

A switched potential, on the other hand, exhibits normal behaviour up to an inner cutoff, r_{on}, and is then scaled by a switching function, F_{switch}, which causes it to drop off to zero at r_{off}:

$$F_{\text{switch}} = \begin{cases} 1 & : \quad r_{ij} \leq r_{\text{on}} \\ \dfrac{(r_{\text{off}} - r_{ij})^2(r_{\text{off}} + 2r_{ij} - 3r_{\text{on}})}{(r_{\text{off}} - r_{\text{on}})^3} & : \quad r_{\text{on}} < r_{ij} < r_{\text{off}} \\ 0 & : \quad r_{ij} \geq r_{\text{off}} \end{cases} \tag{1.64}$$

A variety of other functions can be applied [85] to improve both the overall behaviour of a pair potential, and the continuity of the function and its derivatives around the cutoff distances r_{on} and r_{off}.

1.7 Boundary treatments

The properties of a material's surface can differ greatly from those of its interior. Although some calculations are interested in surface effects, for many calculations — such as those presented in this thesis — such effects are undesirable. This is a particular problem for computational chemistry studies, where the number of atoms studied is small, such that a significant fraction of the system's atoms are on or near the surface.

1.7.1 Periodic boundary conditions

One way of removing surface effects is to use periodic boundary conditions [86] (PBC). A cubic box containing the system is infinitely replicated throughout 3D space, such that for a primary particle at $\boldsymbol{r}$ there are an infinite number of images, at

$$\boldsymbol{r'} = \boldsymbol{r} + \boldsymbol{t}_n \tag{1.65}$$

where $\boldsymbol{t}_n$ is a vector between the primary box and an image box n, given for a cubic box of side length L as:

$$\boldsymbol{t}_n = t_x L\boldsymbol{i} + t_y L\boldsymbol{j} + t_z L\boldsymbol{k} \qquad t_x, t_y, t_z \in \mathbb{Z} \tag{1.66}$$

The procedure is also frequently applied to more complex tessellating shapes, such as cuboids or truncated octahedra. The total electrostatic interaction energy (equation 1.53) for an N-particle system is then written as

$$V_{elec} = \frac{1}{2} \sum_{n}^{'} \sum_{i}^{N} \sum_{j}^{N} \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{|\boldsymbol{r}_i - \boldsymbol{r}_j + \boldsymbol{t}_n|} \tag{1.67}$$

where the sum over n includes the primary box with $n = 0$, but excludes the self-interaction (where $i = j$) for this box only. (A similar expression is used for the van der Waals terms.) Obviously, an infinite or near-infinite number of particles cannot really be considered; other than the computational expense, the sum of nonbond interactions is conditionally convergent — i.e. the result depends on the order in which the terms are summed. Most usually, the *minimum image convention* is used. This enforces a cutoff distance of half the box length; thus any given particle cannot interact with more particles than are present in the primary simulation cell. (Additionally, particles cannot interact with their own images, which prevents undesirable periodic effects such as phonon waves from occurring.) In simple cases, energies and forces can then

be calculated in the usual way, after adding multiples of the box length to distance components to yield the minimum image vector between any two particles. More generally, the images are created explicitly and moved whenever the primary atoms are. If, during the course of a simulation, a primary atom leaves the central simulation cell, it can be swapped with a closer image (for example, when using a cubic boundary, as a particle leaves through one face of the cube, an image will enter through the opposite face). Images are only created up to a certain user-specified distance away from the primary simulation cell, to save on storage and computer time.

1.7.2 Ewald summation

A more rigorous technique than periodic boundary conditions, that gives more accurate long-range forces, is Ewald summation [87]. As with PBC, the system is treated as an infinite number of replicated simulation boxes. The infinite and conditionally convergent sum of electrostatic interactions, equation 1.67, is then solved by splitting the energy into several components:

$$E = E_{\text{dir}} + E_{\text{rec}} - E_{\text{self}} + E_{\text{surf}} \tag{1.68}$$

The bulk of the calculation is accounted for by the terms E_{dir} and E_{rec}. A charge distribution — usually a simple Gaussian — is added to each charge site for the first term, and removed for the second. The individual terms are then convergent; the first, E_{dir}, the so-called 'direct' or 'real-space' term, is evaluated by the CHARMM [10] Ewald implementation for all interactions up to a distance cutoff:

$$E_{\text{dir}} = \frac{1}{2} \sum_{i \neq j}^{N} \frac{q_i q_j}{4\pi\epsilon_0} \frac{\text{erfc}(\sqrt{\alpha} r_{ij})}{r_{ij}} \tag{1.69}$$

where erfc(x) is the *complementary error function*,

$$\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} \mathrm{d}t \tag{1.70}$$

The parameter α determines the width of the added charge distributions, and is chosen such the E_{dir} term converges at the cutoff distance. The second term, E_{rec}, the 'reciprocal' or 'k-space' term, is evaluated in reciprocal space:

$$E_{\mathrm{rec}} = \frac{2\pi}{L^3} \frac{1}{4\pi\epsilon_0} \sum_{k \neq 0}^{k_{max}} \frac{1}{k^2} \exp(-k^2/4\alpha) S(\boldsymbol{k}) S(-\boldsymbol{k}) \tag{1.71}$$

where $S(\boldsymbol{k})$ is a function of a reciprocal lattice vector $\boldsymbol{k}$:

$$S(\boldsymbol{k}) = \sum_{j=1}^{N} q_j \exp(\mathrm{i}\boldsymbol{k} \cdot \boldsymbol{r}_j) \tag{1.72}$$

$\boldsymbol{k}$ itself is given as

$$\boldsymbol{k} = (2\pi/L)\boldsymbol{l} \qquad l_x, l_y, l_z \in \mathbb{Z} \tag{1.73}$$

The number of k-vectors over which to sum, k_{max}, is chosen to balance the computational cost against the convergence of E_{rec}. E_{rec} also incorporates a non-physical interaction of every charge with itself; this must be corrected for by subtracting a self-energy term, E_{self}:

$$E_{\mathrm{self}} = \sqrt{\frac{\alpha}{\pi}} \sum_{k=1}^{N} \frac{q_k^2}{4\pi\epsilon_0} \tag{1.74}$$

Finally, a surface term, E_{surf}, is added. This accounts for the medium that surrounds the collection of periodic boxes. In most cases, this medium is assumed to be a vacuum, necessitating the following:

$$E_{\mathrm{surf}} = \frac{2\pi}{3L^3} \left| \sum_{i=1}^{N} \frac{q_i}{4\pi\epsilon_0} \boldsymbol{r}_i \right|^2 \tag{1.75}$$

The evaluation of the reciprocal space term in traditional Ewald calculations is the most time-consuming for large systems, scaling as $\mathcal{O}(N^2)$ if a fixed cutoff is used [88], and more modern methods aim to replace this term with one that is more efficient. The usual way of doing this is to approximate the charge distribution of the system by a regular lattice or mesh of point charges; the periodic energy of such a mesh can be evaluated rapidly by means of Fast Fourier Transform (FFT) techniques, and mapped back onto the 'real' system by interpolation. Two such approaches are the Particle Mesh Ewald (PME) method [89] and Smooth Particle Mesh Ewald [90]. The use of FFT techniques results in these methods scaling as $\mathcal{O}(N\log N)$.

1.7.3 Stochastic boundary techniques

An alternative to periodic boundaries is to treat the long-range interactions in some approximate fashion. One method suitable for homogeneous systems is the 'reaction field' technique [91–93], whereby the system is assumed to be surrounded by a continuum of constant dielectric. A more commonly-used methodology, particularly for protein simulation, is that of 'stochastic boundaries' [94,95], in which the system is divided into reaction, buffer, and reservoir regions. Generally speaking the part of the protein of particular interest is placed in the reaction region, and solvated by a sphere of water molecules in both the reaction and surrounding buffer regions. (The system is held in this spherical configuration by means of an applied boundary force.) Normal molecular dynamics (see section 1.10.1) are carried out in the reaction region, but to model the average effect of the solvent in the reservoir region (since it is not explicitly treated) a technique such as Langevin dynamics (see equation 3.29) is used in the buffer region. Thus, stochastic boundary simulations solvate a system by means of a solvent droplet, but aim to reduce the solvent surface effects in an average fashion; such simulations are often useful for large protein systems, since to fully solvate them with periodic boundary solvent would require massive numbers of explicit solvent molecules.

1.8 Charge determination

In order to compare results from QM and MM calculations, and to parameterize MM force fields from QM data, it is often useful to know the charges on the QM nuclei, as used in an MM treatment (i.e. with electronic charge included). These charges can also be used to replace the standard charges of an MM force field, to reproduce more accurately the local electrostatics of an important part of the system [96, 97]. Atomic charge cannot be determined exactly in a QM treatment, as it is not an observable; however, several approximate techniques exist. One approach is to derive charges directly from the QM electron density, using a variety of methods to assign density to specific atoms, such as in Mulliken's [98] original technique.

ESP methods, on the other hand, perform a least-squares fit between the molecular electrostatic potential (ESP) as derived from the QM wavefunction, and that derived from an MM force field using point charges on the nuclei; i.e. they minimize the function

$$y(q_1, q_2, \ldots q_n) = \sum_{i=1}^{m} \left[V_i - E_i(q_1, q_2, \ldots q_n)\right]^2 + \sum_{w} \lambda_w g_w \tag{1.76}$$

for a set of charges $\{q_j\}$ on n atoms, considering the potential at m probe points, and incorporating a set of constraints $\{g_w\}$ with their associated Lagrange multipliers λ_w. The QM electrostatic potential is given as

$$V_i = \sum_{A} \frac{Z_A}{|\boldsymbol{r}_i - \boldsymbol{R}_A|} - \sum_{\mu\nu} P_{\mu\nu} \int \frac{\phi_\mu \phi_\nu}{|\boldsymbol{r}_i - \boldsymbol{r'}|} \mathrm{d}\boldsymbol{r'} \tag{1.77}$$

(where A identifies a QM nucleus) and the corresponding MM potential, in atomic units, as

$$E_i = \sum_{j=1}^{n} \frac{q_j}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} \tag{1.78}$$

The solutions are found by solving the set of linear equations

$$\frac{\partial y}{\partial q_k} = 0 \tag{1.79}$$

$$\frac{\partial y}{\partial \lambda_w} = 0 \tag{1.80}$$

A variety of constraints can be applied, although in general only a single constraint is used, to maintain the total charge of the system at the value used in the QM simulation, q_{tot}, i.e.

$$g = q_{tot} - \sum_{j=1}^{n} q_j = 0 \tag{1.81}$$

The resulting linear equations in $(n+1)$ unknowns can be written in matrix form as

$$\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} q_1 \\ \vdots \\ q_n \\ \lambda \end{pmatrix} = \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \\ q_{tot} \end{pmatrix} \tag{1.82}$$

where

$$A_{jk} = \sum_{i=1}^{m} \frac{1}{r_{ij} r_{ik}} \tag{1.83}$$

$$Z_k = \sum_{i=1}^{m} \frac{V_i}{r_{ik}} \tag{1.84}$$

Thus, by obtaining the inverse of matrix A, the charges $\{q_j\}$ can be determined. The various ESP methods differ mainly in the positioning of the m probe points; the CHELP [99] method uses points placed on the surface of spherical shells around the atoms, CHELPG [100] uses a regular grid of points, while MK [101] uses points on and around the molecular surface as defined by Connolly [102].

1.9 Determination of minimum energy system configurations

A variety of methods exist to locate the minimum energy configuration of a system of particles, and are used in order to investigate the energetics of a system, or to remove 'bad contacts' (unphysical configurations that result in high potential energy) prior to performing more expensive calculations (such as those used to determine free energy). If the system is described by the N-dimensional vector of coordinates $\boldsymbol{q}$, then the energy of a trial configuration $\boldsymbol{q}'$ is given relative to this configuration by Taylor expansion, as

$$E(\boldsymbol{q}') = E(\boldsymbol{q}) + (\boldsymbol{q}' - \boldsymbol{q})\dot{E}(\boldsymbol{q}) + \frac{1}{2}(\boldsymbol{q}' - \boldsymbol{q})^2\ddot{E}(\boldsymbol{q}) + \dots \tag{1.85}$$

Here $\dot{E}$ ($\partial E/\partial \boldsymbol{q}$) is a 1-by-$N$ matrix of the first derivatives with respect to the vector $\boldsymbol{q}$, while $\ddot{E}$ is a N-by-N matrix of the second derivatives. These coordinates are often the three Cartesian coordinates for each atom, although many QM packages actually perform minimizations using internal coordinates instead.

1.9.1 Zero-order methods

The simplest strategy for finding the minimum energy is the 'grid search' approach. This is known as a zero-order method, as none of the energy derivatives are used; instead, the energy is simply evaluated at each of a set of points, which form an N-dimensional grid that covers the region of interest. The lowest energy then found is used as the minimum, although it can be refined by a further grid search with a finer grid. Grid searches require many energy evaluations, and are very inefficient for systems where N is non-trivially large. They do, however, have the advantage that no derivatives need be calculated.

A very different type of minimization, which also does not need any energy derivatives to function, is known as simulated annealing [103–105]. This involves

simulating the system with the Monte Carlo (section 1.10.4) technique and gradually lowering the temperature. Such a method aims to overcome a problem inherent to conventional minimization, that of becoming trapped in a local minimum. In the early stages of the simulation, the system has enough energy to traverse the energy barriers that separate local minima. As the temperature of the system is lowered, it becomes more and more preferable to adopt configurations of low energy, until at 0 K the true minimum should be achieved. Such simulations require many energy evaluations, however, and are generally only viable for extremely large systems with many local minima, where conventional minimizations fail.

Other zero-order minimization methods, not used in this work, include the simplex [106] and sequential univariate [107] techniques.

1.9.2 First-order methods

The crude grid search method can be improved upon by using the first derivatives of the energy with respect to the coordinates — i.e. $\dot{E}$ or $\partial E/\partial \boldsymbol{q}$ — in so-called 'first-order' methods. The simplest of these methods is known as the 'steepest descents' approach, which uses the following methodology:

1. Choose an initial step size, S.

2. Generate a new configuration $\boldsymbol{q}_{n+1}$ from the existing configuration $\boldsymbol{q}_n$, by applying the following, where $\dot{E}_n$ is $\partial E/\partial \boldsymbol{q}_n$:

$$\boldsymbol{q}_{n+1} = \boldsymbol{q}_n - S\dot{E}_n \tag{1.86}$$

3. Determine the energy of the new configuration. If it is less than the energy of the old, increase the step size to speed up the minimization process. If it is higher in energy, then a minimum has been overshot, so decrease the step size.

4. Overwrite the old configuration with the new one; if the system is not yet adequately minimized, return to step 2.

If a system is truly minimized, then all of the energy derivatives with respect to the coordinates should be zero, and the energy change between successive minimization steps should also be zero. Thus, minimization algorithms generally successfully terminate when some user-defined energy change or gradient tolerance is met.

Although the steepest descents method is very robust, it is not particularly efficient at reaching the minimum, often oscillating around it. This behaviour can be greatly reduced by using the previous minimization steps as a guide for the current step, in the 'conjugate gradients' method. This is similar to steepest descents, although a new configuration $\boldsymbol{q}_{n+1}$ is now determined from the following formula:

$$\boldsymbol{q}_{n+1} = \boldsymbol{q}_n + S(-\dot{\boldsymbol{E}}_n + \beta_n \dot{\boldsymbol{E}}_{n-1}) \tag{1.87}$$

The factor β_n is set to zero for the first step. For all subsequent steps, the original Fletcher-Reeves [108] algorithm calculates it as

$$\beta_n = \frac{|\dot{\boldsymbol{E}}_n|^2}{|\dot{\boldsymbol{E}}_{n-1}|^2} \tag{1.88}$$

Polak and Ribière [109] proposed an alternative algorithm, which can be more effective for some systems:

$$\beta_n = \frac{\dot{\boldsymbol{E}}_n \cdot [\dot{\boldsymbol{E}}_n - \dot{\boldsymbol{E}}_{n-1}]}{|\dot{\boldsymbol{E}}_{n-1}|^2} \tag{1.89}$$

Conjugate gradient methods generally converge much more quickly than steepest descents. However, they can occasionally get 'stuck', after which a 'restart' is required. Originally, such a restart was implemented simply by setting β_n to zero, as for the first step. However, an improved restart regimen, as suggested by Powell [110], is now more usually used.

Finally, an alternative approach to minimization, which also requires first derivatives, is known as dissipative or quenched molecular dynamics. This is very similar to the simulated annealing technique outlined earlier, except that molecular dynamics (section 1.10.1) is utilised rather than Monte Carlo.

1.9.3 Second-order methods

Second-order methods also use the curvature, i.e. the second derivatives of the energy. The first derivative of the energy at a given point $\boldsymbol{q}'$ can be approximated from that at point $\boldsymbol{q}$ by differentiation of equation 1.85, neglecting terms of third order and above:

$$\dot{E}(\boldsymbol{q}') \approx \dot{E}(\boldsymbol{q}) + (\boldsymbol{q}' - \boldsymbol{q})\ddot{E}(\boldsymbol{q}) \tag{1.90}$$

If $\boldsymbol{q}'$ corresponds to the minimum, then $\dot{E}(\boldsymbol{q}')$ is zero, and thus

$$\boldsymbol{q}' \approx \boldsymbol{q} - \frac{\dot{E}(\boldsymbol{q})}{\ddot{E}(\boldsymbol{q})} \tag{1.91}$$

Newton-Raphson minimization applies this formula directly in order to generate a new configuration, and as before, the procedure can be repeated multiple times before minimization is achieved. Note, however, that the inverse of the N-by-N second derivative matrix (the Hessian) is required. This matrix inversion is rather computationally expensive for large systems, scaling as $\mathcal{O}(N^3)$, and requiring $\mathcal{O}(N^2)$ storage. Calculation of the matrix itself is also costly, as the second derivatives are generally calculated numerically, by calculating the first derivatives for several configurations. CHARMM partially solves these problems by use of its Adopted Basis Newton-Raphson technique [10]. This constructs and uses a much smaller Hessian, which covers only the subset of the system that has moved the most in the previous minimization steps.

1.10 Determination of free energy

Although the energy of a system and its minima, even using a semi-empirical or molecular mechanics treatment, can reveal useful information, for comparison with and prediction of experimental data the entropy must also be considered. The Helmholtz free energy of a system at temperature T is given as

$$A = -k_B T \ln Q \tag{1.92}$$

where k_B is the Boltzmann constant and Q is the *partition function*, given by

$$Q = \sum_i \exp(-E_i/k_B T) \tag{1.93}$$

which in turn is a function of the energies E_i of each configuration i. Thus, it can be seen that the evaluation of the free energy is a difficult undertaking, as the number of configurations that can be adopted by a system is essentially infinite. However, many of these configurations are of such high energy (e.g. particles overlap, or bonds are far from their minimum lengths) that their contribution to Q is negligible. Several techniques are available that can be used to determine free energies by concentrating on configurations with favourable interactions, the simplest of which is perhaps molecular dynamics.

1.10.1 Molecular dynamics

In molecular dynamics, the positions of a system's particles, $\boldsymbol{r}$, are allowed to evolve in time by application of standard Newtonian dynamics. Newton's Second Law relates the force $\boldsymbol{F}$ on a body of mass m to the resultant acceleration, $\ddot{\boldsymbol{r}}$:

$$\boldsymbol{F} = m\ddot{\boldsymbol{r}} \tag{1.94}$$

The force is simply the negative of the energy gradient,

$$\boldsymbol{F} = -\frac{\partial E}{\partial \boldsymbol{r}} \tag{1.95}$$

The position of a particle at short times δt away from a reference time t can be written as a Taylor expansion:

$$\boldsymbol{r}(t+\delta t) = \boldsymbol{r}(t) + \delta t\dot{\boldsymbol{r}}(t) + \tfrac{1}{2}\delta t^2\ddot{\boldsymbol{r}}(t) + \tfrac{1}{6}\delta t^3\dddot{\boldsymbol{r}}(t) + \ldots \tag{1.96}$$

$$\boldsymbol{r}(t-\delta t) = \boldsymbol{r}(t) - \delta t\dot{\boldsymbol{r}}(t) + \tfrac{1}{2}\delta t^2\ddot{\boldsymbol{r}}(t) - \tfrac{1}{6}\delta t^3\dddot{\boldsymbol{r}}(t) + \ldots \tag{1.97}$$

By adding these two expressions together, and neglecting terms of fourth order and higher, the Verlet algorithm [111] is arrived at:

$$\boldsymbol{r}(t+\delta t) = 2\boldsymbol{r}(t) - \boldsymbol{r}(t-\delta t) + \delta t^2\ddot{\boldsymbol{r}}(t) + \mathcal{O}(\delta t^4) \tag{1.98}$$

The Verlet algorithm is easy to implement as, in addition to the current atomic coordinates and accelerations, it is only necessary to store the positions at the last time step; note that the velocities are not required. If these are needed, however (for example, to calculate kinetic energy) then they can be recovered from the following:

$$\dot{\boldsymbol{r}}(t) = \frac{\boldsymbol{r}(t+\delta t) - \boldsymbol{r}(t-\delta t)}{2\delta t} + \mathcal{O}(\delta t^3) \tag{1.99}$$

More commonly, the so-called 'leap-frog' modification is used:

$$\dot{\boldsymbol{r}}(t+\tfrac{1}{2}\delta t) = \dot{\boldsymbol{r}}(t-\tfrac{1}{2}\delta t) + \delta t\ddot{\boldsymbol{r}}(t) \tag{1.100}$$

$$\boldsymbol{r}(t+\delta t) = \boldsymbol{r}(t) + \delta t\dot{\boldsymbol{r}}(t+\tfrac{1}{2}\delta t) \tag{1.101}$$

This algorithm has the same basic storage requirements as the standard Verlet (3 vectors per particle) but has the advantage that the velocities (albeit at the mid-step)

are explicitly included, and can again be recovered at time t with the following:

$$\dot{\boldsymbol{r}}(t) = \frac{\dot{\boldsymbol{r}}(t + \frac{1}{2}\delta t) + \dot{\boldsymbol{r}}(t - \frac{1}{2}\delta t)}{2} \tag{1.102}$$

Computers can only store floating-point numbers to a limited precision, and so this interpolation method gives more accurate velocities than the standard Verlet, which calculates them from the difference of two large values (the particle positions).

The choice of integration time step, δt, is a balance between time and accuracy. Longer time steps obviously mean that fewer energy calculations, usually the most expensive part of any MM or QM/MM simulation, are required; however, they also reduce the accuracy and energy conservation of the simulation, as the forces, generally smoothly varying in time, are only known at the points of energy calculation. (So-called 'predictor-corrector' algorithms [112] can ameliorate this problem to some extent, however.)

Obviously, the time step must be shorter than the timescale of any motions that are modelled. For example, in Car-Parrinello [113] dynamics simulations, the 'motions' are of the wavefunction coefficients themselves, which are indicative of the electron density, and so a very short time step must be used. The fastest motions in most MD simulations are bond stretches and angle bends, and a very short time step must be used in order to capture these motions adequately. However, in most circumstances these fluctuations are not of interest, and in these cases a longer time step can be used if the bonds and angles are fixed during the simulation at their average values. The most common method for achieving this is known as SHAKE [114], which allows time steps of up to around 2 fs. SHAKE is an iterative algorithm, ideally suited for a system containing many constraints, and functions by adjusting the strength of each constraint until every SHAKE-constrained distance and angle at time $(t + \delta t)$ is within a user-defined tolerance of its value at a given time t.

1.10.2 Temperature control

In a physical system, the velocities of the molecules obey a Maxwell-Boltzmann distribution [76]. The average temperature T of the system is given from the velocity distribution, by generalized equipartition [115], as

$$\left\langle p_k \frac{\partial H}{\partial p_k} \right\rangle = k_B T \tag{1.103}$$

where p_k is a generalized momentum. For a system of N particles, each with momentum $\boldsymbol{p}_i$ and mass m_i, this equates to

$$\left\langle \sum_{i=1}^{N} \frac{|\boldsymbol{p}_i|^2}{m_i} \right\rangle = 2\langle K \rangle = 3Nk_B T \tag{1.104}$$

In reality, N_C constraints are placed on the system, reducing the number of degrees of freedom correspondingly, and giving the instantaneous temperature as:

$$T = \frac{1}{(3N - N_C)k_B} \sum_{i=1}^{N} \frac{|\boldsymbol{p}_i|^2}{m_i} \tag{1.105}$$

It can be seen that the instantaneous temperature of the system is proportional to the average squared particle velocity. Thus, at the start of a simulation, the system can be set to a desired temperature by assigning the particles velocities in a suitable distribution (often Gaussian), or if the simulation is started from a 0 Kelvin (minimized) structure, then the system can be 'heated' periodically by simply scaling the calculated velocities. However, simply adding kinetic energy to the system in this way will cause some of it to be redistributed as potential energy, and the system temperature will fluctuate. Thus, a number of time steps must be allowed for the system to equilibrate and the temperature to stabilise.

Even after equilibration, the temperature of a system may continue to fluctuate, largely due to errors introduced by the Verlet algorithm. The temperature can

be controlled during a 'production' dynamics run by simply scaling the calculated velocities periodically. However, it must be remembered that the aim of an MD simulation is usually to produce configurations that belong to the canonical ensemble [88] (a system in which the volume V, the temperature T, and the number of particles N are constant), and there is no rigorous proof that scaling produces such configurations; in fact, it gives rise to discontinuous dynamics [116].

An alternative to scaling the velocities is to introduce a velocity-dependent term to the force on each particle, to represent coupling of this system to an external heat bath at the desired thermostat temperature. One such approach is Berendsen's *weak coupling* algorithm [117] which has the following equation of motion:

$$\ddot{\boldsymbol{r}}_i = -\frac{1}{m_i}\frac{\partial U}{\partial \boldsymbol{r}_i} + \gamma\left(\frac{T_{eq}}{T} - 1\right)\dot{\boldsymbol{r}}_i \tag{1.106}$$

where T_{eq} is the thermostat temperature, and T the instantaneous temperature of the system. γ is a coupling constant, which can be varied to alter the fluctuations in T. Although the statistical mechanical correctness of the method has not been proved, the coupling constant γ can be varied to evaluate and control the effect of the coupling, and once a system has reached thermal equilibrium, a very weak coupling can be used to reduce the perturbation on the system to negligible levels.

A more rigorous treatment is known as Nosé-Hoover thermostatting [118–120]. This involves adding a new degree of freedom to the system to yield extended equations of motion. Although the resultant motion is non-Newtonian, the total system Hamiltonian is conserved and the derived averages of static quantities can be proved to be equal to those of the canonical ensemble.

Nosé's original technique [118] introduces a velocity scaling factor s, which is treated as an additional degree of freedom by means of an extended Lagrangian approach (which is treated in more detail in section 3.3). The variable s is assigned the potential energy $(f+1)kT_{eq}\ln s$, where f is the number of degrees of freedom in the original

system, and kinetic energy $\frac{1}{2}Q\dot{s}^2$. Q is a fictional inertia parameter for the variable s, and controls the timescale of the temperature fluctuation. The total Lagrangian of the system is then defined as

$$\mathscr{L} = \sum_i \frac{m_i}{2} s^2 \dot{\boldsymbol{r}}_i^2 - U + \frac{Q}{2}\dot{s}^2 - (f+1)kT_{eq}\ln s \tag{1.107}$$

The following equations of motion are then obtained:

$$\ddot{\boldsymbol{r}}_i = -\frac{1}{m_i s^2}\frac{\partial U}{\partial \boldsymbol{r}_i} - \frac{2\dot{s}}{s}\dot{\boldsymbol{r}}_i \tag{1.108}$$

$$Q\ddot{s} = \sum_i m_i s \dot{\boldsymbol{r}}_i^2 - \frac{(f+1)kT_{eq}}{s} \tag{1.109}$$

Hoover [119, 120] introduced a set of similar equations, which are used by CHARMM's [10] Nosé-Hoover thermostatting module, and are utilised for charge thermostatting in simulations later in this thesis. These use a scaling parameter S, with associated kinetic energy $\frac{1}{2}Q\dot{S}^2$ and potential $fkT_{eq}S$, and lead to the following simpler equations of motion:

$$\ddot{\boldsymbol{r}}_i = -\frac{1}{m_i}\frac{\partial U}{\partial \boldsymbol{r}_i} - \dot{S}\dot{\boldsymbol{r}}_i \tag{1.110}$$

$$Q\ddot{S} = \sum_i m_i \dot{\boldsymbol{r}}_i^2 - fkT_{eq} \tag{1.111}$$

1.10.3 Ensemble averages

The partition function Q_{NVT} for the canonical ensemble [88] is given in terms of the system Hamiltonian H, particle momenta $\boldsymbol{p}$ and positions $\boldsymbol{r}$:

$$Q_{NVT} = \frac{1}{N!}\frac{1}{h^{3N}}\int\int \mathrm{d}\boldsymbol{p}\mathrm{d}\boldsymbol{r}\exp\left(\frac{-H(\boldsymbol{p},\boldsymbol{r})}{k_B T}\right) \tag{1.112}$$

The Hamiltonian is given for particles of mass m as:

$$H(\boldsymbol{p}, \boldsymbol{r}) = \sum_i \frac{|\boldsymbol{p}_i|^2}{2m} + V(\boldsymbol{r}) \tag{1.113}$$

If the potential energy V is taken to be independent of the momenta, as is usually the case, then the integral in equation 1.112 is separable into position and momentum parts. Furthermore, the integration over momenta can be performed analytically, enabling the total integral to be written as:

$$Q_{NVT} = \frac{1}{N!}\left(\frac{2\pi m k_B T}{h^2}\right)^{3N/2} Z_{NVT} \tag{1.114}$$

where Z_{NVT} is the *configurational integral*:

$$Z_{NVT} = \int \mathrm{d}\boldsymbol{r} \exp\left(\frac{-V(\boldsymbol{r})}{k_B T}\right) \tag{1.115}$$

Thus, the momenta are not required for determination of the partition function. The ensemble average of the potential energy, $\langle V(\boldsymbol{r})\rangle$, is then given as:

$$\langle V(\boldsymbol{r})\rangle = \int \mathrm{d}\boldsymbol{r} V(\boldsymbol{r})\rho(\boldsymbol{r}) \tag{1.116}$$

where the probability of finding the system in position $\boldsymbol{r}$, $\rho(\boldsymbol{r})$, is given as:

$$\rho(\boldsymbol{r}) = \frac{\exp(-V(\boldsymbol{r})/k_B T)}{Z_{NVT}} \tag{1.117}$$

1.10.4 Monte Carlo simulation

It is not often practical, or even possible, to evaluate the integrals for ensemble averages (equation 1.116) analytically in $3N$ dimensions. Thus, approximate methods must be used to determine them. One such method is that of random sampling; configurations $\boldsymbol{r}$ are chosen at random, and are then weighted by their probability, $\rho(\boldsymbol{r})$. This is the

basis of the Monte Carlo method. However, such a method is remarkably inefficient for molecular systems, where the majority of configurations involve unfavourable interatomic interactions (such as strongly repulsive van der Waals forces, or elongated bonds) and thus have a very low probability, contributing little to the ensemble average. An improved method is that of importance sampling, in which configurations are chosen on the basis of their probability, and are then weighted equally. This is known as the Metropolis Monte Carlo [121] method.

Note that the determination of the configurational integral, Z_{NVT}, requires a knowledge of all degrees of freedom of the system, and so again cannot be analytically determined; thus, the absolute probability cannot be determined either. However, the *ratio* of probabilities of two states can be easily calculated. Metropolis Monte Carlo thus proceeds by generating a *Markov chain* of configurations.

Given a starting configuration a, a second configuration b is generated by applying the underlying matrix of the Markov chain, α, where $\alpha(a \rightarrow b)$ is the probability of generating state b from a. The trial configuration b is then accepted into the chain by consideration of the transition matrix π, where $\pi(a \rightarrow b) = \alpha(a \rightarrow b)P(a \rightarrow b)$ and $P(a \rightarrow b)$ is the probability of accepting the trial move. If the move is accepted, the process is repeated with the new configuration b. If it is rejected, configuration a counts twice in the Markov chain.

The matrix π must maintain the equilibrium of a system. Thus, it must obey the balance condition that the number of accepted moves that take the system away from state a must be balanced by the moves that bring it back to state a from other states. In fact, Metropolis Monte Carlo obeys the detailed balance condition [88] that the number of accepted moves from state a to a specific state b is balanced by moves from b to a, which implies that:

$$\rho(\boldsymbol{r}_a)\pi(a \rightarrow b) = \rho(\boldsymbol{r}_b)\pi(b \rightarrow a) \tag{1.118}$$

The original Metropolis scheme used a symmetric α matrix, i.e. $\alpha(a \to b) = \alpha(b \to a)$, which allows the previous equation to be written as:

$$\rho(\boldsymbol{r}_a)P(a \to b) = \rho(\boldsymbol{r}_b)P(b \to a) \tag{1.119}$$

Thus

$$\frac{P(a \to b)}{P(b \to a)} = \frac{\rho(\boldsymbol{r}_b)}{\rho(\boldsymbol{r}_a)} = \exp\left(\frac{-\Delta V(a \to b)}{k_B T}\right) \tag{1.120}$$

where $\Delta V(a \to b) = V(\boldsymbol{r}_b) - V(\boldsymbol{r}_a)$. This is achieved in the Metropolis method by defining the acceptance probability as:

$$P(a \to b) = \begin{cases} 1 & : \quad \Delta V(a \to b) \leq 0 \\ \exp\left(\dfrac{-\Delta V(a \to b)}{k_B T}\right) & : \quad \Delta V(a \to b) > 0 \end{cases} \tag{1.121}$$

The latter case is implemented in code by choosing a random number between 0 and 1, and comparing it to the calculated probability. It should be pointed out at this stage that computer-generated random numbers are usually[2] generated by a deterministic algorithm, and as such are only pseudorandom; a bad random number generator can thus lead to poor Monte Carlo simulations.

The underlying matrix α is constructed by defining a move set, i.e. a set of moves to transfer the system from one state a to another state b. One of the set of the moves is then chosen at random at each Monte Carlo step. Many types of move have been used in simulations; common moves include translations and rotations of rigid molecules, and rotations about dihedrals.

Each move in a Monte Carlo simulation changes only a small part of the system, so generally speaking a much larger number of MC moves are required to achieve the

[2]Some computer systems can use 'true' randomness in their random number generators, such as user input, network traffic, or temperature fluctuations of the hardware; however, true randomness is also undesirable, as it means a given simulation cannot be reproduced exactly.

same sampling as a molecular dynamics simulation on the same system. For example, in a system of N particles, a typical MC move moves only a single particle, while at each MD time step every particle is moved, so it might be imagined that N times as many MC moves would be required to achieve similar sampling. However, this is offset by the reduced cost of calculating the energy delta of each move, as if standard force fields [10, 11] are used, each MC move only requires the evaluation of the energy of interaction between the moved particle and the $N - 1$ other particles, i.e. it scales as $\mathcal{O}(N)$. (Each MD move would require calculation of the interactions between every pair of particles, i.e. it scales as $\mathcal{O}(N^2)$.)

1.10.5 Solvent treatments

Although some simulations are carried out in the gas phase, most systems of experimental interest (such as those studied in this thesis) are solutions — i.e. the system of interest (such as a protein or reactive species) is merely the solute, surrounded by solvent molecules. These molecules can be treated explicitly, in the same way as the solute, using molecular mechanical (section 1.3) or quantum mechanical (section 1.2) force fields. A popular and simple molecular mechanical model for water is the SPC or Simple Point Charge model [122], shown in figure 1.2, which uses point charges on the oxygen and hydrogen atoms, and a van der Waals interaction between oxygen atoms only. The magnitude of the charges, the van der Waals parameters, and the geometry of the molecule were chosen to reproduce experimental properties such as radial distribution functions [123].

The SPC model has the disadvantage that it does not match the experimental geometry of the water monomer; the H–O–H angle is a little too large, and the O–H bonds are too long. This renders the model unsuitable for direct comparison with structures from experiment or high order QM simulation. A similar three-site model, which does match the experimental geometry, is the three-site transferable

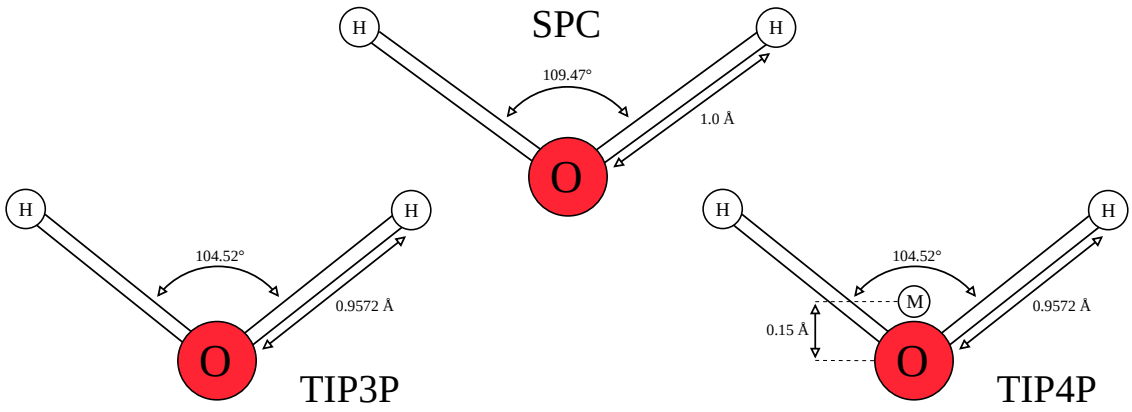

Figure 1.2: The SPC, TIP3P and TIP4P water models.

intramolecular potential, or TIP3P, model [124]. An improvement on TIP3P is the four-site TIP4P. TIP4P keeps the van der Waals interaction site on the oxygen, and the charges on the hydrogens, but moves the oxygen charge to an 'M site', situated on the H–O–H bisector, 0.15 Å towards the hydrogens.

Implicit solvent

For an accurate solvation energy, often a very large number of solvent molecules must be added. Thus, the calculation of electrostatic and van der Waals interactions between solvent molecules, and between solvent and solute, can account for a large fraction of the total computer time. To combat this problem, the solvent can be treated implicitly. Three common implicit solvent treatments are the Born [125] and Onsager [91] models, and the PCM or Polarizable Continuum Model [126].

The Born [125] method simply treats the solute as a point charge within a spherical solvent cavity; the solvent itself is treated as a region of constant dielectric. The electrostatic contribution to the solvation free energy is then simply given as the work required to uncharge the ion in vacuum and then charge it again in the solvent:

$$\Delta G_{\text{elec}} = -\frac{1}{4\pi\epsilon_0}\frac{q^2}{2a}\left(1-\frac{1}{\epsilon}\right) \tag{1.122}$$

where ϵ is the dielectric constant of the solvent, q the charge on the solute, and a the cavity radius.

The Onsager [91] method improves upon the Born by modelling the solute as a dipole, again in a spherical solvent cavity. This dipole induces a dipole in the solvent, which in turn generates an electric field (known as a 'reaction field') given by:

$$\boldsymbol{\phi}_{RF} = \frac{1}{4\pi\epsilon_0}\frac{2(\epsilon-1)}{(2\epsilon+1)a^3}\boldsymbol{\mu} \tag{1.123}$$

where $\boldsymbol{\mu}$ is the dipole moment of the solute. The reaction field interacts with the solute dipole, such that the electrostatic contribution to the solvation free energy is given as:

$$\Delta G_{\text{elec}} = -\frac{1}{2}\boldsymbol{\phi}_{RF}\cdot\boldsymbol{\mu} \tag{1.124}$$

The solute in the Onsager method does not have to be charged, and the method is thus applicable to many more systems than the Born method.

Molecules are rarely spherical, and so better results can be obtained if the shape of the cavity is determined by the van der Waals surface of the solute itself. This is the essence of the PCM method [126]. The cavity surface is divided into a number of tesserae, each of which is assigned a charge proportional to the electric field gradient generated by the solute. These point charges interact electrostatically with each other and also perturb the wavefunction of the solute within the cavity. The charges are modified by an iterative scheme until self-consistency is achieved; this then enables ΔG_{elec} to be evaluated numerically.

1.10.6 Reaction free energies via umbrella sampling

It is often desirable to know how the free energy changes during a chemical reaction. Such information can be obtained by means of techniques such as free energy perturbation [127] or thermodynamic integration [128], although in this work the

'umbrella sampling' technique first proposed by Torrie and Valleau [129, 130] is used. Knowledge of the free energy profile allows comparison with experimentally-determined properties such as equilibrium constants and reaction rates.

To describe the position of the reaction, it is often convenient to use a *reaction coordinate*. This should be chosen such that smoothly varying the reaction coordinate causes the entire reaction profile to be explored. For example, a single bond length or angle, or a ratio of two bond lengths, can be used for simple reactions, although more complex reactions may require multiple coordinates.

When the free energy is given as a continuous function of the reaction coordinate, it is known as a PMF or Potential of Mean Force [131]. The PMF for a reaction coordinate ζ is given as:

$$W(\zeta) = -k_B T \ln \rho(\zeta) \tag{1.125}$$

where the probability of finding the system with coordinate ζ is given by consideration of equation 1.117 as:

$$\rho(\zeta) = \frac{\exp(-V(\zeta)/k_B T)}{Z_{NVT}} \tag{1.126}$$

In principle, the free energy profile could be fully determined by a molecular dynamics (section 1.10.1) or Monte Carlo (section 1.10.4) simulation. However, in a reaction with even a modest variation in free energy, the sampling will be concentrated in the low energy regions, and a very long simulation is required to accumulate reasonable Boltzmann statistics for the high energy portions of the reaction profile. Umbrella sampling tackles this problem by adding a biasing function $w(\zeta)$ to the system energy $V(\zeta)$ to yield a biased energy $V^*(\zeta)$:

$$V^*(\zeta) = V(\zeta) + w(\zeta) \tag{1.127}$$

The form of the biasing function is chosen to enhance sampling in the high energy regions. (The original method [130] used a biasing function to force the distribution to

cover the configuration spaces of two systems, in order to study phase transitions, hence the name 'umbrella sampling'.) A simulation using this biased energy then obviously results in a biased probability distribution, $\rho^*(\zeta)$. Fortunately, the true unbiased probability distribution, $\rho(\zeta)$, can be recovered by use of the following relationship:

$$\rho(\zeta) = \mathrm{e}^{w(\zeta)/k_BT}\langle \mathrm{e}^{-w(\zeta)/k_BT}\rangle \rho^*(\zeta) \tag{1.128}$$

Finally, the PMF is given by:

$$\begin{aligned} W(\zeta) &= -k_BT \ln \rho(\zeta) \\ &= -k_BT \ln \rho^*(\zeta) - w(\zeta) - k_BT \ln\langle \mathrm{e}^{-w(\zeta)/k_BT}\rangle \end{aligned} \tag{1.129}$$

The choice of the biasing function determines the sampling of ζ values. To achieve perfect sampling over the entire range, the function would simply be the negative of the PMF. Obviously, this is impractical as determination of the PMF is the entire aim of an umbrella sampling simulation! Thus, an approximate function must be used. To ease the choice of the functional form and parameterization, an umbrella sampling simulation is usually split into a number of overlapping windows. In each window, the coordinate ζ is constrained near a reference value ζ_0 by using a simple harmonic biasing function:

$$w(\zeta) = \frac{1}{2}k_{\mathrm{umb}}(\zeta - \zeta_0)^2 \tag{1.130}$$

where k_{umb} is a constant which determines the width of the resulting distribution.

At the end of the simulation, a probability distribution is obtained for each window, and thus a PMF. However, the final term in equation 1.129 is not known, and should thus be treated as an arbitrary constant. Furthermore, this constant is different for each window. Thus, the overlap between the windows is utilised to smoothly join the neighbouring free energy curves, to give the PMF for the entire range of ζ. An effective method for joining windows is the Weighted Histogram Analysis Method [132, 133].

However, for simple one-dimensional surfaces, a simple averaging technique is sufficient, and was used in this work, as will be explained in section 5.3.4.

The method of using multiple windows is demonstrated in figures 1.3 and 1.4, using data from a preliminary simulation of the Menshutkin reaction (considered in more detail in Chapter 5). Each window results in a biased probability distribution as a function of the coordinate ζ, shown in figure 1.3. The ζ_0 and k_{umb} values were chosen to ensure sufficient overlap between neighbouring windows. *Relative* free energy curves were then determined for each window by use of equation 1.129, neglecting the final term. These are shown as the coloured lines in figure 1.4, the colours corresponding to those of the original probability distributions. Note that the free energy curve given by a window becomes rather unreliable at each end, where the sampling is low. Finally, each free energy curve was shifted on the y axis until they all joined up smoothly with their neighbours, and the regions of poor sampling were excluded, giving an overall PMF, shown by the black line in the figure. The quality of the final PMF is determined by several factors: the length of simulation used for each window, the number of windows, and the degree of overlap between them.

1.11 Summary

The techniques explained in this chapter all have their own strengths and weaknesses, and a combination of several of them is used in this thesis. For example, Monte Carlo and molecular dynamics can both be used for simulations in the NVT ensemble, but have different characteristics (e.g. MC does not require forces, but equally cannot yield time-dependent properties). A number of drawbacks exist, however, either in the theories or in their application to systems of interest, and possible solutions to some of these issues are outlined in later chapters.

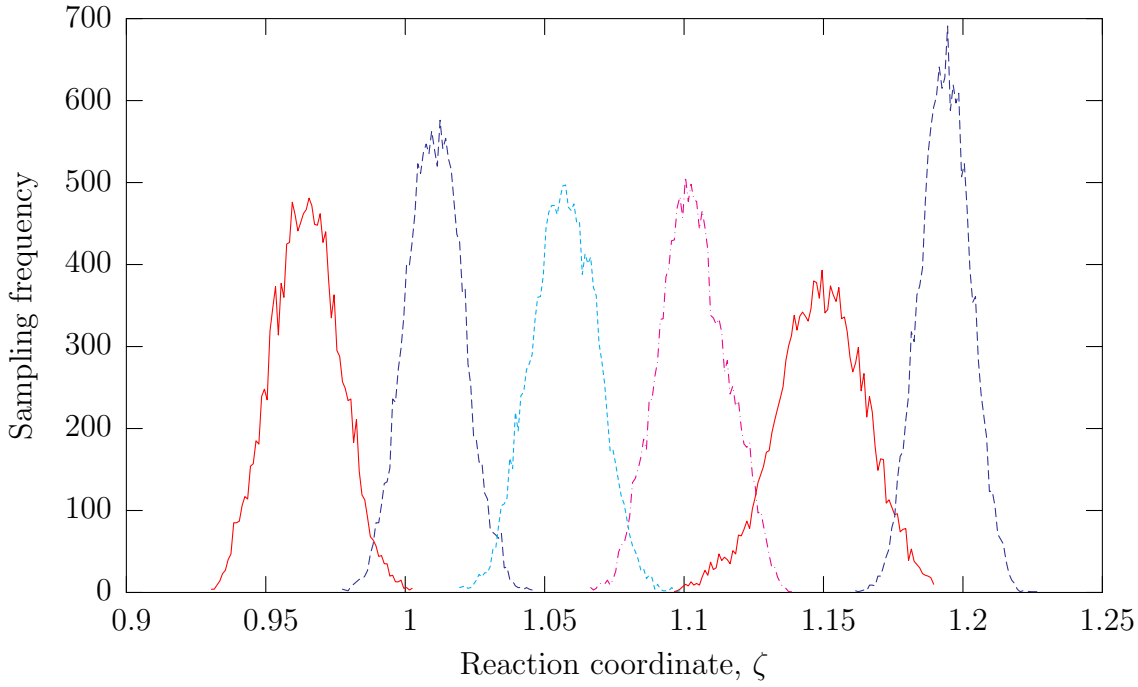

Figure 1.3: Resultant biased probability distributions from a series of umbrella sampling windows along the Menshutkin reaction, each generated using a harmonic umbrella potential (ζ_0 was varied from 0.98 to 1.18 in steps of 0.04). Samples were collected in bins 0.005 wide. ζ itself was defined as the ratio of two bond lengths.

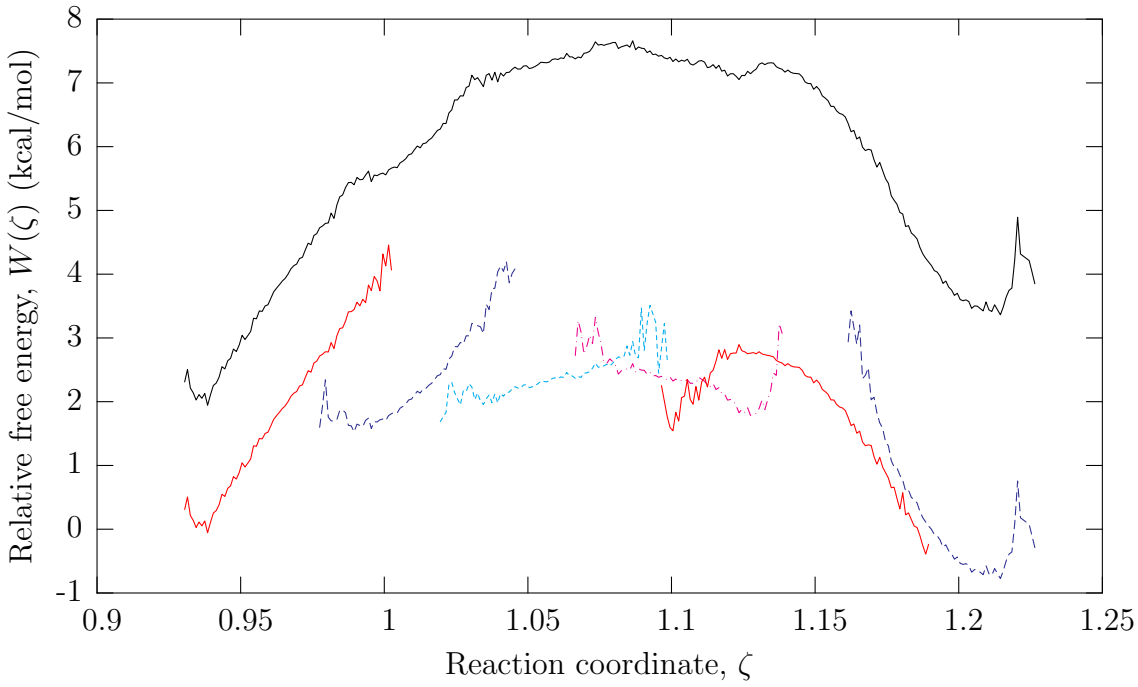

Figure 1.4: Determined relative free energies for each window shown in figure 1.3 (coloured lines) and the joined function above (black line). (The energy zero of each window is different, due to the final unknown term in equation 1.129.)

Chapter 2

Computational infrastructure

2.1 Overview

Computational chemistry makes severe demands on any computing facility. Some methods, such as Density Functional Theory or Configuration Interaction, require the evaluation of a large number of integrals in order to calculate the energy of a configuration. Other methods, such as molecular dynamics or umbrella sampling, need to consider a large number of configurations. Thus the main requirement is for raw processing power, which determines the length and accuracy (e.g. MD time step, level of QM theory) of any simulation. However, disk space and access time are also important factors, particularly for simulation of extremely large protein systems or for nondirect SCF QM [5] calculations. Finally, modern high-performance computing relies on groups of computers working in concert, and thus both the speed of the network hardware and the efficiency of the algorithms used are vital factors in determining the overall performance. In this chapter these factors are explained in more detail. The computer system built for the calculations carried out in this study is described, and shown to perform well.

2.2 Parallelisation of computational problems

The most powerful supercomputers available in the world today are all constructed as arrays of processing units linked by fast interconnects. The full power of such a system is only realised when the task to be performed is able to be split into a number of separate processes, which can each be run on a single processor (also known as a Central Processing Unit, or CPU). Ideally, each process will also not require data from any other process, so that the transmission rate or *bandwidth* of the interconnects does not become rate-limiting and so that tasks will not be forced to wait for others to complete. Obviously, certain problems 'parallelise' more easily and more completely than others - for example, so-called 'embarrassingly parallel' [134] problems simply split the problem set into N parts, and then run each part simultaneously on N processors. Such problems can be efficiently run on thousands or even millions of machines, with little or no communication required between separate processes; two examples are the SETI@home project [135] to analyse detected radio signals for signs of extra-terrestrial intelligent life, and the Cancer Screensaver [136, 137] to pre-screen possible anti-cancer, anthrax and smallpox drugs. Both of these use the otherwise wasted spare processing power of PCs in home and office environments [138]. Methods have also been developed [139] to run molecular dynamics on such massively parallel systems, and used by the Folding@Home project to model directly the folding of simple proteins [140].

For many problems in computational chemistry, however, such as direct modelling of chemical reactions, and for other challenges such as weather prediction, or the accurate simulation of nuclear weaponry, a great deal of communication is required between processes. For these applications, the most powerful supercomputers utilise custom-built processors and networks, for example IBM's experimental Blue Gene [141] system for protein folding, and the MD-GRAPE gravity-pipe hardware [142] for simulation of N-body problems.

To exploit the power of a parallel computing system fully, software must be specially written; it is generally not sufficient to take code designed for running on a single processor ('serial' code) and expect it to run more quickly on a parallel system. This can be seen if the total computational workload, W, is divided into the serial workload W_s and the parallel workload W_p:

$$W = W_s + W_p \tag{2.1}$$

and the *serial fraction*, f, is defined as

$$f = \frac{W_s}{W} = \frac{W_s}{W_s + W_p} \tag{2.2}$$

The parallel workload is so named because it can be split into smaller parts and computed by several processors simultaneously, so reducing the time required to complete the calculation. The serial workload, on the other hand, cannot be split. Thus, if each processor can process ν workload units per unit time, then the time required on N processors is

$$\tau_N = \frac{W_s}{\nu} + \frac{W_p}{\nu N} + \tau_o \tag{2.3}$$

Here τ_o incorporates any overheads of the parallel algorithm. These include inefficiencies in the design of the algorithm and time spent waiting for data to be transmitted across the network or for processors to synchronise.

The speedup of a parallel algorithm is defined as the ratio of the time taken by the algorithm on a single processor, τ_1, to that taken on N processors, τ_N:

$$S_N = \frac{\tau_1}{\tau_N} \tag{2.4}$$

Clearly an efficient parallel algorithm must try to reduce the serial fraction and τo as much as possible, so that the speedup approaches N; it has been shown that

even a small serial fraction can dramatically harm the scalability of a calculation [143] although this is strongly dependent on the nature of the problem [144].

An additional complication is that a parallel calculation may be run on a collection of processors of varying speeds. A properly-designed algorithm should be *load-balancing*, that is to say it should divide the work between the processors in proportion to their power. Ideally this balancing should also be dynamic, such that the work is redivided should one part of the parallel calculation prove to be more demanding than originally estimated.

2.3 Types of parallelism

Parallel processing machines typically consist of processing units, memory, and permanent storage. Obviously the speed of any parallel machine is dependent on the speed and number of the processing units; however, calculation speed is also affected by the speed of communication between separate processing units, between processors and memory, and between processors and permanent storage. 'Speed' is taken to cover both the latency or access time (the delay between a request being made and the data starting to flow) and the bandwidth (the rate of transfer of data).

2.3.1 Vector processors

Some parallel machines are so-called 'vector' machines. A vector machine consists of a number of execution units, each with their own memory or set of registers. They achieve parallelism by running a single instruction simultaneously on each execution unit, each unit being given different data. For example, a matrix multiplication can be 'vectorised' in this way by having each execution unit handle a single row in the matrix. Vector machines are very similar to so-called SIMD (Single Instruction Multiple Data) machines, which also act in this way. Not all SIMD machines are

vector machines, however; most modern processors, such as the AMD Athlon, achieve SIMD-like behaviour by a process known as *pipelining*, whereby instructions are split up into parts and run through a pipeline, such that a new instruction can enter the start of the pipeline while an earlier instruction is still being processed at a later part of the pipeline. (Pipelined processors tend to have separate units for integer and floating point arithmetic in order to achieve this, and may even have more than one of each to improve throughput; multiple arithmetic units may also be employed by use of explicit SIMD instructions. Recent processors also implement *hyper-threading*, which improves pipeline utilisation by allowing instructions from several different processes to be considered in the pipeline scheduling procedure.) To achieve speedup on vector machines, code generally needs to be written such that data is presented to the processor in vector or matrix form. The CHARMM [10] nonbond energy routines, for example, are available in vector versions for machines, such as those manufactured by Cray, which have vector processors.

2.3.2 Shared memory systems

In a basic SMP (symmetric multiprocessing) system, multiple identical processing units share the same memory via a single memory 'bus'. In high-end SMP systems, each processor may have a dedicated high-speed memory cache (to reduce the number of accesses required to the lower-speed main memory) and multiple memory units may be used, connected by multiple memory buses to each processor, known as a 'crossbar' architecture. (A crossbar reduces bus congestion when compared to a single-bus system, in which only one processor can access memory at a time.) In more common SMP systems, however, the single bus approach is generally used, which limits the feasible number of processors in a single machine to four or so. SMP systems are also known as 'shared memory' systems. They are typically easier to program for, as all processes in a parallel calculation have access to the same memory, and the fast memory

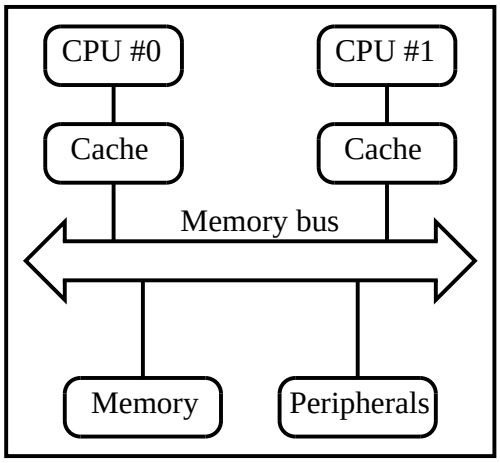

Figure 2.1: Typical architecture of an SMP system.

bus reduces communication overhead. A typical SMP machine is depicted in simplified form in figure 2.1. All communication between the processors and the memory and peripheral devices (such as network or graphics adapters) is shown as passing through a single bus. In practice, multiple buses are employed. For example, in most modern PCs the peripherals are connected to a PCI or ISA bus, which in turn is linked to the memory bus. Note also that the speed of CPU–memory communication is vitally important. Memory is significantly slower than modern CPUs, hence the need for high-speed cache. With several processors making demands on the same slow memory, this problem becomes more severe. Multiple-processor systems also introduce the problem of maintaining *cache coherence*; extra control logic must be incorporated to ensure that, if several processors access the same memory region, they do not encounter out of date information. These considerations tend to make large SMP machines extremely expensive.

2.3.3 Distributed memory systems

Distributed memory systems are those where each processor has its own memory, while individual machines are connected by a network such as Ethernet [145]. (Such a network is generally significantly less expensive than the high-speed memory buses

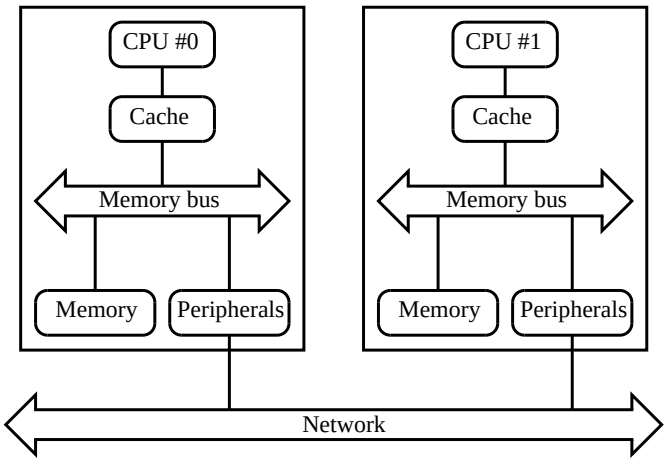

Figure 2.2: Typical architecture of a distributed memory system.

or crossbars used in SMP systems.) A typical distributed memory architecture is shown in figure 2.2. They are harder to write software for, as processes on different machines have access to different memories. Shared memory on such architectures can, however, be simulated, usually in software. Examples of so-called 'distributed shared memory' softwares are the GAMESS-US Distributed Data Interface (DDI) and the Linda [146, 147] shared tuple space[1] used by programs such as Gaussian [148]. Since, in order to access memory physically located on a different machine, a processor must request the data over the network (which is generally significantly slower than the memory bus) such systems are also known as Non-Uniform Memory Access (NUMA) machines.

2.4 Parallel programming

Parallel software on either shared or distributed memory systems performs calculations by running separate subtasks on each processor. On shared memory systems, these subtasks are frequently different threads (sometimes referred to as lightweight processes) of a single process, which share the same memory space. Packages such as

[1]Tuple space can contain zero or more tuples that can be read, modified or added to by every program that shares that space. Tuples are simply lists of values, such as pairs, triples, etc.

PC-GAMESS [149] use an explicit threaded architecture to achieve parallelism, while other packages use language extensions, such as OpenMP [150], to provide implicit threading.

On distributed memory systems, each subtask is a separate process. The processes communicate between each other either using simulated shared memory (described above) or, more usually, using a messaging library, most commonly MPI [151] (Message Passing Interface). When using MPI, processes explicitly send and receive the data they require; for example, in a molecular mechanics energy evaluation, the initial coordinates of every atom may be explicitly broadcast to all of the processes from one control process, and the calculated energies and forces gathered together by this process at the end of the calculation. The simplicity of MPI is such that it can also be used on shared memory systems; OpenMP and other threading approaches, on the other hand, cannot efficiently be used on distributed memory systems. Thus, MPI is very commonly used by parallel applications in the field of computational chemistry. Parallel MPI calculations almost always involve running exactly the same process on every processor; these are known as SPMD (Single Program Multiple Data) systems. (Multiple Program solutions — MPMD — are much less common in a message-passing environment, largely due to the increased difficulty of ensuring synchronisation and preventing deadlocks[2] when different programs are communicating with each other.)

2.5 Design and implementation of a Beowulf-class Linux cluster

The system that was designed and built by the author during the period of this study, and used for the majority of the simulations, bears many similarities to the original

[2] A *deadlock* occurs when process A waits for data from process B, while process B is at the same time waiting for data from A. This circular dependency ensures that neither process can proceed, and is usually indicative of a programming error.

'Beowulf' cluster [152] operated at NASA from 1994. A number of commodity machines (typically AMD Athlon or Intel Pentium class PCs) are connected to each other by means of a Fast Ethernet (100 Mb/sec) switch, and are used solely for computation. For ease of administration, performance, and security, these machines form a private IP network which is not connected directly to the public Internet. Instead, a single machine sits on both the public and private networks, and is used to direct jobs to compute nodes.

Historically, computational chemistry has been conducted on high-end graphics workstations and supercomputers such as the SGI Origin 2000. However, the cost to performance ratio now dramatically favours commodity PCs for many parallel applications. For example, a 195 MHz SGI Origin 2000 processor has a theoretical peak performance [153] of 390 Mflops (i.e. it can perform 390 million distinct floating-point operations per second) while a modern 933 MHz Pentium III class machine can achieve 933 Mflops, for a fraction of the cost. (It should be pointed out, however, that practical performance is rather lower than this; the LINPACK optimized benchmarks referenced above give 344 and 514 Mflops respectively for these machines.) Software availability is also a factor, some codes being available only for certain architectures; however, most of the packages used in this thesis are currently available in source code form, which makes porting to various Unix-like architectures (including Linux [154]) relatively straightforward.

The cluster, like many now in use, employs a mix of shared and distributed memory approaches. It has already been mentioned that large SMP machines are expensive due to the complications of providing a suitable memory bus; however, the additional cost of the bus for an SMP system with only two processors is compensated for by having to provide only one set of memory and peripherals. A dual processor 'box' is also significantly smaller than two single-processor machines, and requires less power and cooling. Thus, calculations run on only two processors see the full benefits of an SMP

machine, while larger calculations must rely on the slower network for at least part of their communication.

The major advantages of the cluster built during this study over a more traditional supercomputer are its low cost and good scalability. The individual machines and Ethernet networking components are inexpensive, the prices having been driven down by the business and home consumer markets. When a single machine develops problems, it can be removed from the cluster without disrupting operation on the remaining machines. Individual machines can be upgraded, or new machines can be added to the cluster, to deal with increased demand. Even workstation computers can be, and are, integrated into the cluster so that their spare computational power can be used when needed. The cluster is also heterogeneous, incorporating machines with both x86 and Alpha processors. Each machine runs its own copy of the RedHat [155] distribution of Linux, which is freely available under the GPL[3] and similar licences.

Linux and other Unix clusters are becoming increasingly common, both for high-performance parallel computing (for example, the LoBoS [156] ('Lots of Boxes on Shelves') series of machines at the NIH, or the Google [157] web search engine) and for high-availability (multiple redundancy) applications in the business arena. They are also starting to be linked together into much larger 'virtual' computers as part of a drive to make supercomputer power a resource available on demand to researchers all over the world. This is known as *grid computing* [158] and is typified by the TeraGrid [159] project, which aims to build a 20 Teraflop supercomputer by linking together smaller supercomputers physically located at four different sites around the USA.

[3]The GNU General Public Licence guarantees that the source code of programs licensed under it, and of derivatives, is always freely available.

2.5.1 Local versus network storage

Some modern clusters do not use hard disks on the compute nodes at all; instead, a system such as PXE[4] is used to boot each node from a kernel image stored on the master machine, and then the operating system is read from a network filesystem. This has the advantage that addition of new execution nodes is trivial, as any setup can be carried out on the master. Every node will also be nearly identical in configuration. The approach also has several disadvantages, however. Networking hardware is expensive, while a hard disk large enough to store an operating system is inexpensive (particularly when bearing in mind that a large fraction of a modern Linux distribution is concerned with graphical environments such as the X Window System, which are not needed on an execution node). Forcing every execution node to request all of its files from a central fileserver wastes precious bandwidth and puts more of a strain on the main server. If the server itself fails, the whole cluster becomes useless. Without disks on the nodes, any calculation output or temporary files must be written to a networked fileserver.

For these reasons, each machine in the cluster as designed has its own hard disk. This disk holds the operating system, and is also used for temporary storage during calculations. This is particularly important for quantum mechanical calculations using the GAMESS-US [79] or Gaussian 98 [148] packages, which can produce temporary files (generally containing two-electron integrals or DFT grid data) of gigabytes in size. The master node is still relied upon for shared user and password information[5], a network filesystem for storing user-built MPI binaries, access to the external network and job scheduling. However, a failure of the master node will only prevent new jobs from being started; existing jobs will continue to run to completion.

[4]The Preboot Execution Environment [160] allows a PC to boot from the network, largely for easier remote management.

[5]A consistent set of user accounts is maintained on every cluster node by using Network Information Services (or NIS+) software.

2.5.2 Factors affecting network performance

It has already been explained that, by connecting processors with a commodity network rather than a dedicated memory bus, significant cost savings are achieved but at the price of reduced performance. However, it was envisaged that the majority of calculations run on the cluster would use 8 processors or less; in fact, a significant fraction of calculations run so far are single-processor only. Obviously, for single-processor (serial) calculations the speed of the network is almost immaterial, it determining only the rate at which data can be obtained from the master node at the start of a calculation and sent back at completion.

The cluster uses standard fast Ethernet, which has a peak transfer rate of 100 Mb/sec, i.e. 100 million binary digits, or 12.5 Megabytes. (In contrast, a typical shared memory machine can communicate over its memory bus at rates greater by one to three orders of magnitude.) However, there are several caveats. Firstly, this is the transfer rate per interface card; each process on a two-processor SMP machine is thus limited to 50 Mb/sec. Secondly, this bandwidth must be shared between all nodes. For example, in a calculation involving 6 nodes, each node may want to communicate with all five of the others. Thus, the 50 Mb/sec is further reduced to 10 Mb/sec, although this can be alleviated by an appropriate choice of network topology and/or parallel algorithm. Thirdly, the operating system takes a non-zero amount of time (the latency) to send a request from the application software to the network hardware itself. This wasted time reduces the data transfer rate still further. Finally, the MPI library introduces extra inefficiencies, generally due to unneeded memory-to-memory copy operations.

The speed of the network can be increased by switching from fast Ethernet to Gigabit Ethernet, which transfers data at 1000 Mb/sec. However, at present Gigabit network hardware is far more expensive than fast Ethernet, and the aforementioned latency problems are far more pronounced. The latency problem can be ameliorated by

avoiding the overheads imposed by the operating system; many standard fast Ethernet cards support the VIA [161] standard for such OS bypass, while for high-throughput applications, Myrinet [162] offers transfer rates of around 1.28Gb/sec (2Gb/sec for the newer Myrinet-2000 offerings) with very low latency.

An inexpensive way of increasing network speed is to utilise channel bonding. Channel bonding simply employs two network cards per machine, and network traffic is divided between them. Thus, the effective bandwidth is doubled, with the only additional expense being additional fast Ethernet network cards and a larger Ethernet switch. (More than two network cards per machine can in fact be used, although again to see the full performance increase, OS bypass is required.)

Channel bonding can be implemented at a variety of levels in a typical network stack. It is most commonly done at the data link level; the OS network driver divides traffic between two or more physical network cards. This has the advantage that the channel bonding is completely transparent to applications; they simply see a network device with twice the normal bandwidth. However, the disadvantage with this approach is that it is ignorant of the network structure; thus, every machine in the cluster must have multiple network cards, and be channel-bonded. Secondly, multiple network switches must be used so that data packets from different network cards on the same machine are not confused. A less expensive approach is to channel bond at the application level; the two network cards are treated distinctly, with separate IP addresses, and the MPI library distributes traffic. This allows channel bonding to be used for part of the cluster, and the same switch can be used for all network cards. It is, however, generally less efficient than link-level bonding, and only certain MPI libraries (such as MP_Lite [163]) support it.

2.5.3 Hardware setup

The setup of the cluster is shown in schematic form in figure 2.3. Note that it was constructed in three parts. In Phase I, an experimental cluster was built by simply adding a second network card to a lab workstation, designating this machine the control node; standard PCs, only without monitors, were then set up as 'headless' compute nodes, and the control node was configured to run jobs on both these dedicated compute nodes and other lab workstations. In Phase II, banks of custom-built dual-processor machines were added in a rackmount configuration; this is far more compact than a collection of generic boxes, and allows for easier maintenance. In the third and to date final phase, additional rackmount machines were added with dual, channel-bonded, network cards. In the current configuration all compute nodes and the control node are headless; almost all administration can be performed remotely over the network. In the case of severe problems (e.g. hardware failure) a monitor and keyboard can be plugged into a misbehaving node in order to diagnose the fault.

2.5.4 Software setup

In order to efficiently distribute calculations between processors, the OpenPBS (Portable Batch System) software [164] was deployed. This receives requests from users at lab workstations for various computational resources (e.g. "10 processors for 50 hours on machines with Gaussian 98 installed"), queues and schedules the submitted jobs, and finally instructs individual compute nodes, or idle workstations, to run the calculations. OpenPBS was chosen as it is available free of charge, has freely-available source code, and is well supported by Grid middleware such as the Globus Toolkit [165]. Source code availability was vital in order to add new features to the code, as described below. Good support by Grid middleware may become important in the future if this currently lab-exclusive compute cluster is ever amalgamated with other such clusters

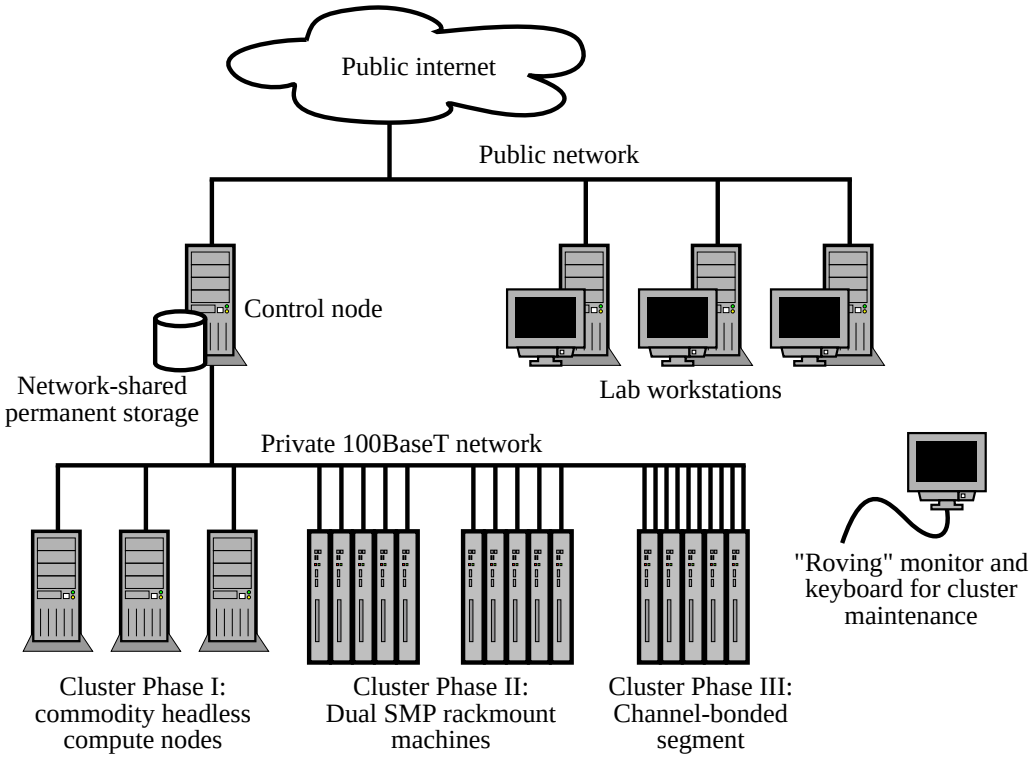

Figure 2.3: The hardware configuration of the compute cluster.

and compute resources on a departmental, university, national or even international scale.

A set of helper scripts was developed in order to copy input files from lab workstations to the control node and then on to the required compute node, and to copy outputs back at the end of a calculation. The codes run during this study are generally quite poorly load-balanced (see section 2.2) instead choosing to spread the work fairly equally between processors (although the first node in any parallel calculation is generally more loaded, as the serial work is also performed on this machine). In order to try to accommodate this on the heterogeneous cluster, the PBS software was modified to include a 'speed' attribute for each compute node. Jobs are then scheduled to run on machines in speed order, fastest first. This minimizes the probability that a calculation will be spread over a combination of fast and slow machines, and even in such an occurrence, ensures that the generally more loaded first node is a faster processor.

The cluster system is balanced between single-processor calculations and parallel jobs by 'packing' the processors requested by a job. After the nodes have been sorted by speed, processors are picked on the machines with the *smallest* number of free processors. For example, if a 2-processor SMP machine is running one single-processor job, and a second is scheduled, it is packed onto the same machine in preference to being assigned to a machine with two processors free. This keeps as many 2-processor machines free as possible, so that parallel jobs can take advantage of the faster SMP parallelism.

Parallel codes are linked against the MPICH [166,167], LAM [168], or MP_Lite [163] MPI libraries, or (in the case of GAMESS-US) use their own parallel implementations on top of the underlying TCP/IP[6] network. All of these codes are locally modified to use the Mpiexec [169] job startup system. This uses a list of compute nodes provided by PBS, and uses PBS directly to start processes on each compute node. (The conventional approach on a Beowulf cluster is to start processes via the rsh or ssh mechanism. This is significantly slower than Mpiexec, and is also open to abuse by malicious users, as it is difficult to limit rsh/ssh access to only those nodes allocated for a job.)

The modifications made to PBS, MPICH, LAM and Mpiexec during the period of this study have been given back to the free software community. In fact, recent versions of MPICH and Mpiexec now directly incorporate some of these fixes and enhancements.

[6]Ethernet communications make no guarantees as to whether information reaches its destination in the correct order, or even whether it arrives at all. TCP, or Transmission Control Protocol, makes these guarantees. IP, or Internet Protocol, underlies TCP and hides the implementation details of Ethernet, allowing information to flow across a variety of possibly disparate, interconnected networks, known as an *internet*.

2.6 Tests of the cluster performance

2.6.1 Base network performance

The theoretical peak performance of the cluster when performing parallel computations was ascertained by use of the NetPIPE program [170]. This simulates an MPI or TCP application by transmitting data in a variety of packet[7] sizes across the network, and determines the mean transfer rate. Three different MPI implementations were used: LAM, MPICH, and MP_Lite. LAM's 'usysv' transport, and MPICH's 'p4-shared' device, were used, both of which employ standard TCP networking for communication between machines, and shared memory for communication between different processors on an SMP node. MP_Lite uses TCP for all communication, whether across the network or between SMP processors, although as previously mentioned, it can channel-bond between two network cards; results from MP_Lite when operating in this mode are denoted by 'tcp $\times$ 2'.

Results are shown for SMP communication in figure 2.4, and for communication over the 100 Mbit Ethernet network in figure 2.5. For comparison, the NetPIPE software was also tested with MPICH on the more recently constructed Oxford University Linux cluster, 'Oswald' [3], which uses Myrinet networking, and the results summarised in figure 2.6. It can clearly be seen that much higher transfer rates are possible over a machine's memory bus than over an Ethernet network, even when channel bonding is employed. It should also be noted that the transfer rate is strongly dependent on the message size. Small messages are sent very inefficiently, as the latencies introduced by the operating system, and the overheads required to wrap, send, receive, and unwrap each packet, account for a significant fraction of the transfer time. For very large packets, the time required to copy them in memory can become rate-limiting. Note

[7]All data are sent across networks in chunks known as 'packets' rather than in a continuous stream. These packets contain header information to identify, for example, their network source and destination. Packets may contain or 'wrap' other packets; for example, when using MPI across an Ethernet network, the Ethernet packets contain TCP packets, which in turn contain MPI packets.

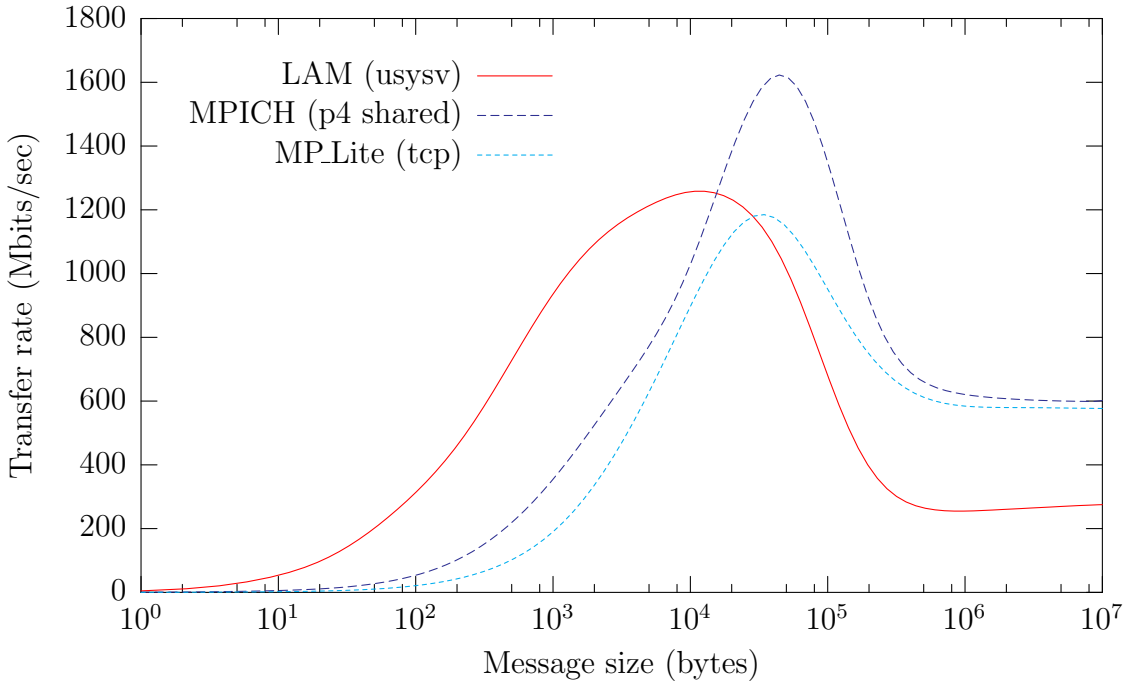

Figure 2.4: 2-processor MPI throughput on an SMP (shared memory) system, for a variety of MPI packet sizes and MPI libraries.

also that the performance of Myrinet is much better than that of 100 Mbit Ethernet, although again maximum performance is only achieved for very large packet sizes. The Ethernet performance of 'raw' TCP (i.e. without using any MPI library) is also shown in figure 2.5, for comparison. Using any of the MPI libraries incurs a small amount of overhead, due to the time required to pack MPI packets within the underlying TCP packets. Note also that channel bonding confers no benefit to performance for packet sizes smaller than about 10^3 bytes. This is because the maximum size of an *Ethernet* packet is 1500 bytes, and there is no point in using multiple Ethernet interfaces until multiple Ethernet packets become a necessity, as Ethernet packets travel at the same speed across the network, regardless of their size. (This is thus true for any type of packet encapsulated within an Ethernet packet, although large packets have to be fragmented in order to travel across Ethernet. This fragmentation and later recombination can also slow communications.)

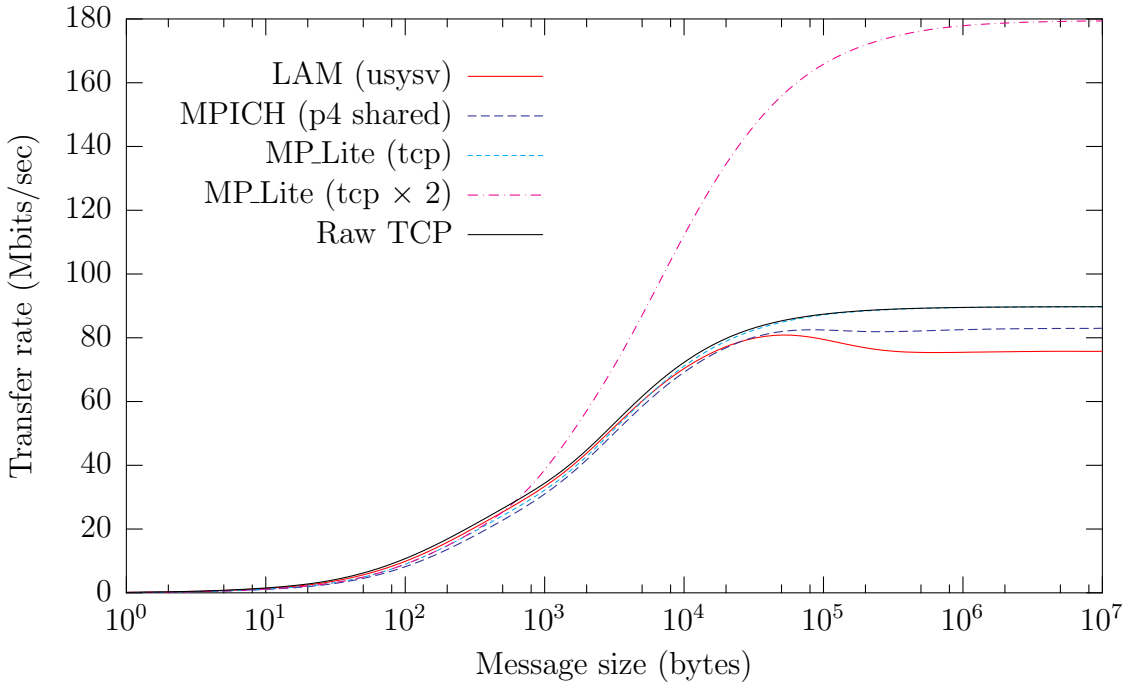

Figure 2.5: 2-processor MPI throughput across a switched 100BaseT network, for a variety of MPI packet sizes and MPI libraries.

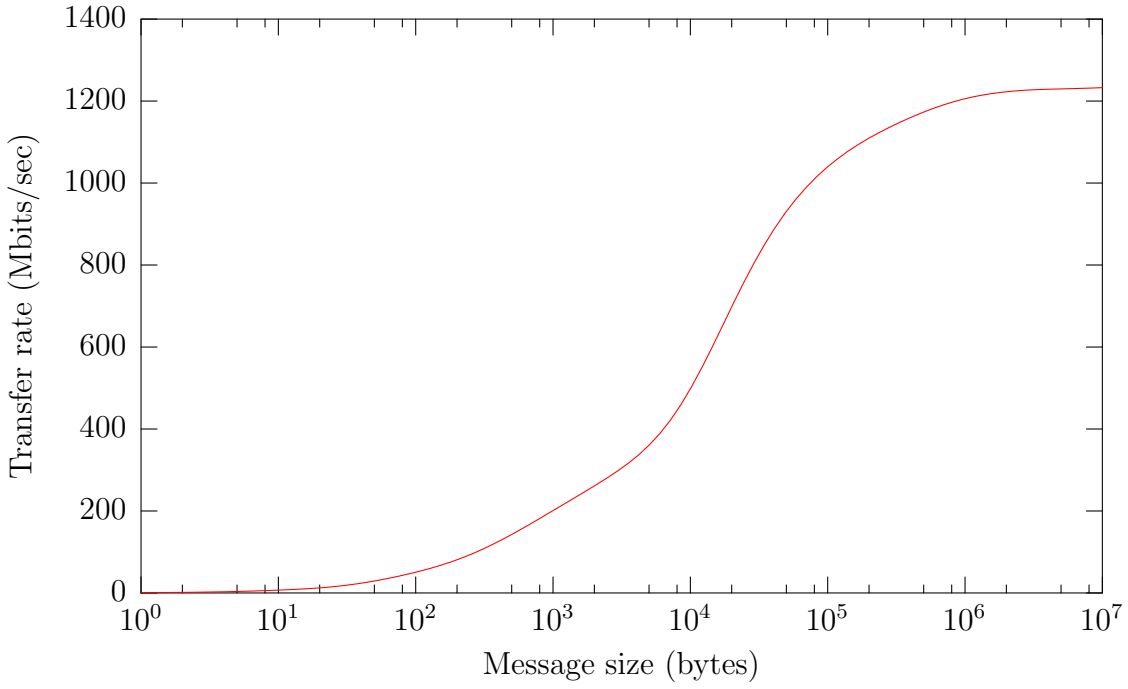

Figure 2.6: 2-processor MPI throughput across a Myrinet network, using the MPICH library, for a variety of MPI packet sizes.

2.6.2 Real-world application test cases

Actual MPI applications would be expected to employ a variety of message sizes during their communications, and thus their performance cannot be deduced from the simple NetPIPE results. Four CHARMM test cases were used to investigate this further. The first of these was a series of minimizations of a 1281-atom QM/MM system; 9 atoms were treated at the HF/6-31G* level of theory, and the remainder, which constituted the water solvating the QM region, with the CHARMM force field. This is particularly representative of the calculations carried out later in this work, as the system is the Menshutkin reaction in the solution phase, considered in Chapter 5. The second test case was used to represent more general usage on a large system, consisting of minimization and molecular dynamics of a 12053-atom enzymatic system, comprised of porcine pancreatic elastase bound to human β-casomorphin-7 [171], and solvated with 2800 TIP3P water molecules [172], using a stochastic boundary approach (see section 1.7.3). The third and fourth test cases focused on smaller systems — minimization and dynamics of a 16-molecule water cluster, and of a 256-molecule periodic water box, respectively.

Each test case was run on one or more SMP cluster nodes, each of which consisted of two Pentium III processors running at 800 MHz. Each test case was repeated for 1, 2, 4, 8 and 16 processors; the timings for the single-processor tests were then used to give parallel speedups for the multi-processor runs, by application of equation 2.4. To account for the difference in bandwidth between SMP and TCP communications, the two-processor test cases were run twice, in the first instance using two processors on one node, and in the second using one processor on each of two nodes. Test cases on more than two processors were instructed to always use two processors per node (so for example, a 16-processor calculation ran on 8 nodes). Other calculations were prohibited from running on the nodes at the same time as the parallel test cases, and each test case was run five times, in order to minimize the slowdown to the test cases

that other processes running on the nodes would cause. The test cases were run for the same parallel libraries previously used for the NetPIPE tests; however, no results were obtained for MP_Lite with channel bonding for the 16 processor case, as insufficient channel-bonded nodes were available. (Channel bonding was also not used for the 2 processor SMP test case, as this involved no communication across the network.)

Results for the QM/MM test case are shown in table 2.1, those for the enzymatic system in table 2.2, for the water cluster in table 2.3, and those for the water box in table 2.4. It can be seen from a quick inspection that the larger system enzymatic and QM/MM test cases parallelise fairly well up to 4 or 8 processors, while the smaller system cluster and periodic box test cases perform extremely poorly. To explain this behaviour, it is beneficial to understand how CHARMM and GAMESS parallelise. Put simply, the CHARMM energy routines transmit the coordinates of all atoms in the system at the start of each energy calculation from the control process to all other processes. After the energy calculation, the energies and forces from each process are combined back onto the control node. Thus the communication for an N-atom system scales as $\mathcal{O}(N)$. However, the time taken for each process to calculate the system energy may scale as $\mathcal{O}(N^2)$, depending on the force field used. Additionally, the smaller the system, the smaller the message size used by the MPI library, which the earlier NetPIPE results (see figure 2.4 and figure 2.5) have shown leads to much lower performance. Thus, the network communication overhead becomes less of an issue with larger CHARMM systems. (GAMESS also requires internode communication to gather matrix elements, but again the communication time scales more favourably than the total energy calculation time.)

It should also be noted that some of the two-processor test cases demonstrate a speedup slightly greater than 2, in disagreement with the simple model of equation 2.3, which assumes the same code is run in both serial and parallel cases. Such discrepancies are generally caused by optimization and caching. The parallel code may be better

Number of processors	MPICH (p4 shared)	LAM/MPI (usysv)	MP_Lite (tcp)	MP_Lite (tcp × 2)
2 (SMP)	2.124 ± 0.001	2.115 ± 0.002	2.023 ± 0.002	-
2 (TCP)	2.109 ± 0.001	2.100 ± 0.001	2.098 ± 10^{-4}	2.051 ± 0.003
4	3.270 ± 0.047	3.274 ± 0.004	3.024 ± 0.215	3.077 ± 0.019
8	4.811 ± 0.013	5.558 ± 0.027	3.598 ± 0.277	3.636 ± 0.008
16	3.866 ± 0.027	5.907 ± 0.036	2.804 ± 0.054	-

Table 2.1: Speedup of a series of minimizations of a 1281-atom QM/MM system, with various MPI libraries. (Each run lasted about 30 CPU minutes.)

Number of processors	MPICH (p4 shared)	LAM/MPI (usysv)	MP_Lite (tcp)	MP_Lite (tcp × 2)
2 (SMP)	1.919 ± 0.018	1.877 ± 0.020	1.920 ± 0.022	-
2 (TCP)	1.950 ± 0.014	1.864 ± 0.011	1.947 ± 0.018	1.948 ± 0.014
4	2.787 ± 0.041	3.413 ± 0.055	2.785 ± 0.028	2.773 ± 0.040
8	5.008 ± 0.038	5.861 ± 0.045	5.690 ± 0.046	5.901 ± 0.066
16	5.078 ± 0.035	8.963 ± 0.060	8.012 ± 0.059	-

Table 2.2: Speedup of MM minimization and dynamics of a 12053-atom enzymatic system, with various MPI libraries. (Each run lasted about 4 CPU hours.)

optimized than the corresponding serial code by design of the algorithm, at compile time when the Fortran or C code is converted to machine code, or at run time when the code is pipelined by the CPU. A calculation run on two processors rather than one also benefits from twice the total amount of on-CPU cache; put another way, the problem size per processor is smaller, such that the data may fit entirely into the cache, or at least the number of cache 'misses' is reduced.

The performance of MPICH and LAM is similar for the systems tested, although LAM generally performs better whenever SMP communication is required. This is consistent with the NetPIPE results shown in figure 2.4. MP_Lite, however, tends to perform rather poorly in comparison, probably because it relies on TCP for all communication, which is less than ideal for intranode data transfer. Channel bonding gives little benefit (or is even detrimental to performance) for the cluster and box test

Number of processors	MPICH (p4 shared)	LAM/MPI (usysv)	MP_Lite (tcp)	MP_Lite (tcp × 2)
2 (SMP)	1.082 ± 0.005	2.077 ± 0.009	0.560 ± 0.007	-
2 (TCP)	0.298 ± 0.002	0.291 ± 0.004	0.132 ± 0.016	0.116 ± 0.006
4	0.070 ± 10^{-4}	0.185 ± 10^{-4}	0.017 ± 10^{-4}	0.016 ± 10^{-4}
8	0.006 ± 10^{-5}	0.085 ± 10^{-4}	0.040 ± 10^{-4}	0.040 ± 10^{-4}
16	0.002 ± 10^{-6}	0.040 ± 10^{-5}	0.018 ± 10^{-5}	-

Table 2.3: Speedup of MM minimization and dynamics of a 16-molecule water cluster, with various MPI libraries. (Each run lasted about 5 CPU minutes.)

Number of processors	MPICH (p4 shared)	LAM/MPI (usysv)	MP_Lite (tcp)	MP_Lite (tcp × 2)
2 (SMP)	1.994 ± 0.010	2.021 ± 0.013	1.891 ± 0.008	-
2 (TCP)	1.774 ± 0.013	1.751 ± 0.012	1.258 ± 0.030	1.295 ± 0.072
4	0.705 ± 0.009	2.192 ± 0.010	0.675 ± 0.010	0.618 ± 0.073
8	0.439 ± 0.018	1.852 ± 0.009	1.140 ± 0.009	1.142 ± 0.009
16	0.124 ± 0.001	1.253 ± 0.007	0.648 ± 0.005	-

Table 2.4: Speedup of MM minimization and dynamics of a 256 water molecule periodic box, with various MPI libraries. (Each run lasted about 5 CPU minutes.)

cases, as the network packets used are so small. For the larger systems, however, it does begin to confer some small benefits.

No parallel library gives a speedup on 16 processors even close to 16, even for the large enzymatic system. (It is expected, however, that larger MM systems would parallelise more efficiently.) This is no doubt due in part to overloading of the network, but the major cause is likely to be inefficiencies in the CHARMM parallel algorithm.

2.7 Summary

In order to tackle modern computational chemistry problems efficiently, advantage must be taken of parallelism. There are many methods for running parallel calculations, but inexpensive commodity Linux clusters are becoming an increasingly viable

alternative to traditional supercomputers. As part of this work, such a cluster was designed and constructed, and has been shown to perform well for certain classes of problems on up to 8 processors. It has additionally been successfully used by other researchers in the group for energy and dynamics simulations of systems with MM, QM, or hybrid QM/MM force fields [20, 173–175].

Many computational chemistry codes are still not parallelised, and so a large proportion of calculations would be expected to run on a single processor. For example, the MOPAC code interfaced to CHARMM for semi-empirical QM/MM is not parallelised at all, and so a semi-empirical QM/MM calculation on multiple processors is rather inefficient, as only the MM part of the calculation experiences a parallel speedup. Again, Linux clusters excel here, as money is not wasted on a needlessly high bandwidth interconnect.

For this work, most of the simulations that will be described in later chapters were run serially, largely due to limitations in the parallel algorithms available. On the other hand, for QM/MM simulations using *ab initio* codes such as GAMESS-US, which are well parallelised, the LAM MPI library was almost exclusively used.

The PBS system used in this configuration allows the integration of otherwise unused desktop machines into the compute cluster, and readies the computing resources for possible future integration into a computing grid.

Linux clusters are not a panacea. Commodity PC and networking hardware works well only for certain classes of problem; some problem types, however, require very fast communications, and can actually demonstrate speedups of less than unity on a Linux cluster. For these, traditional shared-memory supercomputers are generally required, although commodity SMP machines may perform adequately for parallelisation across a small number of processors. In a future grid, it may well be possible to 'purchase' CPU time, memory, and other resources on such a supercomputer in return for similar resources on a Linux cluster, and research in this area is ongoing [165].

Chapter 3

Implementation of fluctuating charge force fields

3.1 Introduction

As previously discussed, microscopic systems are generally modelled by application of a molecular mechanics force field, or by solution of the quantum mechanical wavefunction. Molecular mechanics methods have the advantage that they are computationally inexpensive, and can therefore be applied to very large systems and run for long timescales. By suitable parameterization from both experimental data and QM calculation, they can be and are applied to a wide range of systems of biochemical interest. QM methods, on the other hand, require far less parameterization (dependent on the choice of *ab initio*, DFT or semi-empirical method) but can yield vital information on electronic properties and bond breaking or forming processes.

Hybrid QM/MM methods seek to combine elements of both QM and MM approaches in order to balance computational cost against accuracy. It must be remembered, however, that the two techniques can differ dramatically in their level of approximation. A number of approaches seek to smooth the join between the QM and MM regions, generally by using an intermediate method in the buffer region between

the two systems. For example, the 3-layer ONIOM [176, 177] method typically uses a semi-empirical QM potential in the region between the *ab initio* and MM systems, while other methods mix DFT and semi-empirical treatments in the QM Hamiltonian [178], or use a frozen density approximation in the buffer between the DFT region and the MM system [179].

Another approach is to improve the accuracy of the MM potential. Conventional MM methods use point charges, usually centred on the nuclei; these point charges are fixed, their values being determined by a best fit to experimental and/or QM electrostatic interactions. Such fixed point charge models are unable to polarize in response to changes in the environment, and as such lack transferability. For example, the dipole moment of water changes from 1.85 D [180] in the gas phase to 2.6 D in solid ice [181, 182], while in solution it is believed to be between 2.5 and 2.7 D [183–188]; a fixed-charge model, however, cannot change its dipole moments and is therefore generally parameterized to reproduce the bulk solvent behaviour only. This can lead to inaccuracies when the model is used to simulate non-bulk systems, such as the solvation of highly-charged species.

3.1.1 Dipole polarizable methods

The deficiencies of fixed-charge models can be partially overcome by allowing for polarization in some approximate way. A variety of methods have been proposed by researchers; several of the methods can be described as 'dipole polarizable', in that they retain the fixed charges of conventional solvent models, but add point polarizabilities [25–29]. Some of these models reproduce experimental properties to a high degree of accuracy in both the gas and solution phase. However, their disadvantage is that while charge-charge interactions are simply described by Coulomb's law, addition of multiple dipoles necessitates the calculation of charge-dipole and dipole-dipole interactions. The increased number of interactions, which are at any rate of a more complex functional

form than Coulomb's law, results in much increased computational expense; one such method has been reported to increase the cost by a factor of 4 [26].

3.1.2 Electronegativity equalisation approaches

An alternative class of polarizable force fields does not add any extra charge sites or dipoles. Instead, polarization is modelled by allowing the values of the charges themselves to alter in response to their environment. These generally apply the 'electronegativity equalisation' approach proposed by Sanderson [31,32], which suggests that when a molecule is formed from a set of atoms or other functional groups, all of which have their own electronegativities, electron transfer will occur such that the electronegativities of all of the groups become equal. Parr *et al.* [9] have subsequently shown from Density Functional theory that the Mulliken electronegativity [189,190], χ, is simply the negative of the chemical potential of an atom's electron cloud, μ, which itself is given as the derivative of the energy E with respect to the number of electrons N, at a constant external potential ν:

$$\mu = \left(\frac{\partial E}{\partial N}\right)_{\nu} = -\chi \tag{3.1}$$

Thus, the charge transfer involved in electronegativity equalisation can be alternatively viewed as occurring from regions of low chemical potential to those of high potential, such that μ is equalised for all atoms, in an analogous way to the equalisation of chemical potentials of mixtures in macroscopic chemistry.

Many methods for charge determination, utilising either full or partial electronegativity equalisation, have been proposed [191–194]. The Fluctuating Charge (FlucQ) model [30, 33–35] is one such technique, and is considered in further detail in this work.

3.2 Determination of polarization energy

The Fluctuating Charge model begins by considering the Taylor expansion of the energy of an atom A as a function of its partial charge, Q_A:

$$E_A(Q_A) = E_A(0) + Q_A \left(\frac{\partial E}{\partial Q_A}\right)_{Q_A=0} + \frac{1}{2}Q_A^2 \left(\frac{\partial^2 E}{\partial Q_A^2}\right)_{Q_A=0} + \dots \tag{3.2}$$

For ease of parameterization, terms of higher than second order are ignored. The response of atomic energy to polarization can thus be determined if the first and second derivatives of the energy with respect to the partial charge are known. These are treated as adjustable parameters of the model, and are written as:

$$E_A(Q_A) = E_A(0) + \tilde{\chi}_A^0 Q_A + \frac{1}{2}J_{AA}^0 Q_A^2 \tag{3.3}$$

Using this formula, the ionisation energy (IE) and electron affinity (EA) can be written in terms of the newly-introduced parameters:

$$E_A(+1) = E_A(0) + \tilde{\chi}_A^0 + \frac{1}{2}J_{AA}^0 \tag{3.4}$$

$$E_A(-1) = E_A(0) - \tilde{\chi}_A^0 + \frac{1}{2}J_{AA}^0 \tag{3.5}$$

$$IE = E_A(+1) - E_A(0) = \tilde{\chi}_A^0 + \frac{1}{2}J_{AA}^0 \tag{3.6}$$

$$EA = E_A(0) - E_A(-1) = \tilde{\chi}_A^0 - \frac{1}{2}J_{AA}^0 \tag{3.7}$$

Thus it can be seen that

$$\tilde{\chi}_A^0 = \frac{1}{2}(IE + EA) \tag{3.8}$$

$$J_{AA}^0 = (IE - EA) \tag{3.9}$$

i.e. that $\tilde{\chi}_A^0$ is the Mulliken electronegativity [189, 190] of the uncharged atom, while J_{AA}^0 is a measure of its hardness [9]. The *instantaneous* (i.e. charge-dependent) electronegativity is then defined as:

$$\tilde{\chi}_A = \frac{\partial E}{\partial Q_A} \tag{3.10}$$

Observe that this equation is similar to equation 3.1, where the relation $Q = Z - N$ has been applied, with N the number of electrons, and Z the atomic number of the atom. Since the total energy E is a sum of individual energy terms (equation 1.54), the electronegativity will also be a sum, including contributions from all energy terms that have a charge dependence, such as the QM/MM energy.

In a system containing multiple partial charges, the minimum energy configuration corresponds to all atoms having the same instantaneous electronegativity. Additionally, a constraint must be placed upon the system in order to maintain the total charge. For example, in an uncharged system, charge transfer could be allowed without restriction between molecules, such that the following applies to atoms α in molecules i:

$$\sum_{i=1}^{N_{molec}} \sum_{\alpha=1}^{N_{atom}} Q_{i\alpha} = 0 \tag{3.11}$$

Alternatively, charge transfer could be allowed only within each molecule, such that the following is true for all molecules i:

$$\sum_{\alpha=1}^{N_{atom}} Q_{i\alpha} = 0 \tag{3.12}$$

It should be noted that the Fluctuating Charge model places no kinetic restrictions on charge transfer. Thus charge transfer can occur over unrealistically large distances, and the polarizability of large systems can be overestimated. In this work, therefore,

intermolecular charge transfer is not permitted and the constraints in equation 3.12 are applied.

The energy of an uncharged atom in equation 3.2, $E_A(0)$, is not determined by the FlucQ method. It is, however, a constant, and can thus be set to an arbitrary value in any calculation of relative energy. It is convenient to set this such that the polarization energy of the gas phase monomer is zero; the electrostatic energy of a system containing N_{molec} molecules, each with N_{atom} atoms, is then given as:

$$E_{elec} = \sum_{i=1}^{N_{molec}} \sum_{\alpha=1}^{N_{atom}} \left(\tilde{\chi}^0_{i\alpha} Q_{i\alpha} + \tfrac{1}{2} J^0_{\alpha\alpha} Q^2_{i\alpha} \right) + \sum_{i\alpha<j\beta} J_{\alpha\beta}(\boldsymbol{r}_{i\alpha,j\beta}) Q_{i\alpha} Q_{j\beta} - N_{molec} E_{gp} \quad (3.13)$$

where E_{gp} is the electrostatic energy of a single isolated gas phase molecule, and $J_{\alpha\beta}(\boldsymbol{r}_{i\alpha,j\beta})$ is a term for the electrostatic interaction between two charges. For intermolecular charge interactions, this is the standard Coulomb term, $1/r$, but screening must be taken into effect for intramolecular interactions, as will be seen in section 3.9.

It should also be pointed out that this energy contribution will give rise to an electronegativity contribution, incorporating self-energy terms and both intramolecular and intermolecular interactions, given as:

$$\frac{\partial E_{elec}}{\partial Q_{i\alpha}} = \tilde{\chi}^0_{i\alpha} + J^0_{\alpha\alpha} Q_{i\alpha} + \sum_{i\alpha<j\beta} J_{\alpha\beta}(\boldsymbol{r}_{i\alpha,j\beta}) Q_{j\beta} \quad (3.14)$$

3.3 Molecular dynamics via extended Lagrangian methods

Direct solution of the FlucQ equations leads to a set of coupled linear equations. Solution of these equations to yield the minimum energy charges for each configuration is costly, as it requires a matrix inversion operation for all but the simplest systems.

This dramatically increases the time required to calculate the energy of the system, however [28]. An alternative approach is to perform a full optimization at the start of a calculation, and then to treat the charges as dynamical variables in the same way as the nuclear degrees of freedom, propagating them with Newtonian dynamics by use of an extended Lagrangian formalism. Such a technique has previously been used for propagation of the Kohn-Sham density matrix (rather than performing a full SCF at each step) in Car-Parrinello molecular dynamics [113], and for Nosé's thermodynamically rigorous thermostat [118].

In Lagrangian mechanics [195] the Lagrangian, $\mathscr{L}$, of the system is defined as

$$\mathscr{L} = T - U \tag{3.15}$$

where T is the kinetic and U the potential energy. The Lagrangian equation of motion of some property x is defined as

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \mathscr{L}}{\partial \dot{x}}\right) = \frac{\partial \mathscr{L}}{\partial x} \tag{3.16}$$

In *extended* Lagrangian mechanics the usual kinetic and potential energies of the nuclear (or Cartesian) degrees of freedom are added to by defining kinetic and potential energies for other dynamical variables [196]. The kinetic energy, T_x, of the property x is given by

$$T_x = \frac{1}{2} M_x \dot{x}^2 \tag{3.17}$$

Note that a fictional mass associated with the degree of freedom, M_x, has been introduced. This mass serves to determine the balance between the time fluctuations of the property x and those of the nuclear degrees of freedom.

In the case of the FlucQ model, a fictional charge mass term, M_Q, is used, giving the system's Lagrangian as

$$\mathscr{L} = \sum_{i=1}^{N_{molec}} \sum_{\alpha=1}^{N_{atom}} \frac{1}{2} m_{i\alpha} \dot{\boldsymbol{r}}_{i\alpha}^2 + \sum_{i=1}^{N_{molec}} \sum_{\alpha=1}^{N_{atom}} \frac{1}{2} M_Q \dot{Q}_{i\alpha}^2 - U - \sum_{i=1}^{N_{molec}} \lambda_i \sum_{\alpha=1}^{N_{atom}} Q_{i\alpha} \tag{3.18}$$

This includes the nuclear kinetic energy, the charge kinetic energy, the total potential energy (including the self-energy of the charges) and a number of constraints to maintain the charge neutrality of the system, as per equation 3.12, each characterised by a Lagrange multiplier [107], λ_i.

Thus, by application of equation 3.16 for the Cartesian coordinates, $\boldsymbol{r}_{i\alpha}$, the standard Newtonian equation of motion is obtained:

$$m_{i\alpha} \ddot{\boldsymbol{r}}_{i\alpha} = -\frac{\partial U}{\partial \boldsymbol{r}_{i\alpha}} \tag{3.19}$$

Solution for the charge coordinates, $Q_{i\alpha}$, yields the following:

$$M_Q \ddot{Q}_{i\alpha} = -\frac{\partial U}{\partial Q_{i\alpha}} - \lambda_i = \tilde{\chi}_{i\alpha} - \lambda_i \tag{3.20}$$

In order to solve the constraint imposed by equation 3.12, it should be noted that, for all i, the constraint implies:

$$\sum_{\alpha=1}^{N_{atom}} \ddot{Q}_{i\alpha} = 0 \tag{3.21}$$

Substitution of equation 3.20 into the above then yields

$$M_Q \sum_{\alpha=1}^{N_{atom}} \ddot{Q}_{i\alpha} = -\sum_{\alpha=1}^{N_{atom}} \tilde{\chi}_{i\alpha} - N_{atom} \lambda_i = 0 \tag{3.22}$$

and thus

$$\lambda_i = -\frac{1}{N_{atom}} \sum_{\alpha=1}^{N_{atom}} \tilde{\chi}_{i\alpha} \tag{3.23}$$

Finally, by substitution for λ_i into equation 3.20, the following equation of charge motion is obtained:

$$M_Q \ddot{Q}_{i\alpha} = -\frac{1}{N_{atom}} \sum_{\beta=1}^{N_{atom}} (\tilde{\chi}_{i\alpha} - \tilde{\chi}_{i\beta}) \tag{3.24}$$

It can thus be seen that the charge acceleration of a given atom, $\ddot{Q}_{i\alpha}$, is given by the difference between the instantaneous electronegativity of that atom, $\tilde{\chi}_{i\alpha}$, (as defined by equation 3.10) and the average of the $\tilde{\chi}_{i\beta}$ for all atoms β in the molecule. Thus, the charge forces act to equalise the electronegativities of all charge sites within the molecule.

If, on the other hand, intermolecular charge transfer is allowed, as per equation 3.11, then the resulting equation of charge motion is

$$M_Q \ddot{Q}_{i\alpha} = -\frac{1}{N_{molec} N_{atom}} \sum_{j=1}^{N_{molec}} \sum_{\beta=1}^{N_{atom}} (\tilde{\chi}_{i\alpha} - \tilde{\chi}_{j\beta}) \tag{3.25}$$

(In this case, the electronegativities of all charge sites in the *entire system* are equalised by the charge forces.)

With these expressions for the charge accelerations $\ddot{Q}_{i\alpha}$, the charges can be propagated during a dynamics simulation by standard Verlet; thus equation 1.98 becomes

$$Q_{i\alpha}(t + \delta t) = 2Q_{i\alpha}(t) - Q_{i\alpha}(t - \delta t) + \delta t^2 \ddot{Q}_{i\alpha}(t) \tag{3.26}$$

while the charge velocity can be approximated as

$$\dot{Q}_{i\alpha}(t) = \frac{Q_{i\alpha}(t + \delta t) - Q_{i\alpha}(t - \delta t)}{2\delta t} \tag{3.27}$$

The charges also have an instantaneous 'temperature' defined by their velocities. By consideration of equation 1.105 this temperature can be written as

$$T = \frac{1}{(N - N_C)k_B} \sum_{i=1}^{N_{molec}} \sum_{\alpha=1}^{N_{atom}} M_Q \dot{Q}_{i\alpha}^2 \tag{3.28}$$

Here N is the total number of fluctuating charges, and N_C the number of charge constraints, which is equal to N_{molec} when no intermolecular charge transfer is allowed, as one constraint is applied to each molecule (equation 3.12).

3.4 Charge minimization

By propagating the charges via modified Newtonian-like dynamics, they can be made to respond to changes in their environment. However, in order for these charges to be representative, they should be in their minimum energy configuration before dynamics is started. (Any other configuration would invariably lead to a high charge temperature, and although thermostatting could be used to force this temperature down to an acceptable level, equilibration would take more time steps.) This is exactly analogous to minimizing the Cartesian degrees of freedom before running conventional dynamics.

One approach is to lower the charge energy by dissipative Langevin dynamics. The Langevin equation of motion differs from the Newtonian in that the average effects of the environment are added; a particle experiences both a frictional drag proportional to its velocity through the medium, and stochastic forces due to collisions with other particles:

$$\dot{\boldsymbol{p}}_{i\alpha}(t) = \boldsymbol{F}(t) - \zeta \boldsymbol{p}_{i\alpha}(t) + \breve{\boldsymbol{p}}_{i\alpha}(t) \tag{3.29}$$

Here the momentum $\boldsymbol{p}_{i\alpha}$ of particle $i\alpha$ is affected by the force $\boldsymbol{F}$ (as in Newtonian dynamics), the friction constant ζ, and random forces represented by $\breve{\boldsymbol{p}}_{i\alpha}$.

The method used in this work is an application of the equations originally developed by Ermak [86, 197], which are Verlet-like, but aim to model Brownian-type dynamics.

When applied in charge space, they take the form:

$$Q(t+\delta t) = Q(t) + c_1 \delta t \dot{Q}(t) + c_2 \delta t^2 \ddot{Q}(t) + \delta Q^G \tag{3.30}$$

$$\dot{Q}(t+\delta t) = c_0 \dot{Q}(t) + (c_1 - c_2)\delta t \ddot{Q}(t) + c_2 \delta t \ddot{Q}(t+\delta t) + \delta \dot{Q}^G \tag{3.31}$$

In this equation δQ^G and $\delta \dot{Q}^G$ represent random forces, the analogues of the random force $\breve{\boldsymbol{p}}_{i\alpha}$ in equation 3.29. Such variables are generally random variables with a Gaussian distribution, having a zero mean and a variance proportional to the simulation temperature [198]. In dissipative dynamics, the desired temperature is 0 K, and thus for this work both of these variables are taken to be numerically zero. The coefficients c_0, c_1 and c_2 are defined as

$$c_0 = \mathrm{e}^{-\zeta \delta t} \tag{3.32}$$

$$c_1 = (\zeta \delta t)^{-1}(1 - c_0) \tag{3.33}$$

$$c_2 = (\zeta \delta t)^{-1}(1 - c_1) \tag{3.34}$$

Charge minimization was carried out in this work by choosing a friction constant ζ and a time step δt, and then carrying out dissipative Langevin dynamics on the charge degrees of freedom, keeping the Cartesian coordinates fixed. This process was terminated when the charge temperature dropped to below a given threshold, or a maximum number of iterations was exceeded.

An alternative method, since charge forces are available, is to simply treat the charges as extra degrees of freedom of the system, such that the energy is defined as a function of both the Cartesian coordinates and the charges. Then any of the standard CHARMM minimization algorithms described in section 1.9 can be used in order to obtain the minimum energy of the system, by minimizing the Cartesian and charge coordinates simultaneously.

3.5 Thermostatting

To maintain the charge temperature of the system, a Nosé-Hoover or Berendsen thermostat, or simple charge velocity scaling (as described in section 1.10.2) can be used. Such a thermostat is independent of any applied to the nuclear degrees of freedom. The forms of the equations of motion match those in equations 1.106 and 1.110, with the obvious difference that the dynamical parameter is not a particle's position $\boldsymbol{r}_{i\alpha}$ but its charge $Q_{i\alpha}$. For example, by comparing equation 3.24, the unperturbed charge motion equation, with the Berendsen equation of motion (equation 1.106), the following equation of motion for thermostatted charges is obtained:

$$\ddot{Q}_{i\alpha} = -\frac{1}{N_{atom} M_Q} \sum_{\beta=1}^{N_{atom}} (\tilde{\chi}_{i\alpha} - \tilde{\chi}_{i\beta}) + \gamma \left(\frac{T_{Qeq}}{T_Q} - 1 \right) \dot{Q}_{i\alpha} \tag{3.35}$$

The charge acceleration $\ddot{Q}_{i\alpha}$ is thus dependent on the charge velocity $\dot{Q}_{i\alpha}$. By writing the charge acceleration as a sum of a velocity-independent and a velocity-dependent term,

$$\ddot{Q}_{i\alpha} = A + B\dot{Q}_{i\alpha} \tag{3.36}$$

and by using the definition of $\dot{Q}_{i\alpha}$ in equation 3.27, equation 3.26 then becomes:

$$Q_{i\alpha}(t+\delta t) = 2Q_{i\alpha}(t) - Q_{i\alpha}(t-\delta t) + \delta t^2 A + \frac{\delta t}{2} B \left[Q_{i\alpha}(t+\delta t) - Q_{i\alpha}(t-\delta t) \right] \tag{3.37}$$

By collecting terms, a modified Verlet equation is obtained:

$$Q_{i\alpha}(t+\delta t) = \frac{2Q_{i\alpha}(t) - (1 + \frac{1}{2}\delta t B) Q_{i\alpha}(t-\delta t) + \delta t^2 A}{1 - \frac{1}{2}\delta t B} \tag{3.38}$$

This reduces to the standard Verlet, equation 3.26, in the limiting case where the velocity-dependent term B approaches zero.

The term B is generally a function of the instantaneous temperature, T. However, the temperature is not known when the Verlet equation is applied; thus, an iterative

procedure must be used in order to propagate the charges and apply the thermostat. The basic procedure is as follows:

1. Using an estimate for the instantaneous temperature (e.g. from the last dynamics step) calculate B and thus $\ddot{Q}_{i\alpha}$.

2. Calculate the new charges $Q_{i\alpha}(t + \delta t)$ with equation 3.38.

3. Calculate the new charge velocities $\dot{Q}_{i\alpha}$ using equation 3.27.

4. Derive the current charge temperature from equation 3.28.

5. Compare the derived charge temperature with the initial estimate. If convergence has not been achieved, use the derived temperature as the new guess, and return to step 1.

The procedure for Nosé-Hoover thermostatting is identical, except that the equation of motion (equation 3.35) is derived from the Nosé-Hoover expression, equation 1.110, rather than the Berendsen.

3.6 Monte Carlo implementation

A FlucQ Monte Carlo move is implemented following Martin *et al.* [199–201] by adding a small amount of charge, δq, to one charge site j, and removing a compensatory amount of charge from all other charge sites in the molecule α, to maintain charge neutrality:

$$Q_{i\alpha} = \begin{cases} Q_{i\alpha} + \delta q & : \quad i = j \\ Q_{i\alpha} - \delta q/(N_{atom} - 1) & : \quad i \neq j \end{cases} \tag{3.39}$$

δq is chosen from the linear random distribution $[-Q_{max} : Q_{max}]$ where Q_{max} is a user-specified parameter, which is used to control the acceptance rate of MC/FQ moves.

3.7 Incorporation into QM/MM models

In order to link the FlucQ model with QM/MM, the contribution from the QM/MM interaction to the electronegativities must be accounted for — i.e. the derivative of the QM/MM energy with respect to MM charge is required. This turns out to be fairly straightforward, as the expressions for the QM/MM energy (equations 1.58 and 1.61) are linear in charge [202]. For example, for an *ab initio* system, the contribution to the FlucQ forces, by consideration of equation 1.58, is given by

$$\frac{\partial E_{QM/MM}}{\partial Q_{i\alpha}} = \sum_{\mu} \sum_{\nu} \frac{P_{\mu\nu} I^{i\alpha}_{\mu\nu}}{Q_{i\alpha}} + \sum_{m} \frac{Z_m}{r_{m,i\alpha}} \tag{3.40}$$

The interaction term for a semi-empirical system is similarly derived [203].

It should be pointed out at this stage that although the one-electron integrals, $I^{i\alpha}_{\mu\nu}$, are calculated at the start of the QM calculation, before the SCF, they cannot be used until the density matrix, P, is obtained at the end of the SCF procedure. Furthermore, the integrals are not stored in a form that is useful to FlucQ; consider the core Hamiltonian matrix, equation 1.17, which has elements containing the following energy contributions:

$$H^{core}_{\mu\nu} = \sum_{m} I^{m}_{\mu\nu} + \sum_{i,\alpha} I^{i\alpha}_{\mu\nu} + \ldots \tag{3.41}$$

The first term accounts for the interaction between the electron density and the QM nuclei m, while the second is an analogous term for the interaction with the MM partial charges $i\alpha$. Note that the sum over all MM atoms is stored in the core matrix, whereas the sum over all basis function pairs is required for FlucQ. Thus, FlucQ needs either to store all calculated $I^{i\alpha}_{\mu\nu}$ values for use at SCF completion, or alternatively they can be recalculated at the end of the energy calculation. This is the same trade-off between memory usage and computational expense that is considered for direct SCF calculations (section 1.2); however, one-electron integrals are relatively inexpensive to recalculate. Both approaches were implemented and evaluated for this

work; the additional overhead of recalculating the one-electron integrals was deemed to be acceptable for systems containing a small number of MM atoms, and preferable to using the large amounts of memory that would be needed to store these integrals for the large basis sets used in this work.

3.8 Ewald summation

By consideration of equation 1.71, the Ewald reciprocal space contribution to the electronegativities, $\partial E_{\mathrm{rec}}/\partial q_i$, can be written as

$$\begin{aligned}\frac{\partial E_{\mathrm{rec}}}{\partial q_i} &= \frac{2\pi}{L^3}\frac{1}{4\pi\epsilon_0}\sum_{k\neq 0}^{k_{max}}\frac{1}{k^2}\exp(-k^2/4\alpha)\left[S(\boldsymbol{k})\frac{\partial S(-\boldsymbol{k})}{\partial q_i}+S(-\boldsymbol{k})\frac{\partial S(\boldsymbol{k})}{\partial q_i}\right] \\ &= \frac{2\pi}{L^3}\frac{1}{4\pi\epsilon_0}\sum_{k\neq 0}^{k_{max}}\frac{1}{k^2}\exp(-k^2/4\alpha)\left[S(\boldsymbol{k})\mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}_i}+S(-\boldsymbol{k})\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}_i}\right]\end{aligned} \tag{3.42}$$

By use of the identity $\mathrm{e}^{\mathrm{i}\theta}\equiv\cos(\theta)+\mathrm{i}\sin(\theta)$ the imaginary parts can be eliminated to yield

$$\begin{aligned}\frac{\partial E_{\mathrm{rec}}}{\partial q_i} = \frac{2\pi}{L^3}\frac{1}{4\pi\epsilon_0}\sum_{k\neq 0}^{k_{max}}\frac{2}{k^2}\exp(-k^2/4\alpha)\Big[&\cos(\boldsymbol{k}\cdot\boldsymbol{r}_i)\sum_{j=1}^{N}q_j\cos(\boldsymbol{k}\cdot\boldsymbol{r}_j)+ \\ &\sin(\boldsymbol{k}\cdot\boldsymbol{r}_i)\sum_{j=1}^{N}q_j\sin(\boldsymbol{k}\cdot\boldsymbol{r}_j)\Big]\end{aligned} \tag{3.43}$$

CHARMM [10] Ewald code was modified to incorporate this contribution to the electronegativities, and also the charge derivatives of the other Ewald energy components, which are trivial to derive as these are linear or quadratic in the charge. The more advanced Particle Mesh Ewald code also used by CHARMM was not modified in this work, however.

3.9 Intramolecular interaction

The interaction between two atoms i and j in the same molecule, cannot use the standard Coulomb formula, as the atoms are generally sufficiently close for orbital overlap to be non-negligible. Instead, a screened Coulomb term is used, as proposed by Rappé and Goddard [194], corresponding to the interaction between Slater-type orbitals on the two atoms:

$$J_{ij}(r) = \int \mathrm{d}\boldsymbol{r}_i \mathrm{d}\boldsymbol{r}_j |\phi_i(r_i)|^2 \frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j - \boldsymbol{r}|} |\phi_j(r_j)|^2 \tag{3.44}$$

where the Slater-type orbitals themselves are given by

$$\phi_i(r) = A_i r^{n_i - 1} \mathrm{e}^{-\zeta_i r} \tag{3.45}$$

where A_i is a normalization constant, n_i the principal quantum number, and ζ_i the orbital exponent.

Calculation of these integrals on the fly is expensive, and greatly slows an energy calculation. Fortunately, the model is usually applied to rigid solvent molecules, such that all interatomic distances within each molecule are fixed. Thus, J_{ij} values can be precalculated at these fixed distances at the start of the calculation, to avoid any integral evaluations during the run.

FlucQ parameters were determined by Rick *et al.* [30] for the SPC [122] and TIP4P [124] water models shown in section 1.10.5 (giving rise to the SPC-FQ and TIP4P-FQ models, respectively) so that the models reproduced gas-phase dipole moments and the energy, pressure and pair correlation functions of the bulk liquid; these parameters are summarised in table 3.1. Note that the FlucQ model is fully parameterized by the orbital exponents ζ_i, the charge inertia parameter M_Q, and the relative electronegativity, $\tilde{\chi}_O - \tilde{\chi}_H$. Since the FlucQ model alters the electrostatic properties of the water molecule, Rick *et al.* also modified the van der Waals O–O

	SPC	TIP4P	SPC-FQ	TIP4P-FQ
r_{OH} (Å)	1.0	0.9572	1.0	0.9572
r_{OM} (Å)		0.15		0.15
θ_{HOH} (degrees)	109.47	104.52	109.47	104.52
Q_O (e)	-0.82	-1.04		
$\tilde{\chi}_O - \tilde{\chi}_H$ (kcal/mol/e)			73.33	68.49
ζ_O (a_0^{-1})			1.61	1.63
ζ_H (a_0^{-1})			1.00	0.90
M_Q (ps/e)^{2}kcal/mol			6.9×10^{-5}	6.0×10^{-5}
ϵ (kcal/mol)	0.1554	0.1550	0.2941	0.2862
σ (Å)	3.166	3.154	3.176	3.159

Table 3.1: Parameters used for the fluctuating charge models, compared to those for the standard SPC and TIP4P models. Geometrical, FlucQ, and Lennard-Jones parameters are shown.

interaction to compensate. (Note also that their notation is used throughout this work, in which the O subscript is used to denote parameters which affect the oxygen charge, in both the SPC-FQ and TIP4P-FQ models, despite the fact that in TIP4P-FQ this charge is actually physically located on the M site.)

3.9.1 Flexible water models

For water molecules where the O–H and H–H distances can change due to bond stretches and bends, the FlucQ model as originally developed suffers from a severe disadvantage, as the evaluation of the intramolecular interaction integrals (equation 3.44) is computationally demanding. These integrals can, however, be avoided by fitting a simpler function of r to the $J_{ij}(r)$ curve. From inspection of figure 3.1, this curve can be seen to tend to the pure Coulomb $1/r$ interaction as $r \to \infty$, and to a Gaussian-like curve as $r \to 0$. Thus, for this work the curve is fitted

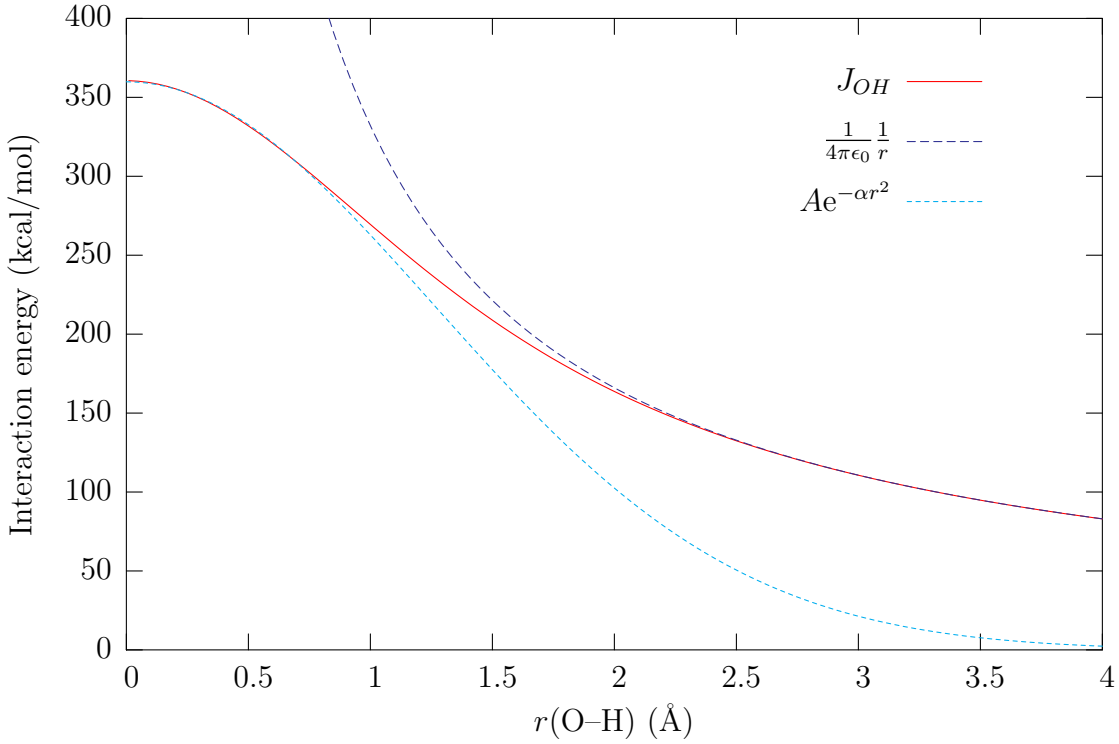

Figure 3.1: The screened Coulomb interaction term between oxygen and hydrogen atoms in the FlucQ model, J_{OH}. The standard Coulomb interaction and a single best-fit Gaussian for the function at small r are also shown for comparison.

to the function

$$J'_{ij}(r) = F(r)\sum_{k}^{N}\left(A_k \mathrm{e}^{-\alpha_k r^2}\right) + (1 - F(r))\frac{1}{4\pi\epsilon_0}\frac{1}{r} \tag{3.46}$$

where the number N of Gaussian functions in the linear combination can be increased to improve the fit at small r, and $F(r)$ is a switching function characterised by a switching distance, R_{sw}:

$$F(r) = \frac{(r/R_{sw})^8 + 2}{(r/R_{sw})^8 + 1} - 1 \tag{3.47}$$

Optimum values of the parameters A_k, α_k and R_{sw} are then determined by minimizing the function

$$\chi^2 = \int [J_{ij}(r) - J'_{ij}(r)]^2 \mathrm{d}r \tag{3.48}$$

	J_{HH}	J_{OH}
A_1 (kcal/mol)	321.01	260.08
α_1 (Å^{-2})	0.21735	0.16252
A_2 (kcal/mol)	40.500	100.32
α_2 (Å^{-2})	1.5526	0.77669
R_{sw} (Å)	1.5497	1.5296
RMS (kcal/mol)	0.11787	0.11872

Table 3.2: TIP4P-FQ intramolecular interaction parameters.

In this work, a linear combination of two Gaussians was used. The integral in equation 3.48 was approximated by the sum over points in the range $0 < r \leq 4.0$ Å, spaced by 0.01 Å, and the function optimized by Fletcher-Reeves conjugate gradient minimization [108, 204]. The procedure continued until the determined RMS deviation between the Slater integration $J_{ij}(r)$ and the approximate function $J'_{ij}(r)$ over the range in question changed by less than 10^{-5} kcal/mol between successive minimization steps. The determined parameters for the relevant interactions in TIP4P-FQ water are displayed in table 3.2.

As a test of the approximate function, it was applied to the water dimer. The geometry and charges of the dimer were optimized with the flexible TIP4P-FQ model by 1000 steps of Adopted Basis Newton Raphson minimization. The minimum energies, O–O distances, CPU time required, and charges on one of the water molecules are shown in table 3.3 for the two cases where the intramolecular interactions are calculated by integration or by the approximate function. As can be seen, in this case the approximation cuts the CPU time required by over 98%, while yielding properties which differ from those of the 'exact' simulation by a little under 1%. The accuracy of the approximate function could perhaps be improved further if necessary by adding more Gaussian functions to the linear combination, or by precalculating values of the integrals and storing them in a lookup table for later use via linear or cubic spline [205] interpolation.

	Integration	Approximate function
Energy, kcal/mol	-64.136	-63.690
r(O–O), Å	2.923	2.927
q_O, e	-0.942	-0.935
q_{H_1}, e	0.527	0.523
q_{H_2}, e	0.415	0.412
Approximate runtime	100.0%	1.6%

Table 3.3: Flexible TIP4P-FQ minimized water dimer properties, comparing integration of the intramolecular interaction with an approximate function.

3.10 FlucQ energy terms

In summary, the FlucQ model adds two energy terms to a standard MM force field — a polarization term and, at least during molecular dynamics, a charge kinetic energy term. The former is given as:

$$E_{pol}^{FQ} = \sum_{i=1}^{N_{molec}} \sum_{\alpha=1}^{N_{atom}} \left(\tilde{\chi}_{i\alpha}^{0} Q_{i\alpha} + \frac{1}{2} \sum_{\beta=1}^{N_{atom}} J_{\alpha\beta}(\boldsymbol{r}_{i\alpha,i\beta}) Q_{i\alpha} Q_{i\beta} \right) - N_{molec} E_{gp} \tag{3.49}$$

It can be seen that is essentially the same as equation 3.13, except that the intermolecular contributions have been removed; these are calculated by the standard Coulomb term. Also, the parameter $J_{\alpha\alpha}^{0}$ has been set equal to $J_{\alpha\alpha}(\boldsymbol{r} = 0)$.

The charge kinetic energy term is simply given by:

$$E_{kin}^{FQ} = \sum_{i=1}^{N_{molec}} \sum_{\alpha=1}^{N_{atom}} \frac{1}{2} M_Q \dot{Q}_{i\alpha}^{2} \tag{3.50}$$

3.11 Code implementation in CHARMM

In order to carry out the work described in this thesis, the CHARMM [10] package was extended by the author to calculate FlucQ energies and forces, and to perform minimizations, molecular dynamics, and Monte Carlo simulations. These modifications are included in CHARMM releases c28a3 and later; for a full description of the

CHARMM implementation, and the core FlucQ code, please refer either to CHARMM itself, or to the CD which accompanies this thesis.

3.11.1 Energy and force determination

The time-consuming part of the FlucQ algorithm is the determination of the nonbonded and QM/MM contributions to the instantaneous electronegativity. The former are calculated by a modification to the standard CHARMM nonbond routines, which by default loop over all relevant atom pairs $\{i, j\}$ and calculate an interaction energy E_{ij} and its first derivatives with respect to the Cartesian coordinates for each. These are accumulated to give forces on each atom, and a system total electrostatic energy. The code is modified when running with FlucQ to take these E_{ij} components immediately after calculation and accumulate them into per-atom arrays. (This part of FlucQ thus takes advantage of any parallelisation in the CHARMM nonbond routines.) A similar procedure is followed during the evaluation of the Ewald real-space and k-space energy terms, again modifying the base CHARMM code for efficiency and parallelisation. A specific FlucQ energy routine then accumulates these atomic energies from all parallel nodes, and calculates the first derivatives with respect to charge by simply dividing by the charge.[1]

The next step is the incorporation of QM/MM contributions; these are calculated by FlucQ after the QM SCF is complete. All of the relevant one-electron integrals are calculated by routines written for this work, which were derived from the standard one-electron routines of the various QM packages. Thus these are also parallelised in the cases where the parent codes are themselves parallelised — i.e. GAMESS-US, GAMESS-UK and CADPAC exhibit parallel speedup, while MOPAC does not. The relevant density matrix elements are then factored in to yield $\partial E_{QM/MM}/\partial q_i$ for each

[1] Obviously, this approach will fail for cases where the charge is numerically zero. However, such cases are rare, and CHARMM does not calculate electrostatic interactions in such cases anyway, so special-case code would be required for this eventuality.

MM atom i. This quantity is in fact the electrostatic potential due to the QM density at the MM atom, and thus these routines are also used for the determination of the electrostatic potential at arbitrary points, for example in the CHELPG algorithm.

At the end of the CHARMM energy calculation, FlucQ code gathers and sums the calculated electronegativities from all processors. The self-energy and intramolecular contributions are then added, by application of equation 3.14. Charge forces are then calculated by equation 3.24 or equation 3.25.

3.11.2 Charge dynamics

The CHARMM dynamics routines are modified in this work for FlucQ such that when new Cartesian velocities are calculated by the various CHARMM integrators (leap-frog Verlet, velocity Verlet, Langevin etc.) the charge forces calculated by the energy routines are used by FlucQ to determine charge velocities and accelerations. It is also at this point that the various thermostats (if used) are applied, and the dynamical energy terms (FlucQ charge kinetic energy, thermostat kinetic and potential energy) are calculated. This particular part of the code is not parallelised, as the computational work required is minor when compared to the previous calculation of the energy, and so little benefit would be observed. When the CHARMM integrator is used to generate new Cartesian coordinates with the previously-calculated forces, FlucQ is also called to similarly propagate the charges.

3.11.3 General modifications

A variety of other minor modifications were necessary for FlucQ. For example, trajectory files read from or written to during FlucQ-enabled dynamics or Monte Carlo include not only the standard Cartesian coordinates of all atoms, but also their charges. Similarly, restart files include charges and charge velocities. Monte Carlo moves are enabled by implementation of equation 3.39. The CHARMM minimization routines are

altered to minimize a function not of $3N$ Cartesian variables, for an N-atom system, but a function of $4N$ variables, which includes the charge of each atom. (In order to determine overall RMS forces, a conversion of 1 kcal/mol/e to 1 kcal/mol/Å is used.)

The performance of the FlucQ code was optimized by avoiding square root and division operations (which are particularly computationally demanding) wherever possible, particularly within the energy calculation and other loops. The former were removed by dealing in squared quantities wherever practical, and the latter by multiplying by the precalculated reciprocal if possible. Conditionals were eschewed within loops to increase the efficiency of pipelining operations, and to make the code vectorizable in some cases (see section 2.3.1). All FlucQ code was also added by means of conditional compile directives, so that a version of CHARMM could be built without FlucQ present if desired, for performance comparisons.

3.12 Summary

The FlucQ method of Rick *et al.* [30] provides an elegant and computationally inexpensive mechanism for determining the charges of a system by electronegativity equalisation [31], and maintaining these during dynamics. As part of this work, the FlucQ algorithm was added to CHARMM [10] by the author, and extended for use not only in molecular dynamics simulations but also in minimization, by means of dissipative Langevin dynamics or standard CHARMM first-order optimization methods, and in Monte Carlo. The implementation is completely general, simply adding extra terms and parameters to the CHARMM force field, and could thus be used for the application of the fluctuating charge methodology to any system for which parameters are available. The time-consuming integral evaluation required for the simulation of flexible molecules with FlucQ, a major drawback, was avoided by fitting an empirical function to the interaction. The performance of both the FlucQ methodology and the TIP4P-FQ water model is evaluated in Chapter 4.

Chapter 4

Fluctuating charge test cases

4.1 Overview

The FlucQ model was originally parameterized to give an accurate estimation of the properties of bulk solvent [30], and the simplicity of the extended Langrangian formalism means that it can be rapidly applied to such systems. It is also straightforward to extend the model to handle QM/MM interactions. However, the ubiquitous nature of solvent means that the model must also be able to deal well with solute-solvent interactions, and ideally it would reproduce QM behaviour for all systems, from bimolecular complexes through clusters to the bulk. The TIP3P and TIP4P solvent models, and to a lesser extent SPC, are already in wide use for this purpose, however, and thus FlucQ must be carefully compared with these existing models to evaluate its strengths and weaknesses.

4.2 Bulk solvent test cases

An essential test for any solvent model is the determination of bulk solvent properties such as average dipole moment, radial distribution functions, and dielectric constants. In this section, therefore, the TIP4P-FQ model is tested for use with both dynamics and

Monte Carlo methods, and some subtleties of simulating bulk solvent with fluctuating charges are discussed.

4.2.1 Use of cutoff approximations

When using simple periodic boundary conditions, all pair potentials must be truncated at half the box length or less, as per the minimum image convention (see section 1.7.1). As was mentioned in section 1.6, the inaccuracies introduced by the imposition of this cutoff can be ameliorated by employing switching or shifting functions. However, the situation is complicated slightly for electrostatics when dealing with methods which deal with electron density, such as FlucQ or QM approaches, as not only does the choice of cutoff function affect the energy of interaction between the electron density and its environment (as in a standard MM force field), it also affects the nature of that electron density itself.

The interaction of FlucQ with various CHARMM electrostatic cutoff functions was thus investigated. For each type of cutoff, a standard pre-equilibrated cubic box of 256 TIP4P waters, with a side length of 19.716 Å, was used. Simple periodic boundaries were imposed, and pair potentials were not evaluated for any atoms separated by more than 9.5 Å, as per the minimum image convention. The FlucQ polarization energy was minimized for the whole system by applying dissipative Langevin dynamics, using a frictional coefficient of 1600 ps^{-1} and a time step of 1 fs, until the charge temperature dropped below 10^{-6} K. The water box was then equilibrated by applying molecular dynamics for 10 ps, using random starting Cartesian velocities assigned from a Gaussian distribution consistent with simulation at a temperature of 298 K. The temperature of these degrees of freedom was thermostatted at 298 K by application of Berendsen's weak coupling algorithm [117] with a coupling constant (γ in equation 1.106) of 0.001 ps^{-1}. That of the charges was kept at 1.0 K by a Nosé-Hoover thermostat [118–120] with an inertia parameter (Q, from equation 1.111) of 10.0 ps^2 kcal/mol. The equilibrated

system was then simulated for a further 50 ps with molecular dynamics. Van der Waals interactions, being independent of the FlucQ charges, and dropping off much more rapidly with distance than electrostatics, were simply treated in all cases with a simple shifting function, such that the energies dropped smoothly to zero at 9.5 Å.

Firstly, simulations using electrostatic force shift potentials (where the electrostatic forces are modified at all distances so as to drop smoothly to zero at the cutoff distance, 9.5 Å) were compared to those with force switch potentials (where the forces are 'switched' from their normal value at the inner cutoff distance, 9.0 Å, to zero at the outer cutoff, 9.5 Å). Electrostatic *energy* shifts and switches were not considered in this work, as they have been previously shown to give poorer results in general [85] and specifically for water simulation [206, 207].

Secondly, atom-based and group-based cutoffs were contrasted, where the interatomic distance used for the cutoff determination (for both electrostatic and van der Waals interactions) is calculated as the true atom-atom distance, or the distance between the centres of mass of the two atoms' parent molecules, respectively. Note that group-based force shifts were not used, as these are not supported by CHARMM.

It can be seen from figure 4.1 that the dipole distribution produced from an atom-based force shift simulation is around 0.15 Debye lower on average than that produced with a group-based force switch. (The latter, an average of about 2.6 Debye, is consistent with that determined by Rick *et al.* [30]. In comparison, the TIP3P and TIP4P models yield dipoles of 2.35 and 2.18 Debye, respectively [124].) The average dipole moment is consistently lower when using a force shift presumably because the FlucQ forces on each water molecule are consistently lower, due to the nature of the shifting function, such that each water molecule is less polarized by its environment and is more gas-like. (A force shift has also been shown to be damaging to the short-range interactions of charged groups in standard MM simulations [85].)

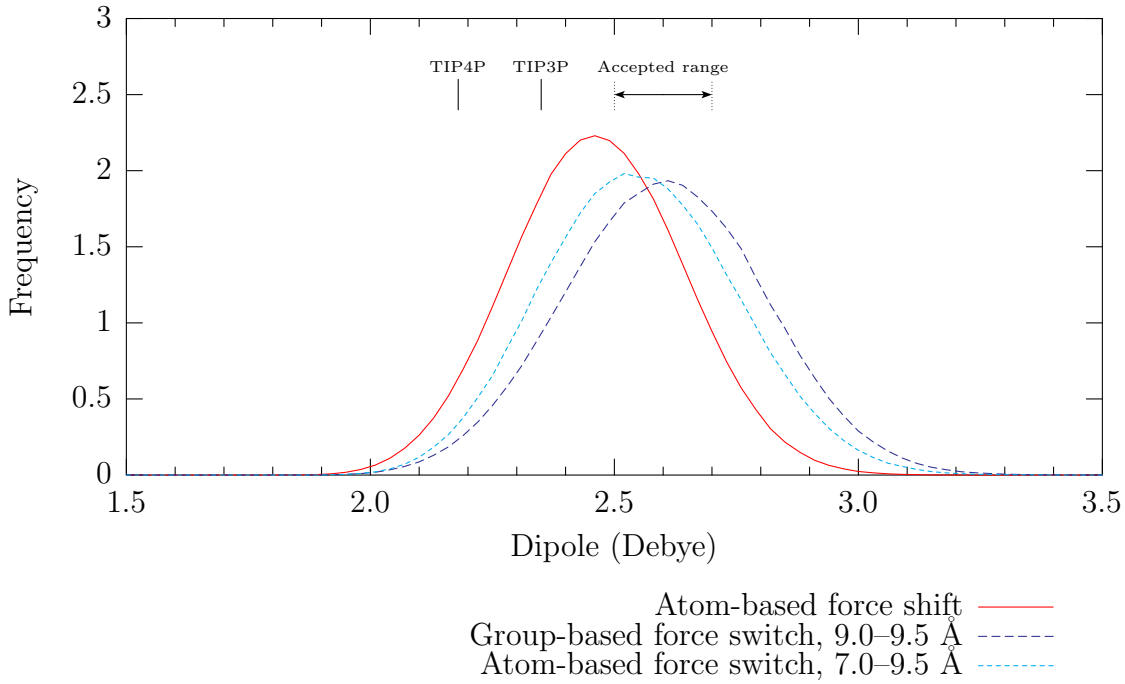

Figure 4.1: FlucQ dipole distributions using different cutoff methods. For comparison, the solution-phase dipoles for the TIP3P (2.35 D) and TIP4P (2.18 D) models, and the generally-accepted range for the solution-phase dipole (2.5–2.7 D, from section 3.1), are also shown.

Although a group-based force switch gives good results, the atom-based equivalent is more problematic. Simulations using a switching function acting between an inner cutoff of 9.0 Å or 8.5 Å and an outer cutoff of 9.5 Å failed due to the FlucQ charge temperature approaching infinity. This is due to bad interactions with water molecules within the switching region; for example, the oxygen charge can fall within the inner cutoff, while the hydrogens fall outside of the outer cutoff, such that the FlucQ forces are more representative of an isolated anion than of a neutral molecule. These incorrect forces lead to elevated charge temperatures, which only worsen the problem. (This does not occur with a group-based force switch, since all of the atoms in the water molecule experience the same switch function.) The problem is alleviated by using a wider switching region, as shown in figure 4.1 for a atom-based force switch between 7.0 and 9.5 Å. Note, however, that the dipole is still slightly underestimated.

For comparison, a simulation was run with standard TIP4P water and a force shift cutoff of 9.5 Å. This was carried out in exactly the same way as the FlucQ simulations, except that the TIP4P van der Waals parameters were employed, and static TIP4P charges were used on each of the water molecules. The corresponding FlucQ simulation was thus revealed to require roughly 4% more CPU time than the static charge treatment, which itself took just over 2 hours on a single 800 MHz Pentium III machine.

4.2.2 Thermostat test cases

To test the applicability of the various thermostats, the same box of 256 TIP4P-FQ waters used above was simulated with molecular dynamics for 20 ps, using a group-based force switch to cut off the long-range electrostatic interactions. Before starting the dynamics run, the charge degrees of freedom were minimized as described above. The Cartesian degrees of freedom were maintained at 298 K by Berendsen weak coupling with a 0.001 ps^{-1} coupling constant, while in the initial run the charge degrees of freedom were allowed to evolve naturally with no thermostatting. The resultant fluctuations in the charge temperature are shown in figure 4.2; the corresponding fluctuations in the thermostatted Cartesian degrees of freedom are shown in figure 4.3. As might be expected for a pre-equilibrated system, the Cartesian temperature of the system remains stable for the duration of the simulation, fluctuating by less than 1%. However, the charge temperature increases gradually during the run. This is partly an indication of incomplete equilibration of the charge degrees of freedom; furthermore, as the charge temperature increases, the charge distribution deviates more and more from the minimum energy configuration and the validity of the extended Lagrangian approximation is reduced. Thus, thermostatting is clearly necessary for runs of more than a few tens of picoseconds.

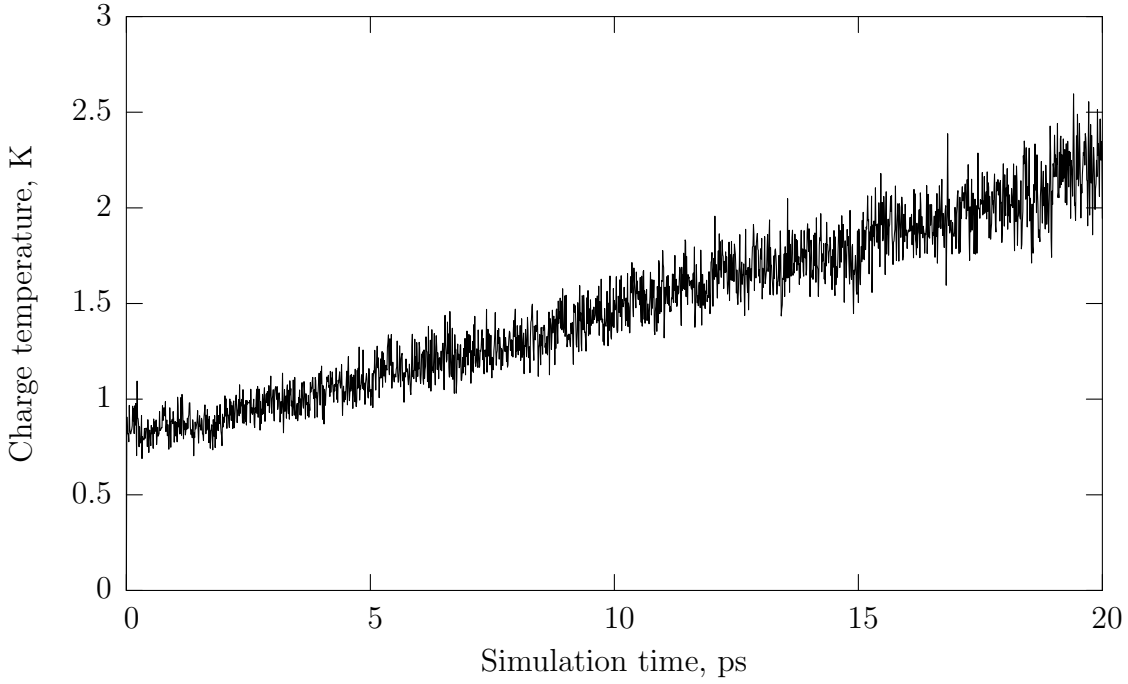

Figure 4.2: Charge temperature evolution with no thermostat applied.

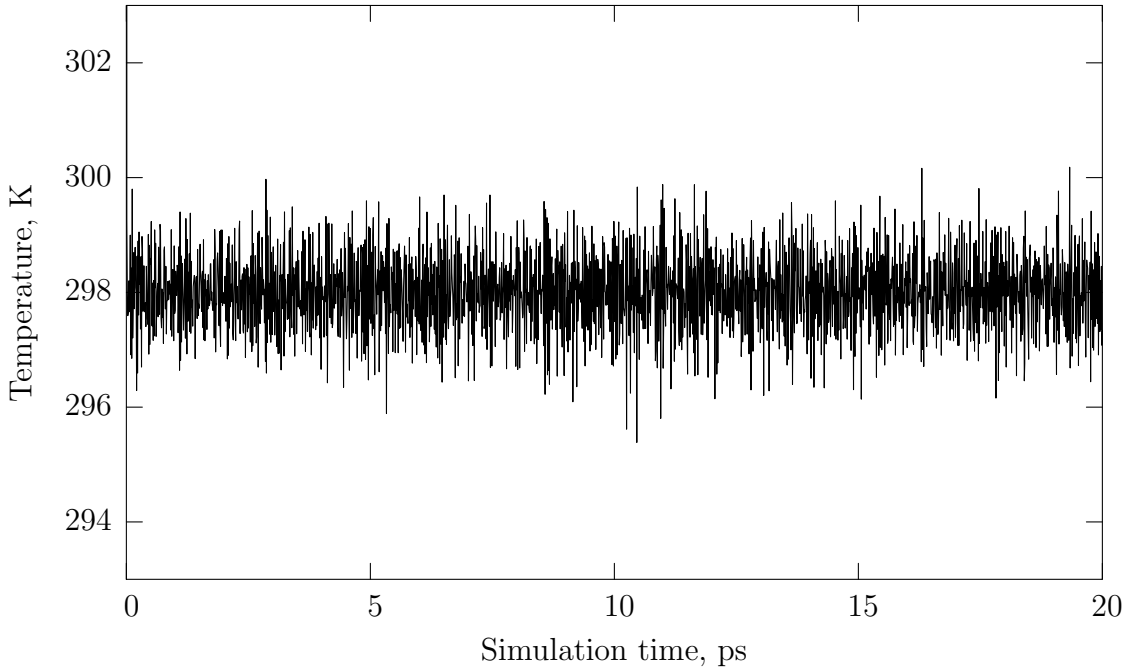

Figure 4.3: Evolution of the temperature of the Cartesian degrees of freedom, with a Berendsen thermostat of coupling strength 0.001 ps^{-1} applied.

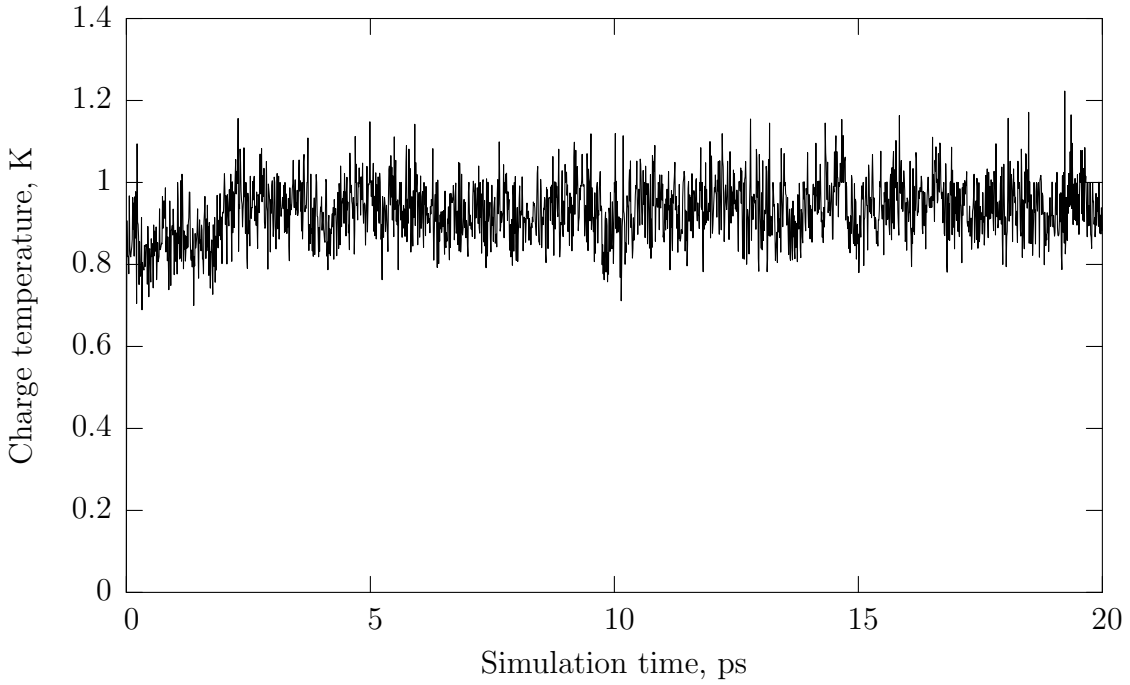

Figure 4.4: Charge temperature evolution with the charge velocities scaled to 1.0 K every 50 time steps.

The simulation was repeated, with the charge velocities scaled every 50 time steps to maintain a charge temperature of 1.0 K. Velocity scaling was only applied if the original charge temperature was greater than 1.0 K, i.e. it was not used to heat the system. The evolution of the charge temperature under this thermostat is shown in figure 4.4. It can clearly be seen that velocity scaling is very effective at maintaining the system temperature. However, very little control is offered over the strength of the coupling, and the dynamics are non-physical, so while it is suitable for initial equilibration, it is of little utility for production work.

The simulation was repeated with the charge degrees of freedom constrained to 1.0 K by a Berendsen-style weak coupling thermostat, with a coupling strength of 1.0 ps^{-1}. As can be seen from figure 4.5, application of this thermostat causes the charge temperature to remain relatively stable at around 1.0 K.

Finally, the simulation was repeated using a Nosé-Hoover thermostat for the charge temperature, again using a target temperature of 1.0 K, and a Nosé-Hoover inertia

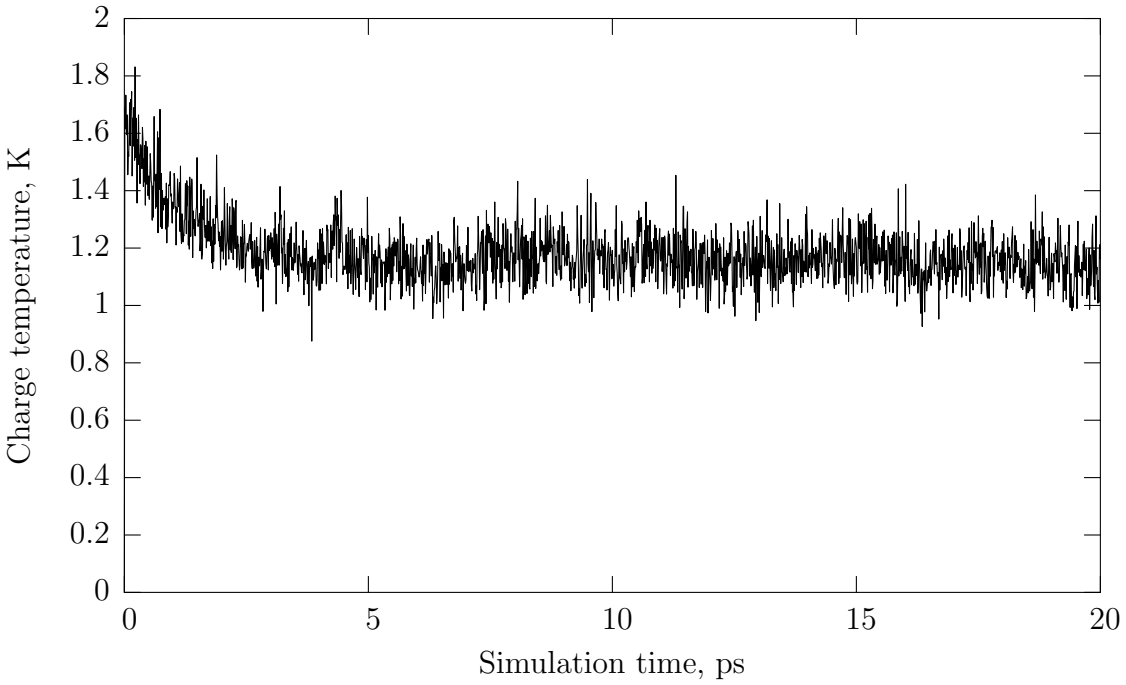

Figure 4.5: Charge temperature evolution with a Berendsen-style weak coupling thermostat of coupling strength 1.0 ps^{-1} applied.

parameter of 10.0 ps^2 kcal/mol. As can be seen from figure 4.6, this thermostat allows fluctuations similar in magnitude to the Berendsen method, the charge temperature increasing to 1.6 K for a short time, but the temperature again settles down to close to 1.0 K before the end of the simulation. This thermostat also has the advantage that, being thermodynamically rigorous, the energy removed from the charge degrees of freedom to those of the thermostat can be exactly quantified. The evolution of the thermostat parameters S and $\dot{S}$ is shown in figure 4.7, where it can be seen that both the potential energy (by reference to section 1.10.2, a linear function of S) and the kinetic energy (dependent on $\dot{S}$) of the thermostat increase during the course of the simulation, as the increase in charge temperature observed in figure 4.2 is absorbed into the thermostat's heat bath.

The time taken to apply a thermostat to the Cartesian or charge degrees of freedom is minimal when compared to that taken to evaluate the energy of the system. The iterative procedure outlined in section 3.5 is perhaps not the most efficient, but both

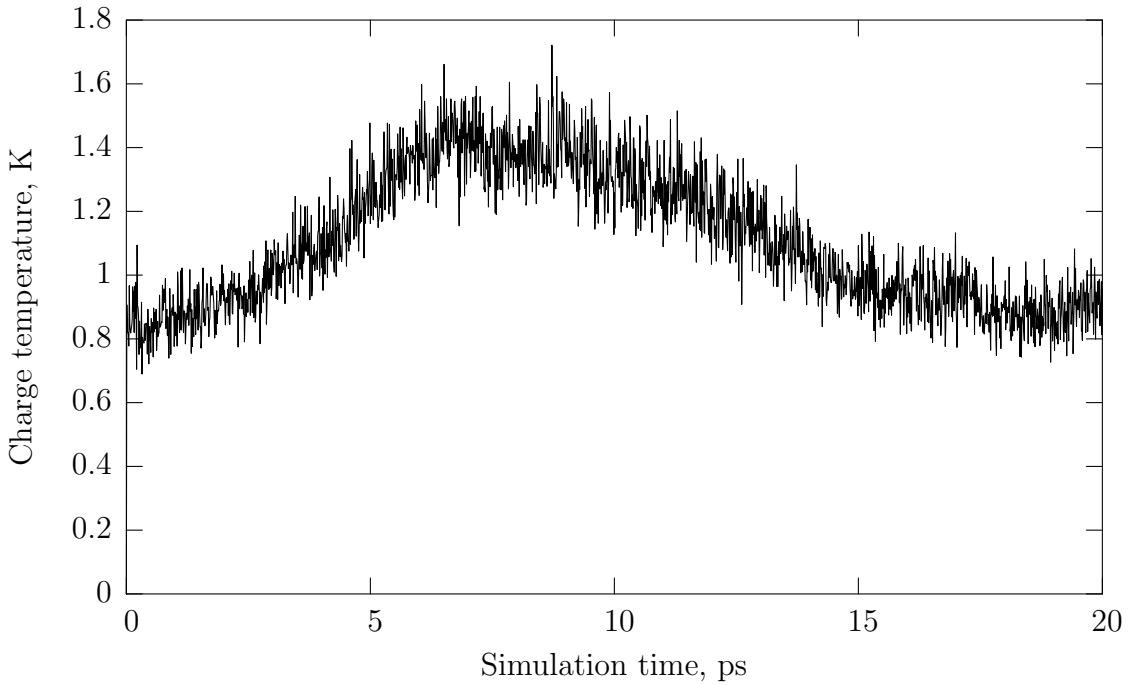

Figure 4.6: Charge temperature evolution with a Nosé-Hoover thermostat applied, with inertia parameter 10.0 ps^2 kcal/mol.

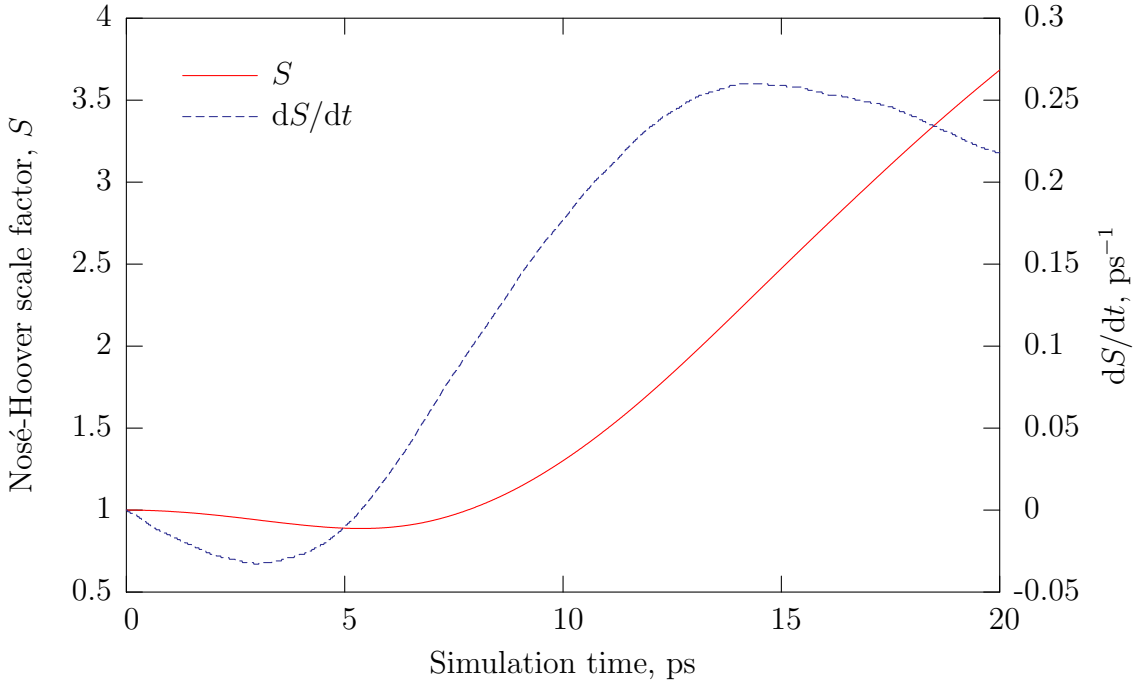

Figure 4.7: Evolution of the Nosé-Hoover charge temperature thermostat parameters S and dS/dt.

the Berendsen and Nosé-Hoover thermostats required only two or three iterations on average for convergence. In fact, it was observed that the execution times of the above simulations differed only by one or two percent. Additionally, the different thermostats did not noticeably change the properties of the simulated system; for example, the distribution of water dipole moments for each thermostat matched that produced in figure 4.1 for a group-based force switch. Thus there is little justification on efficiency grounds for using a less rigorous thermostat; both the Berendsen and Nosé-Hoover thermostats appear to perform well, and so the Nosé-Hoover was used for further FlucQ dynamics simulations in this work.

4.2.3 Simulation using Ewald summation

The problems introduced by the various cutoff methods are partly alleviated by using the Ewald sum (section 1.7.2) to evaluate pairwise interactions, although the technique is also far more computationally expensive. An Ewald simulation was run with an identical setup to that used for the investigation of cutoffs in section 4.2.1, i.e. a cubic box of 256 TIP4P-FQ waters, with the Cartesian degrees of freedom thermostatted at 298 K and the FlucQ charges at 1.0 K. 10 ps of equilibration were followed by 50 ps of production dynamics. Van der Waals interactions were shifted to zero at 9.5 Å, while electrostatics were evaluated using the Ewald sum. To avoid multiple calculations of the complementary error function, equation 1.70, during evaluation of the k-space part of the Ewald sum (equation 1.71) values were tabulated at the start of the calculation; table values were then used via a cubic spline interpolation during the simulation. Real-space interactions were evaluated up to a cutoff distance of 9.5 Å, and k-space interactions for the remainder; conservative values, as suggested by the CHARMM documentation, were chosen for the Ewald parameters to optimize this. A Gaussian width parameter, $\sqrt{\alpha}$, of 0.42 was used, while the l_x, l_y and l_z components of the $\boldsymbol{k}$ vectors were allowed to vary up to 11, with a spherical cutoff of 300 applied in k-space

(i.e. $l_x^2 + l_y^2 + l_z^2 \leq 300$). The simulation was repeated with standard TIP4P parameters and static charges for comparison. The extra calculations required to calculate the charge forces for FlucQ increased the calculation time (when compared to TIP4P, which took approximately 45 hours on an 800 MHz Pentium III PC) by roughly 10.5%.

In order to analyze the bulk solvent structure, configurations were recorded every 10 time steps during each simulation. These were then used to calculate radial distribution functions, $g(r)$, where $g(r)$ is defined as the ratio of the number of atoms found at a distance r from some reference atom to those found in the ideal gas at equal density [86].

The distribution functions for O–O, O–H, and H–H pairs are shown in figures 4.8, 4.9 and 4.10 respectively. These functions are compared with neutron diffraction data collected by Soper and Phillips [123]. It can clearly be seen that TIP4P and TIP4P-FQ give very similar bulk solvent structure. This is hardly surprising, as TIP4P is optimized for bulk properties. However, TIP4P-FQ gives a slightly better fit to the experimental data than TIP4P, for both peak positions and intensities, particularly for the all-important first solvation sphere. For example, both TIP4P and TIP4P-FQ underestimate the oxygen-oxygen distance of the first solvation sphere (figure 4.8) when compared to experiment, although TIP4P-FQ comes slightly closer to the experimental value than TIP4P. Additionally, Rick *et al.* [30] have shown that TIP4P-FQ gives dielectric constants, diffusion constants, Debye relaxation times and NMR relaxation times in closer agreement with experiment than TIP4P.

4.2.4 Monte Carlo simulation

To test the applicability of FlucQ to Monte Carlo simulations, the box of 256 TIP4P waters employed in section 4.2.1 was used. Once again, periodic boundaries were used with a cubic box of side 19.716 Å, while the van der Waals interactions were shifted to zero at 9.5 Å, and a group-based force switch was used on the electrostatic interactions between 9.0 Å and 9.5 Å.

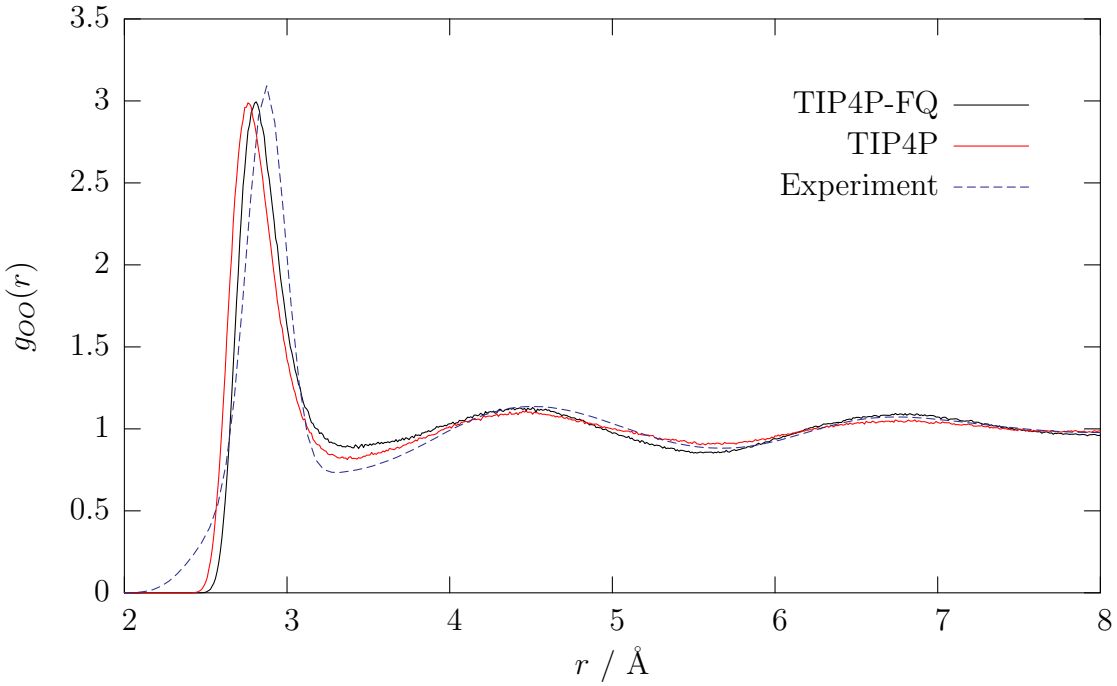

Figure 4.8: Oxygen-oxygen radial distribution functions for TIP4P and TIP4P-FQ, compared to experimental data.

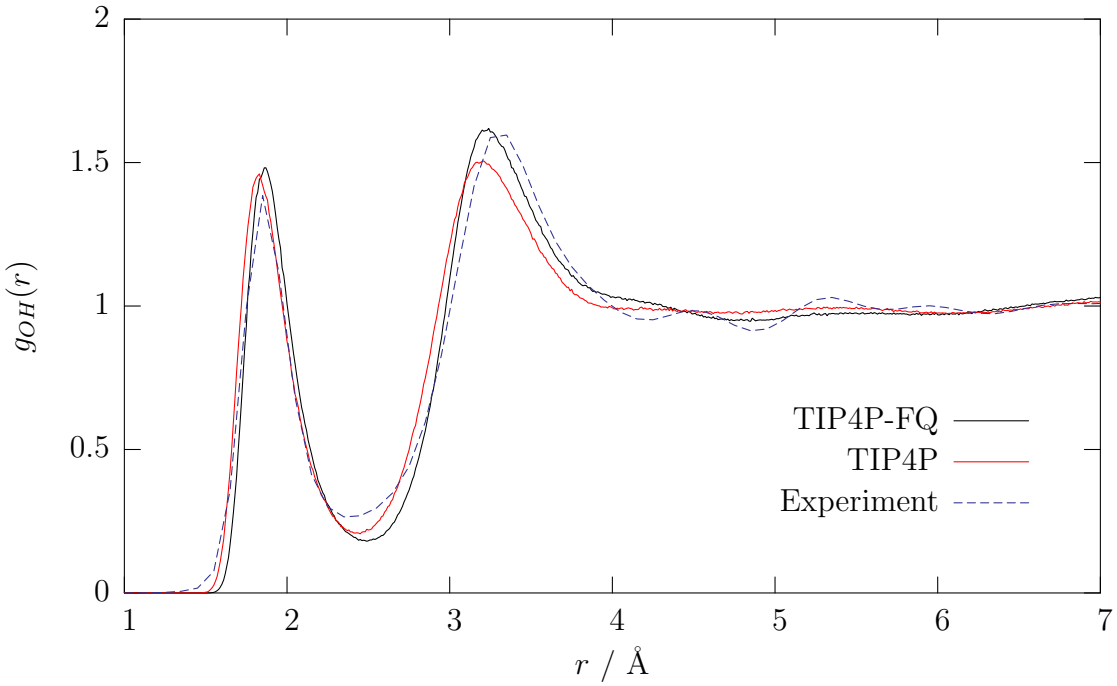

Figure 4.9: Oxygen-hydrogen radial distribution functions for TIP4P and TIP4P-FQ, compared to experimental data.

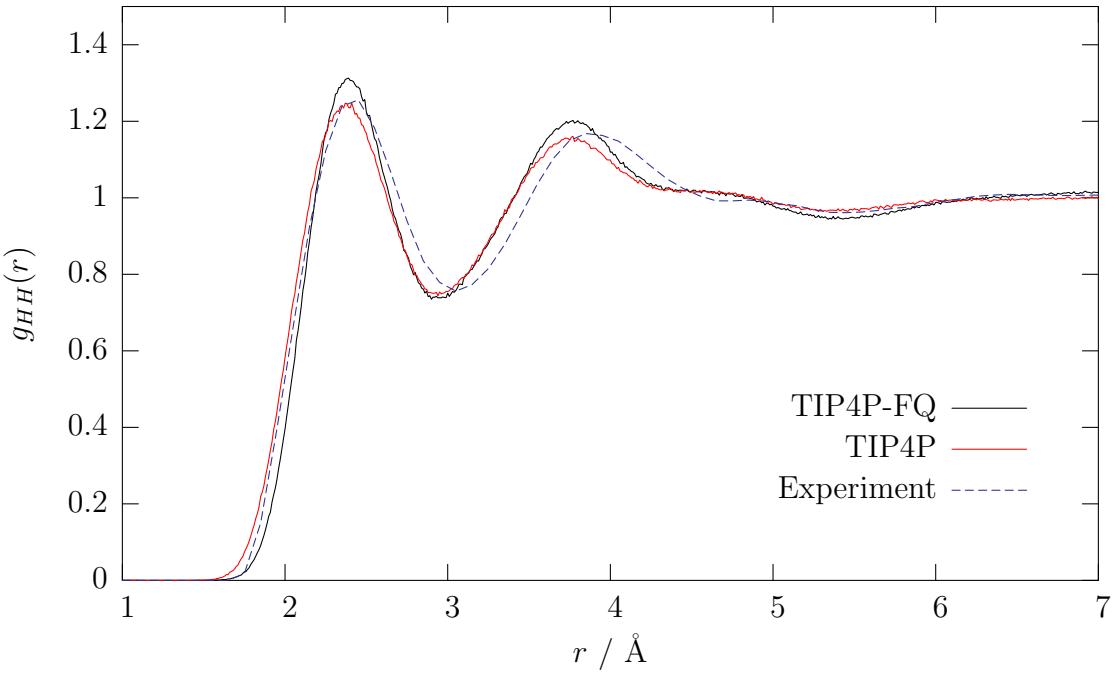

Figure 4.10: Hydrogen-hydrogen radial distribution functions for TIP4P and TIP4P-FQ, compared to experimental data.

Three moves were used to sample the configurational space of the system. The first was a simple H_2O rigid translation, where to perform the move a random water molecule and a random direction vector were chosen, and then every atom in that molecule was translated a random distance along the vector. The maximum displacement along the vector was chosen so as to maintain the acceptance rate of these moves at around 50%. The second move was a rigid rotation, for which a random water molecule was rotated a random angle about a random direction vector, which itself passed through the centre of mass of the molecule. Again, the acceptance rate was controlled by the choice of the maximum rotation angle. Finally, a FlucQ charge move, as defined in section 3.6, was used to sample the charge degrees of freedom. One of the three charge sites in a random water molecule was chosen at each move, and a random amount of charge removed and equally redistributed between the other charge sites. The acceptance rate was controlled by the choice of the maximum amount of charge to transfer.

Move type	Max. move size	Weight	Temperature
H_2O rigid translation	0.25 Å	0.5	298 K
H_2O rigid rotation	25.0°	0.5	298 K
FlucQ charge perturbation[a]	0.02 e	5.0	1.0 K

[a] TIP4P-FQ simulations only.

Table 4.1: The move set used in Monte Carlo simulations of bulk water.

Monte Carlo does not use forces or Newtonian dynamics, and thus has no concept of Cartesian or charge velocity. This removes the complication of having to apply thermostats to the degrees of freedom; instead, the temperatures simply determine the acceptance criterion in the Metropolis method, equation 1.121. As with the molecular dynamics simulations, a temperature of 298 K was used for the Cartesian degrees of freedom and one of 1.0 K was used for the charges.

Rigid translation and rotation moves were selected in Monte Carlo simulations with equal probability. Electrons move much more rapidly than the nuclei; thus, a higher weight was chosen for the FlucQ charge moves, such that on average 5 FlucQ moves were performed for every Cartesian (translation or rotation) move.

The maximum move sizes for the Monte Carlo moves were initially estimated, followed by a test run to give the acceptance rates for each move. The move sizes were then either increased (if too many moves were being accepted) or decreased (if the acceptance rate was far below 50%). This process was repeated until the acceptance rate for every move was roughly 50%. The final determined weights, maximum move sizes, and temperatures for all of the moves are shown in table 4.1.

The system was equilibrated for 2×10^5 steps, and then a production run of 10^7 steps was carried out. At each step, a move type was chosen at random, taking into account the move weights, the energy change of a trial move was calculated, and the move accepted or rejected by applying the Metropolis acceptance criterion with the given move's temperature. For comparison, a similar simulation of TIP4P water was carried out without FlucQ, by using the fixed TIP4P charges and not applying the FlucQ

charge move. Again, equilibration for 2×10^5 steps was followed by a production run, although only 2×10^6 production steps were required, due to the absence of the FlucQ move, to achieve the same sampling of the Cartesian degrees of freedom.

The resulting O–O pair distribution functions for Monte Carlo TIP4P and TIP4P-FQ simulations are shown in figure 4.11; the molecular dynamics Ewald results of figure 4.8 are also shown for comparison. The dipole distribution obtained from the MC simulation is shown in figure 4.12, together with that for the group-based force switch MD simulation from figure 4.1. It can clearly be seen that the Monte Carlo simulations give very similar results to the molecular dynamics. However, the average dipole for MC/FQ is very slightly lower than for MD/FQ, and the distribution function peak heights are slightly different. This is most likely due in part to insufficient sampling of the charge degrees of freedom. However, another explanation, suggested by Chen *et al.* [200], is that this sequential MC algorithm, where moves affect either the Cartesian or the charge degrees of freedom, is not physically accurate, since in reality any move of the nuclei would have a concomitant electron density move. This possibility was allowed for in their Adiabatic Nuclear and Electronic Sampling MC [200] or ANES-MC algorithm, in which every trial move of the nuclei is modified by performing one or more electronic moves, such that each move perturbs both the nuclei and the charges. This method was further applied to phase equilibrium calculations [201, 208].

A major drawback of the MC/FQ method is the much increased computational cost. While molecular dynamics simulations have been shown to be slowed by no more than about 10%, the much larger number of MC steps required to incorporate the FlucQ moves increased the time for the production run of this system by over 400% (from 75 minutes to just over 6 hours, on a 800 MHz Pentium III). It should be pointed out, however, that this overhead could be reduced by a more efficient implementation in code. Any move of the charges does not affect the interatomic distances; thus, the van der Waals interactions (equation 1.52) are unchanged and so do not need to be

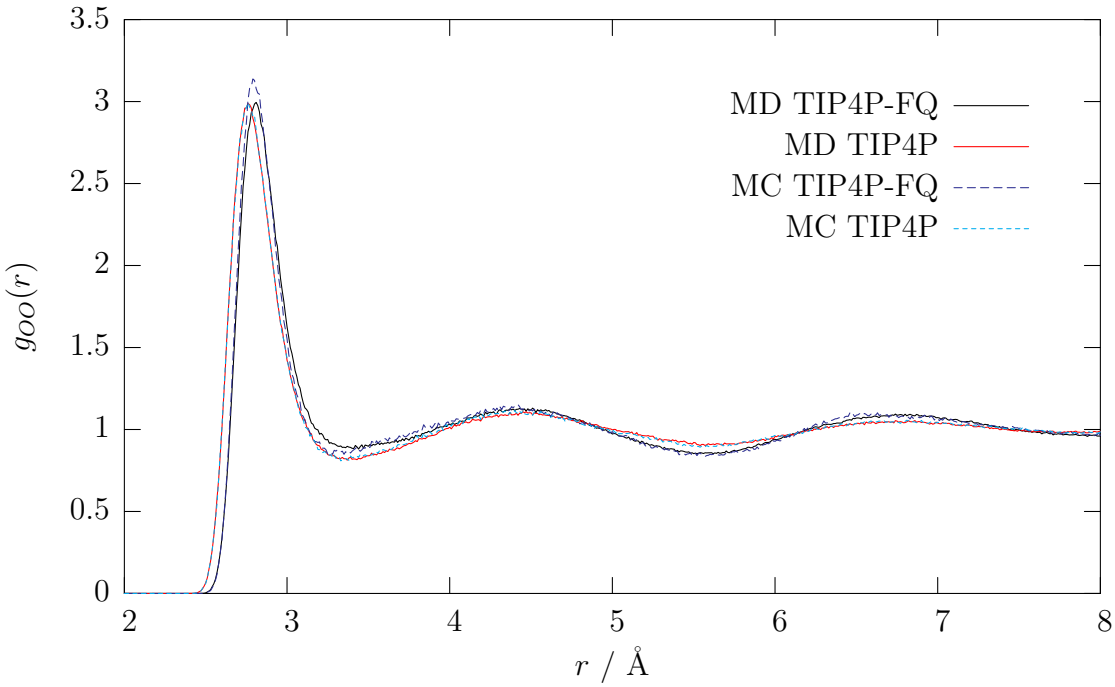

Figure 4.11: Oxygen-oxygen radial distribution functions for TIP4P and TIP4P-FQ, for both Monte Carlo and molecular dynamics simulations.

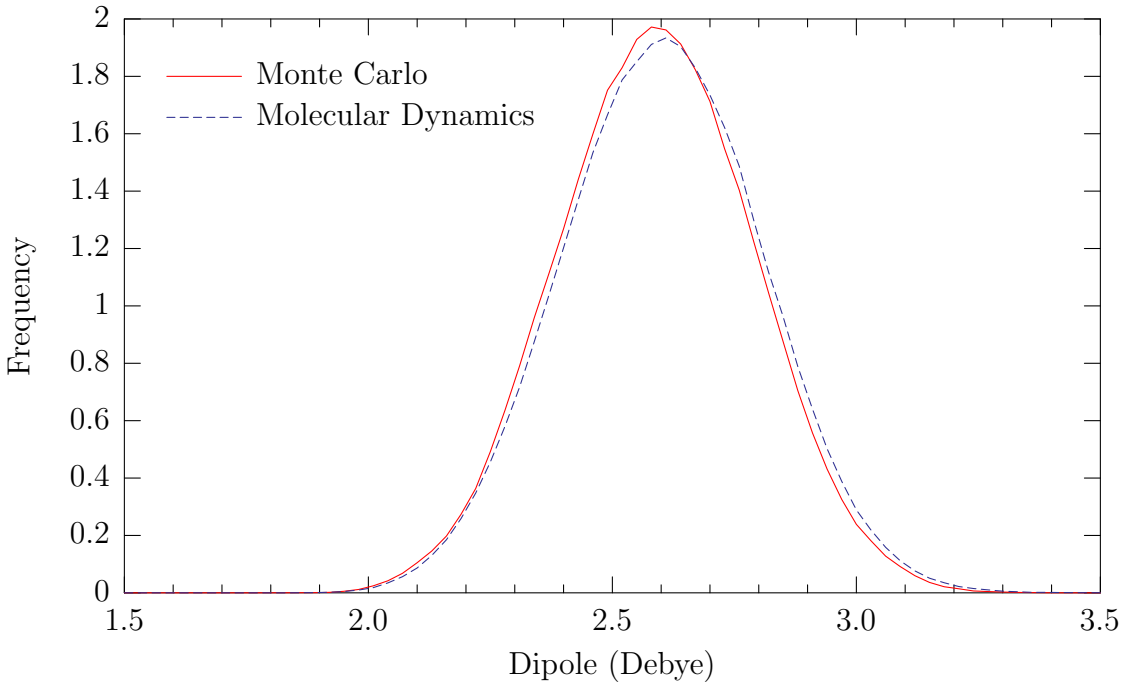

Figure 4.12: Dipole distributions using Monte Carlo and molecular dynamics.

calculated at all. The Coulombic interactions can also be optimized; by consideration of equation 1.53 it can be seen that the total Coulombic energy between a particle i and the remaining particles j in a system is given by:

$$V_{elec}(i) = \frac{1}{4\pi\epsilon_0} \sum_{j \neq i} \frac{q_i q_j}{r_{ij}} \tag{4.1}$$

Thus, the change in the Coulombic energy when a charge move $q_i \rightarrow q_{i'}$ is applied, given that all r_{ij} remain unaffected, is:

$$\Delta V_{elec}(i) = V_{elec}(i') - V_{elec}(i) = \frac{1}{4\pi\epsilon_0} \sum_{j \neq i} \frac{(q_{i'} - q_i) q_j}{r_{ij}} \tag{4.2}$$

Two separate evaluations of the Coulomb energy can thus be replaced with a single calculation, using the charge delta. However, one complete evaluation of all the relevant distances r_{ij} is still required, and this distance calculation is most likely the most expensive part of the energy calculation, due to the computational difficulty of evaluating square roots. Therefore, such optimizations would be unlikely to reduce the overhead of MC/FQ by more than a half, at most, thus still rendering the approach considerably more costly than the MD equivalent.

4.3 The water dimer

TIP4P-FQ has been seen to give excellent results for bulk solvent, performing very efficiently with MD and to a lesser extent with MC. However, its ability to model solute-solvent interactions must also be evaluated. An ideal starting point for this is the water dimer in the gas phase, which has been extensively studied both experimentally [209, 210] and theoretically [211–214].

4.3.1 Dipole determination

In order to seriously compare FlucQ or conventional static charge water models with *ab initio* calculations, a level of QM theory that gives water properties consistent with experiment should be used. A demanding test of a basis set is the accurate determination of the gas phase water dipole moment, and it has been shown [215] that only particularly large basis sets reproduce this quantity accurately. (Additionally, the dipole of an uncharged system is an observable, unlike point charge, so allowing direct comparison between MM and QM models.) A number of such basis sets have been formulated to complement the more commonly-used Pople basis sets such as 3-21G. One such basis set is of double-zeta quality, with polarization functions on heavy atoms, and is known as DZVP [216]. An even larger, triple-zeta basis developed by Dunning [217] is known as TZV. Researchers have added polarization functions derived by Roos and Siegbahn [218] on all atoms to this basis to generate TZVP, and have also added the field-induced polarization (FIP) functions of Zeiss *et al.* [219], again on all atoms, to generate TZVP+. The DZVP and TZVP+ basis sets have been shown to give results in very good agreement with the experimental water dipole moment [220].

A comparison of various QM methods and basis sets is given in table 4.2. The dipole moment of the water monomer, constrained to its experimental geometry, was calculated with Gaussian 98. Hartree Fock calculations were compared to both local (LDA) and nonlocal (B3LYP) DFT treatments with the same basis sets. It can quickly be seen that all of the Pople basis sets greatly overestimate the experimental value of 1.85 Debye, with the exception of STO-3G, which actually does reasonably well. However, it is likely, given the small size of this basis, that this is merely a fortuitous cancellation of errors. It can be seen by looking at the results for the split-valence 6-31G basis that the dipole moment is generally improved by the addition of polarization functions, but worsened by adding diffuse functions. There is also some benefit in

Basis set	HF	LDA	B3LYP
STO-3G	1.7258	1.7295	1.6792
3-21G	2.4350	2.3149	2.2555
6-31G	2.6301	2.5266	2.4605
6-31+G	2.7448	2.7277	2.6503
6-31++G	2.7338	2.7124	2.6365
6-31G*	2.2248	2.1259	2.0782
6-31G**	2.1844	2.0695	2.0289
6-31+G*	2.3316	2.3196	2.2594
6-31++G**	2.2860	2.2526	2.2017
6-311++G**	2.2390	2.2129	2.1635
DZVP	2.2768	2.2680	2.2033
TZVP	2.1963	2.1619	2.1145
TZVP+	2.0132	1.8801	1.8762

Table 4.2: Dipole moments (in Debye) calculated for the gas phase water monomer in its experimental geometry, for a variety of basis sets and QM methods. (The experimental value is 1.85 D.)

going to the triple-zeta 6-311G basis, which is perhaps not surprising as the move from double to triple zeta improves the description of the electron density.

DZVP and TZVP can also be seen to overestimate the water dipole. However, the addition of FIP functions to TZVP results in very reasonable estimations of the water dipole moment. It can also be seen that electron correlation is important for all basis sets; moving from Hartree Fock to LDA and then to B3LYP consistently results in a drop in the dipole moments, and better agreement with experiment, for all but the STO-3G case. Thus, it was decided to use the B3LYP DFT functional and the TZVP+ basis set for investigation of other water properties.

4.3.2 Minimum-energy properties

The water dimer was minimized at B3LYP/TZVP+ level, with the geometry of each water molecule constrained to experimental values, until forces had dropped to below

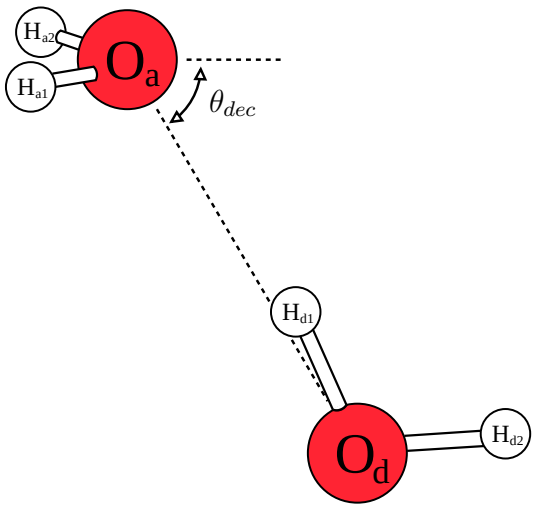

Figure 4.13: Geometry of the gas phase water dimer. The lower molecule acts as a hydrogen bond donor, donating hydrogen H_{d1} to the acceptor oxygen O_a. The donor lies within the same plane as O_a and the H_{a1}–O_a–H_{a2} bisector. The O–O bond is declined by an angle θ_{dec} below the plane of the acceptor water.

0.01 kcal/mol/Å. The resulting geometry is shown in figure 4.13. The system is of C_s symmetry, and contains a single hydrogen bond; the acceptor, O_a, of the hydrogen has two symmetry-equivalent protons, H_{a1} and H_{a2}. The planes of the two water molecules are perpendicular, with the donor water lying in the same plane as the H_{a1}–O_a–H_{a2} bisector. The O_d–H_{d1}–O_a hydrogen bond is found to be close to, but not actually at, linearity. This geometry is consistent with that observed experimentally in molecular beam experiments [209].

Determined energetic and geometric parameters from the minimization are shown in table 4.3, with experimental values for comparison where available. Charges were approximated for each of the atoms in the QM system by a CHELPG fit to the molecular electrostatic potential (see section 1.8). The B3LYP/TZVP+ treatment yields a dimer O–O distance of 2.926 Å. This is very close to the value of 2.919 Å determined from multi-determinant (CISD) treatments [211], suggesting that B3LYP accurately reproduces the correlation energy of the system. (It should be pointed out, however, that only the O–O bond length and θ_{dec} were optimized in the CISD

	Water model			
	B3LYP/TZVP+	TIP4P	TIP4P-FQ	Experiment[a]
$d(\mathrm{O_a \cdots O_d})$ (Å)	2.93	2.75	2.93	2.98
θ_{dec} (degrees)	56.97	49.15	27.18	57 ± 10
$\angle \mathrm{O_a \cdots O_d–H_{d2}}$ (degrees)	108.98	105.80	104.38	103 ± 10
System dipole (Debye)	2.66	2.72	3.41	2.63
ΔE_{bind} (kcal/mol)	-4.79	-6.23	-4.46	-5.44 ± 0.70
q_{O_a} (e)	-0.725	-1.040	-0.972	
$q_{H_{a1}}$ (e)	0.379	0.520	0.486	
$q_{H_{a2}}$ (e)	0.375	0.520	0.486	
q_{O_d} (e)	-0.770	-1.040	-0.958	
$q_{H_{d1}}$ (e)	0.376	0.520	0.540	
$q_{H_{d2}}$ (e)	0.365	0.520	0.419	

[a]See Odutola and Dyke [209] and Curtiss *et al.* [210].

Table 4.3: Water dimer properties obtained with each molecule constrained to experimental geometry.

study, the hydrogen bond being assumed to be linear.) It is somewhat shorter than the experimentally-determined value of 2.98 Å. The determined angles and binding energy, however, fall within the error limits of the experimental data, and that, taken with the close agreement of the dipole moment of the dimer between experiment and theory, shows that B3LYP/TZVP+ is a very reasonable model for the water dimer. Therefore, an accurate MM treatment might be expected to reproduce many of the QM-derived properties, including those for which experimental values are not available, such as point charges.

For comparison, the water dimer was modelled with the TIP4P-FQ and TIP4P models, and again minimized, with each molecule constrained to experimental geometry, until forces dropped below 0.01 kcal/mol/Å. The resulting properties are also shown in table 4.3 in comparison with the QM. Most notably, TIP4P greatly underestimates the O–O distance, and overestimates the binding energy. TIP4P-FQ predicts an O–O distance of 2.93 Å, in excellent agreement with QM and experiment.

The predicted binding energy is rather low, however, although it is in reasonable agreement with the QM result.

Looking at the point charges, it can be seen that TIP4P, with its fixed charges, rather poorly reproduces the QM results. TIP4P-FQ, on the other hand, shows similar H–H polarization in the donor water to QM. Some polarization would be expected, due to the strong electron-repelling negative charge on O_a. Less polarization is seen between the acceptor hydrogens, as they are both at similar distances from O_d. Both TIP4P and TIP4P-FQ predict larger point charges than QM, and correspondingly larger system dipoles. The dipole is particularly poorly predicted by TIP4P-FQ, probably because of the large charge difference, $0.121e$, that TIP4P-FQ suggests between the two donor hydrogens, as opposed to $0.011e$ from QM. This large charge difference gives rise to a large dipole on the donor water, and thus a greater tendency for that on the acceptor to orient, suggesting an explanation for the rather small θ_{dec} predicted by FlucQ.

4.3.3 Determination of the water-water binding energy profile

To examine the water-water interaction more closely, it was plotted as a function of the dimer O–O distance, where the latter was varied from 1.50 to 5.00 Å in steps of 0.05 Å. The geometry of each water-water complex was generated using the previously minimized dimer structure at B3LYP/TZVP+ level, and simply adjusting the O–O distance. The geometry of each monomer, and the angles θ_{dec} and O_a–O_d–H_{d2}, were kept equal to that of the minimized structure. The single-point energy at B3LYP/TZVP+ level was evaluated for each configuration, to give the energy profile shown in figure 4.14. A similar profile for TIP4P water was obtained by evaluating the CHARMM energy of each structure. Finally, that for TIP4P-FQ was determined by minimizing the system charges of each configuration by dissipative Langevin dynamics, with frictional coefficient 1600 ps^{-1} and time step 1 fs, until the charge temperature

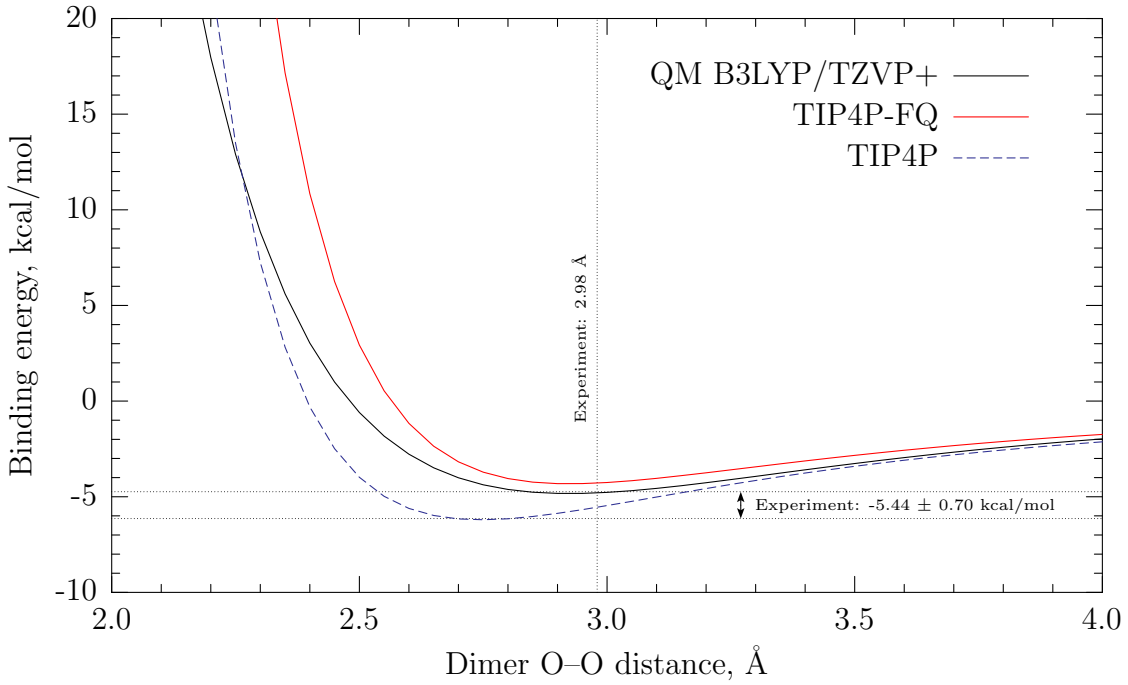

Figure 4.14: The binding energy of the gas-phase water dimer, as a function of the oxygen-oxygen separation, calculated for molecules in their experimental geometry using DFT, FlucQ, and the standard TIP4P water model. Experimental data (from table 4.3) are shown for comparison.

dropped below 10^{-6} K, and then evaluating the energy. These two profiles are also shown in figure 4.14.

The observations made about the binding energy and minimum O–O distance can be more clearly seen from this energy profile. TIP4P-FQ fits rather closely with the QM curve, although the interaction energy is always slightly underestimated. TIP4P, on the other hand, fits neither the shape nor the depth of the QM interaction.

It has already been seen that FlucQ performs excellently for bulk solvent simulations, and these results show that it also does very well at determining dimer binding energy and O–O distance. However, it is not clear from this simple examination of the total binding energy whether this agreement is due to FlucQ accurately reproducing the polarization of the system, or whether it is merely fortuitous parameterization of the force field. One means of investigating this further is to consider the QM electrostatic energy explicitly.

4.3.4 Energy decomposition

The QM interaction energy was approximately decomposed into electrostatic and van der Waals contributions by determining point charges on each of the nuclei at each point on the previously-obtained dimer interaction surface, by CHELPG fitting to the QM electrostatic potential; an approximate electrostatic interaction energy was then determined for each configuration by simple application of the Coulomb equation to these charges. The van der Waals term was then defined simply as the difference between this energy and the total binding energy; these two contributions are shown in figure 4.15. The electrostatic components of the TIP4P and TIP4P-FQ dimers are also shown in this figure; it can easily be seen that both conventional static charge and FlucQ treatments overestimate the magnitude of the electrostatic interaction, at least at the distances likely to be seen in simulations. This is consistent with the larger charges seen with the MM models for the minimum-energy dimer, in table 4.3. A solution to this problem is considered in section 4.6.

4.3.5 Charge transfer

In a quantum mechanical treatment, the total charge on each water molecule is not required to be zero; it is entirely possible that electron density can flow between molecules. Standard MM force fields, lacking an explicit treatment of electron density, cannot model this. The conventional FlucQ method, whereby charge is conserved within molecules, will obviously also fail to treat it. The charge transferred from the hydrogen bond donor to the acceptor, as a function of the O–O dimer distance, is shown in figure 4.16. It can be seen that the amount of charge transfer is rather small, becoming significant only for O–O distances less than 2.0 Å. By using equation 3.25 for FlucQ charge motion, charge transfer can be allowed in the FlucQ formalism. In this case, TIP4P-FQ predicts that a charge of $0.104e$ is transferred from donor to acceptor, giving rise to a minimum energy dimer O–O distance of 2.81 Å and a binding

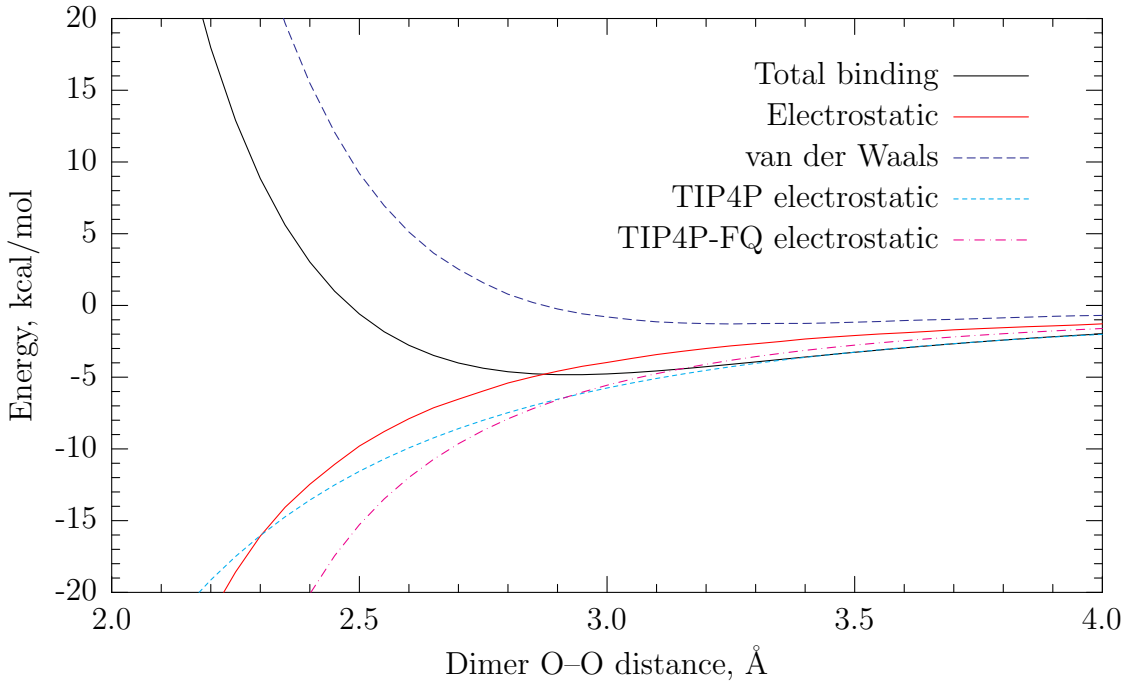

Figure 4.15: Decomposition of the gas-phase water dimer B3LYP/TZVP+ binding energy into electrostatic and van der Waals contributions. For comparison the electrostatic component of binding with TIP4P and TIP4P-FQ is also shown.

energy of 6.16 kcal/mol. TIP4P-FQ thus over-predicts the amount of charge transfer, causing large errors in the dimer properties; this is because the TIP4P-FQ model was parameterized under the assumption that charge transfer would not occur. Allowing charge transfer in systems larger than dimers would be even more problematic, due to FlucQ's lack of any restriction on the distance over which charges can be transferred. Thus, given the small amount of charge transfer that is observed at a QM level, it is probably reasonable to ignore this effect.

4.3.6 QM/MM studies

Having seen that both TIP4P-FQ and B3LYP/TZVP+ are reasonable models for the water dimer system, they were used together for QM/MM optimizations of the complex, with one water being treated quantum mechanically, and the other with MM. The water dimer is often used as a test case for the QM/MM method [17, 24, 221, 222]. The two

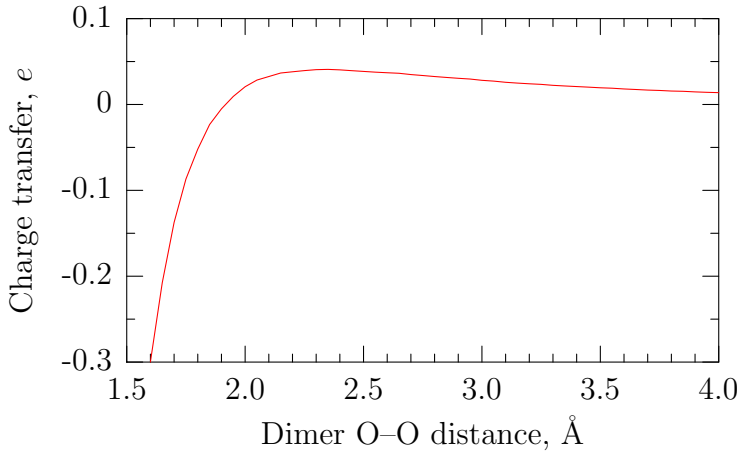

Figure 4.16: QM charge transfer (in units of electron charge) from the donor water to the acceptor, as determined from CHELPG analysis.

possible QM/MM partitions, where the acceptor water is treated with QM and the donor with MM, and the reverse case, were both minimized using B3LYP/TZVP+ for the QM region and TIP3P, TIP4P or TIP4P-FQ for the MM region, until forces dropped to below 0.001 kcal/mol/Å. TIP3P van der Waals parameters were used for the quantum oxygen in all cases. Determined energies, geometries, and charges (those of the QM atoms were derived by a CHELPG fit) of the final minimized structures are shown in tables 4.4 and 4.5.

First, by reference to table 4.4, it should be clear that the three water models all predict geometries in very good agreement with experiment. The O–O distance is, however, uniformly underestimated; this is partly because the MM models use larger charges on H_{d1} and H_{d2} than the QM charges on the acceptor hydrogens, and partly due to the use of van der Waals parameters optimized for bulk water-water interactions in an MM treatment, rather than the QM/MM. These larger charges also give rise to binding energies larger than the pure QM value. When the acceptor is in the MM region, as shown in table 4.5, the TIP3P and TIP4P models underestimate the O–O distance and over-predict the binding energy to a worse extent. TIP3P is particularly poor, with a very low θ_{dec} angle of 13.19°. TIP4P-FQ, on the other hand, is rather consistent between the MM donor and acceptor cases, with very similar binding energies

	TIP3P	TIP4P		TIP4P-FQ		Experiment[a]
ΔE_{bind} (kcal/mol)	-6.18	-6.54	(-6.23)	-5.08	(-4.46)	-5.44 ± 0.70
$d(\mathrm{O_a \cdots O_d})$ (Å)	2.81	2.79	(2.75)	2.81	(2.93)	2.98
θ_{dec} (deg.)	63.84	65.47	(49.15)	61.43	(27.18)	57 ± 10
$\angle \mathrm{O_a \cdots O_d}$–$\mathrm{H_{d2}}$ (deg.)	101.17	105.32	(105.80)	106.22	(104.38)	103 ± 10
q_{O_a} (e)	-0.762	-0.765		-0.778		
$q_{H_{a1}}$ (e)	0.381	0.383		0.389		
$q_{H_{a2}}$ (e)	0.382	0.383		0.389		
q_{O_d} (e)	-0.834	-1.040		-0.967		
$q_{H_{d1}}$ (e)	0.417	0.520		0.560		
$q_{H_{d2}}$ (e)	0.417	0.520		0.407		

[a] See Odutola and Dyke [209] and Curtiss *et al.* [210].

Table 4.4: Water dimer properties obtained with QM/MM methods; the donor water was modelled with one of three different MM models, while the acceptor was treated quantum mechanically, at B3LYP/TZVP+ level. Pure MM data (from table 4.3) are shown in brackets for comparison.

and O–O distances predicted in each case; its only failing is a rather small θ_{dec} in the latter case, although it is only very slightly worse than TIP4P in this respect, and much better than TIP3P. This suggests that the FlucQ model is able to reproduce the change in water polarization as it is moved from an acceptor position to a donor, at least in an average way, and that the van der Waals parameters for the QM/MM interaction are reasonably good.

4.4 Other bimolecular systems

The predicted geometries and energies of simple ion-water complexes have been previously used as tests of QM/MM methods [17, 24, 221]. Such systems are good models for more complex solvated systems, while the ions themselves have important roles in many biological systems. FlucQ is compared to the TIP3P and TIP4P models, with both MM and QM/MM methods used to treat the ions.

	TIP3P	TIP4P		TIP4P-FQ		Experiment[a]
ΔE_{bind} (kcal/mol)	-8.11	-6.95	(-6.23)	-5.18	(-4.46)	-5.44 ± 0.70
$d(\mathrm{O_a \cdots O_d})$ (Å)	2.69	2.72	(2.75)	2.83	(2.93)	2.98
θ_{dec} (deg.)	13.19	42.01	(49.15)	36.24	(27.18)	57 ± 10
$\angle \mathrm{O_a \cdots O_d}$–$\mathrm{H_{d2}}$ (deg.)	104.99	106.63	(105.80)	105.99	(104.38)	103 ± 10
q_{O_a} (e)	-0.834	-1.040		-0.979		
$q_{H_{a1}}$ (e)	0.417	0.520		0.489		
$q_{H_{a2}}$ (e)	0.417	0.520		0.489		
q_{O_d} (e)	-0.754	-0.749		-0.739		
$q_{H_{d1}}$ (e)	0.418	0.407		0.398		
$q_{H_{d2}}$ (e)	0.335	0.342		0.341		

[a] See Odutola and Dyke [209] and Curtiss *et al.* [210].

Table 4.5: Water dimer properties obtained with QM/MM methods; the donor water was treated quantum mechanically, at B3LYP/TZVP+ level, while the acceptor was modelled with one of three different MM models. Pure MM data (from table 4.3) are shown in brackets for comparison.

4.4.1 Water/chloride

The TIP4P-FQ method has been shown to perform reasonably well for this system. In particular, it predicts for clusters of the form $\mathrm{Cl(H_2O)_n^-}$ that the chloride is solvated on the outside of the waters, while non-polarizable solvent models suggest interior solvation for $n \geq 18$ [223]; more recent studies with Car-Parrinello and dipole-polarizable water models [224] also predict exterior solvation, at least for $n = 6$.

Using the TIP3P, TIP4P, and TIP4P-FQ water models, a system consisting of a single water molecule bound to a chloride ion was minimized without constraints until forces dropped below 0.001 kcal/mol/Å. For each model, full MM optimizations, in which the chloride ion was treated as a point negative charge, were complemented with QM/MM simulations, using B3LYP/TZVP[1] for the chloride in the QM region. In both cases, standard CHARMM van der Waals parameters [10, 225] were used for the chloride ion. For comparison, the system was also optimized in a fully quantum

[1] TZVP+ was not used, as FIP functions [219] are not available for chlorine.

Model	ΔE_{bind} (kcal/mol)	$d(\mathrm{O}\cdots\mathrm{Cl}^-)$ (Å)	$\angle$O–H$\cdots$Cl$^-$ (degrees)
TIP3P/MM	-15.96	3.07	143.2
TIP4P/MM	-15.86	3.02	161.0
TIP4P-FQ/MM	-17.55	2.95	177.0
TIP3P/QM	-15.58	3.11	116.3
TIP4P/QM	-14.65	3.14	149.3
TIP4P-FQ/QM	-14.01	3.12	174.3
QM	-17.74	3.11	167.9
Experiment[a]	-13.4		

[a]See Yamdagni and Kebarle [226]

Table 4.6: Determined energetic and geometrical properties for the H_2O/Cl^- system. QM calculations used the TZVP basis set for Cl^-, and TZVP+ for water, and utilised the B3LYP functional.

mechanical treatment, using the B3LYP functional and the TZVP+ basis on the water, and TZVP on the chloride. The determined properties are shown in table 4.6.

It can be seen from the determined O–H–Cl angle that the *ab initio* results, and those for TIP4P-FQ, predict a near-linear hydrogen bond with the chloride ion. This is in good agreement with previous studies [227], which predict a C_s symmetry for the complex. The geometries from TIP4P and TIP3P simulation are less good, particularly when the chloride ion is treated quantum mechanically, underestimating the O–H–Cl angle, an effect seen in previous DFT/MM studies [24]. TIP3P/QM is particularly poor, predicting a C_{2v} geometry (this has also been observed in TIP3P/AM1 optimizations [17]). All of the methods overestimate the experimental binding energy, most likely due to an incomplete treatment of the diffuse Cl^- electron density at both MM and QM levels.

The predicted binding energies from QM/MM simulation are consistently lower than the pure MM, and the O–Cl distances correspondingly greater, no doubt caused by the non-spherical charge distribution of the QM chloride being less polarizing than the point charge of the MM treatment. In table 4.7 CHELPG charges from the QM

System	Model	q_O	q_{H1}	q_{H2}	q_{ion}
	TIP4P-FQ/MM	-1.044	0.728	0.316	-1.000
H_2O/Cl^-	TIP4P-FQ/QM	-1.029	0.692	0.337	-1.000
	QM	-0.822	0.388	0.317	-0.883
	TIP4P-FQ/MM	-1.285	1.019	0.266	-1.000
H_2O/F^-	TIP4P-FQ/QM	-1.273	1.035	0.239	-1.000
	QM	-0.963	0.544	0.283	-0.863
	TIP4P-FQ/MM	-1.202	0.601	0.601	1.000
H_2O/Na^+	TIP4P-FQ/QM	-1.165	0.582	0.582	1.000
	QM	-0.991	0.499	0.499	0.993

Table 4.7: Point charges determined for systems optimized with the TIP4P-FQ solvent model, using either MM or QM to describe the bound ion, compared to CHELPG charges from full QM minimizations. All QM calculations used the B3LYP DFT functional, with the TZVP+ basis for water and F^-, TZVP for Cl^-, and DZVP for Na^+.

optimization are displayed in comparison with the FlucQ charges for the same H_2O/Cl^- system. It can be seen that FlucQ reproduces the direction of the H–H polarization, although it does rather over-predict its magnitude, an effect seen previously with the water dimer. Charge transfer is obvious in this system, with $0.12e$ being transferred from the chloride ion to the water in the QM treatment, an effect that FlucQ cannot reproduce.

4.4.2 Water/fluoride

The SPC-FQ water model has been used to study water/fluoride clusters [202, 228]. Again, a near-linear O–H–F hydrogen bond is predicted, with the complex adopting C_s symmetry.

OPLS parameters [11] were used for the fluoride van der Waals interaction, due to the lack of such parameters in the CHARMM force field. Minimized structures were obtained as for the H_2O/Cl^- system, using B3LYP/TZVP+ for QM simulation. FlucQ has been noted to over-polarize in the presence of the very compact F^- ion [228],

Model	ΔE_{bind} (kcal/mol)	$d(\mathrm{O}\cdots\mathrm{F}^-)$ (Å)	$\angle$O–H$\cdots$F$^-$ (degrees)
TIP3P/MM	-25.34	2.54	159.1
TIP4P/MM	-27.08	2.47	169.7
TIP4P-FQ/MM[a]	-39.07	2.39	179.7
TIP3P/QM	-24.63	2.59	151.4
TIP4P/QM	-24.82	2.59	163.4
TIP4P-FQ/QM[a]	-38.41	2.48	179.0
QM	-36.29	2.43	177.4
Experiment[b]	-23.3		

[a]SHAKE [114] constraints were necessary for successful optimization.
[b]See Arshadi *et al.* [230]

Table 4.8: Determined energetic and geometrical properties for the H_2O/F^- system. QM calculations were carried out at B3LYP/TZVP+ level.

and this can be seen from the highly over-predicted binding energies in table 4.8 and the large charges in table 4.7. In fact, the F^- so polarized the TIP4P-FQ water that SHAKE constraints were required on the O–H bonds and angles; without these, significant lengthening of the O–H bond nearer the fluoride ion occurred, causing a larger hydrogen charge, which in turn further lengthened the bond, and so on, leading to a 'polarization catastrophe' [229]. This effect is due to a combination of too strong a polarization effect, and a lack of van der Waals parameters for the hydrogen atom. Even with SHAKE applied, the FlucQ charges are much larger than the CHELPG; in particular, the hydrogen nearer the fluoride is given a charge of greater than $1.0e$, which would correspond to production of an H^+ ion. However, the FlucQ binding energies and geometries are in very reasonable agreement with the full QM treatment. They do overestimate the experimental binding energy of 23.3 kcal/mol [230], which the TIP3P and TIP4P models reproduce rather well, particularly in QM/MM simulations, although a recent theoretical study [231], using Dang's dipole polarizable water model [29], has called this value into question, claiming that a binding energy 20% greater than the experimentally-claimed value would be more reasonable.

Model	ΔE_{bind} (kcal/mol)	$d(\mathrm{O}\cdots\mathrm{Na}^+)$ (Å)	$\angle$H–O$\cdots$Na$^+$ (degrees)
TIP3P/MM	-26.12	2.22	144.7
TIP4P/MM	-23.69	2.24	144.1
TIP4P-FQ/MM	-22.94	2.27	144.2
TIP3P/QM	-26.19	2.22	144.8
TIP4P/QM	-23.75	2.24	144.2
TIP4P-FQ/QM	-20.42	2.28	144.1
QM	-24.39	2.23	127.5
Experiment[a]	-24.0		

[a]See Džidić and Kebarle [232]

Table 4.9: Determined energetic and geometrical properties for the H_2O/Na^+ system. QM calculations used the DZVP basis set for Na^+, and TZVP+ for water, and utilised the B3LYP functional.

4.4.3 Water/sodium

For simulation of the H_2O/Na^+ system, standard CHARMM van der Waals parameters [10, 225] were used for the sodium ion, while the DZVP basis was used in QM simulations to model the ion, as TZVP+ parameters were not available for this atom type. The complex adopts a simple C_{2v} geometry, with Na^+ bound to the water oxygen and lying on the H–O–H bisector, opposite the hydrogen atoms. Table 4.9 shows that TIP4P gives binding energies in very close agreement with theory and experiment, while TIP3P tends to overestimate it and under-predict the O–Na distance. The reverse is the case for TIP4P-FQ. However, all the models give reasonable agreement; they all overestimate the H–O–Na angle to roughly the same extent, due to the MM treatment yielding a smaller water H–O–H angle than the QM.

In summary, the TIP4P-FQ method reproduces water-ion energetic and geometrical properties at a similar or greater quality to TIP3P and TIP4P, exhibiting a high degree of consistency between MM and QM/MM simulations. The H–H polarization effect of the water molecule is also reproduced by FlucQ, although its magnitude does not exactly match that given by QM methods. A more rigorous test of the technique

involves moving beyond simple gas-phase minimizations to simulation of species in bulk solvent.

4.5 Solvated QM/MM water

Several researchers have investigated the utility of combined QM/MM methods by applying them to the solvation of water itself. A single water molecule is designated as the solute, and is treated with quantum mechanics, while the remainder of the system acts as the solvent, and is treated classically. Wei and Salahub [220] used both DZVP and TZVP+ basis sets, and the LDA and nonlocal Perdew [233] DFT functionals, for the QM region, evaluating the energy of configurations obtained from a classical simulation using the SPC water model. Tuñón *et al.* [234] extended on this by using the full QM/MM energy to generate their trajectories. Similar studies have also been carried out using QM/MD [235] and QM/MC [236].

The ability of the TIP4P-FQ model to accurately reproduce the solvation environment of a single water was investigated in this work by treating a single water molecule at B3LYP/TZVP+ level, due to the proven accuracy of this method in reproducing the dipole moment of the gas phase monomer (see table 4.2). This water molecule was constrained to experimental geometry, and allowed to interact with 255 rigid TIP4P-FQ water molecules. Standard TIP3P van der Waals parameters were used for the QM/MM interaction.

Before simulation, the system was equilibrated by a procedure very similar to that used in section 4.2.1. A pre-equilibrated box of 256 TIP4P waters was used; simple cubic boundaries, with side length 19.716 Å, were imposed, and electrostatic forces were switched to zero between 7.0 and 9.5 Å. A single water near the centre of the periodic box was chosen to act as the solute, and was simulated during the equilibration phase with the TIP3P model and an MM force field. The FlucQ polarization energy was minimized, and the system then equilibrated, at a Cartesian temperature of 298 K

and a charge temperature of 1.0 K, for 50 ps. The solute water was then treated with DFT rather than TIP3P, and a further 10 ps of production QM/MM dynamics were carried out, with starting velocities taken from the final step of the equilibration run. This yielded a trajectory of 10000 configurations, for each of which the dipole of the QM water was determined from the QM wavefunction.

For comparison, the simulation was repeated using the standard TIP4P water model for the solvent rather than TIP4P-FQ. A third simulation removed the M sites from the solvent molecules, and treated the solvent with the TIP3P model, for comparison with previous work [234–236].

As previously mentioned, the three full QM/MM simulations all used standard TIP3P van der Waals parameters for the QM/MM interaction. Researchers have commented on the need for optimized van der Waals parameters for QM/MM interactions [17, 235, 237]. Such optimization was not carried out for this system; instead, three further simulations were carried out, in which the production trajectory was derived from a standard MM force field. Once again, a single water molecule was chosen to act as the 'solute'; however, it was treated with the same solvent model as the other 255 solvent molecules. Otherwise, the initial 50 ps equilibration was identical. This equilibration was then followed by 50 ps of production MM dynamics; at every 5th time step, Cartesian and charge coordinates were saved, to again yield a trajectory of 10000 configurations. The energy of each of these configurations was then reevaluated using QM for the solute water, the MM charges and the M site for this water first having been removed. (This took approximately 24 hours of walltime when spread over 4 800 MHz Pentium III processors.)

The advantage of the latter method for deriving the trajectory is that the first solvation sphere of the solute is likely to be well reproduced, as the solvent models are parameterized to yield accurate O–O radial distribution functions [30]. The error in the mean dipole is also likely to be lower, as longer simulations can be carried out (in

Water model	QM/MM trajectory	MM trajectory
TIP3P	2.817	2.575
TIP4P	2.551	2.513
TIP4P-FQ	2.701	2.643

Table 4.10: Average QM water dipoles, in Debye, for simulations of the B3LYP/TZVP+ water monomer solvated by 255 MM waters, using the TIP3P, TIP4P and TIP4P-FQ water models. 10000 configurations were generated by dynamics using either QM/MM or MM treatments.

this case, 50 ps compared to 10 ps) such that adjacent configurations in the generated trajectory are far less correlated with each other. Additionally, only the QM/MM *energy* needs to be evaluated for each configuration; first derivatives are not required, which saves computer time. However, the disadvantage is that the solvent cannot relax in response to fluctuations in the electron density of the solute molecule, as these fluctuations are not known until after the trajectory is generated.

The mean QM dipole moments for the TIP3P, TIP4P and TIP4P-FQ water models, from both the QM/MM and MM trajectories, are shown in table 4.10, while the distribution functions for solvent water oxygens around the single solute (QM) oxygen for the three solvent models are shown in figures 4.17, 4.18 and 4.19 respectively.

Looking firstly at the TIP3P data, it can be seen that the dipole obtained from the MM trajectory (2.575 D) agrees rather well with the study by Wei and Salahub [220], which obtained 2.47 D. (The discrepancy is explained by their use of the SPC water model rather than TIP3P, and a different, Perdew, functional [233].) When QM/MM is used to generate the trajectory, the solvent is able to relax in response to the QM water's electron density, and a much larger dipole, of 2.817 D, is obtained. This in turn is in good agreement with the DFT QM/MM study of Tuñón *et al.* [234], which predicted 2.89 D. This latter value is rather larger than the accepted range for the solution phase dipole (around 2.5–2.7 D), and figure 4.17 demonstrates that the first peak of the O–O RDF is rather poorly reproduced by the QM/MM trajectory. (Tu and Laaksonen [235] suggested that this was caused by poor Lennard-Jones parameters for the QM/MM

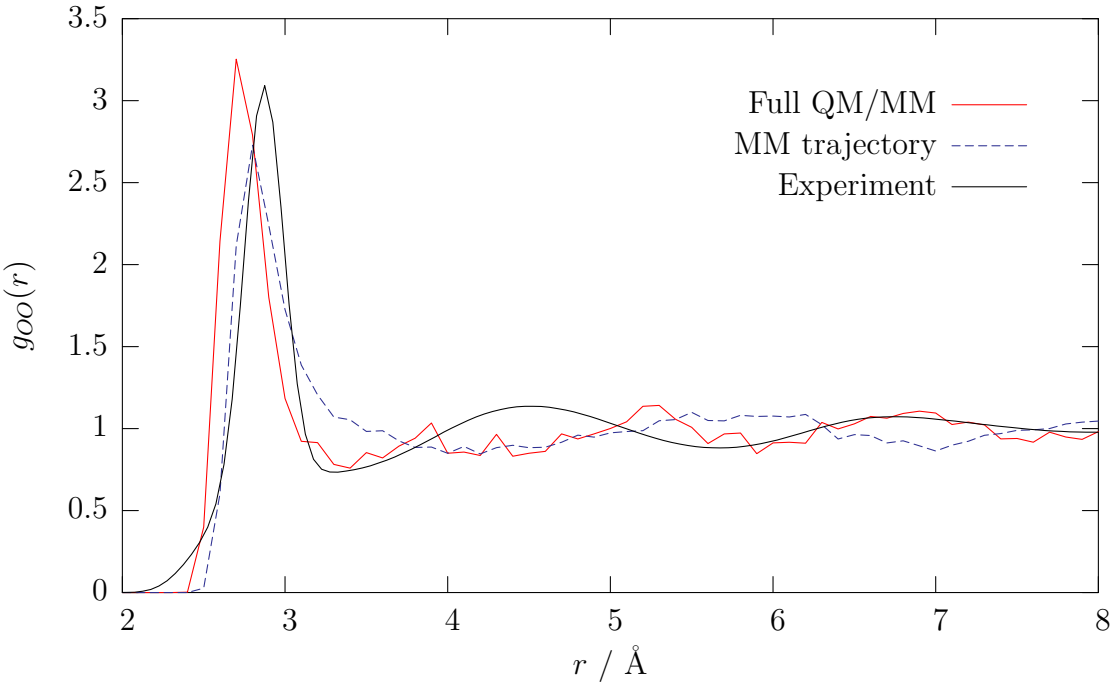

Figure 4.17: Quantum oxygen-solvent oxygen radial distribution functions, for full QM/MM simulation or reevaluation of the energies of an MM trajectory, compared to experimental data. The QM water was solvated by 255 TIP3P molecules.

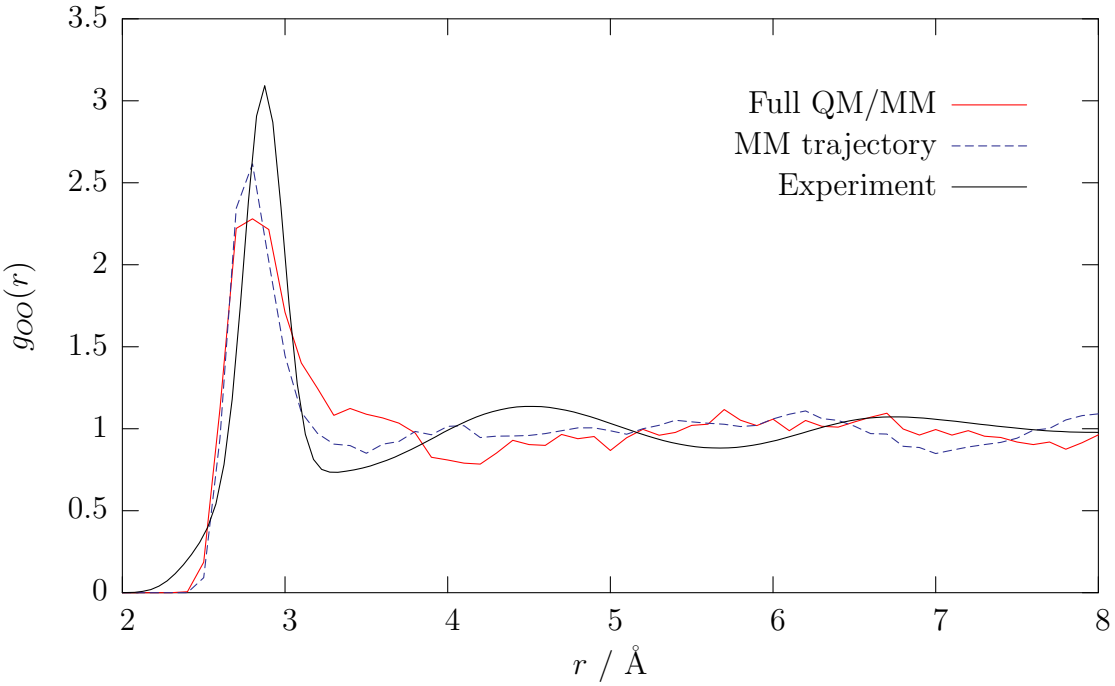

Figure 4.18: Quantum oxygen-solvent oxygen radial distribution functions, for full QM/MM simulation or reevaluation of the energies of an MM trajectory, compared to experimental data. The QM water was solvated by 255 TIP4P molecules.

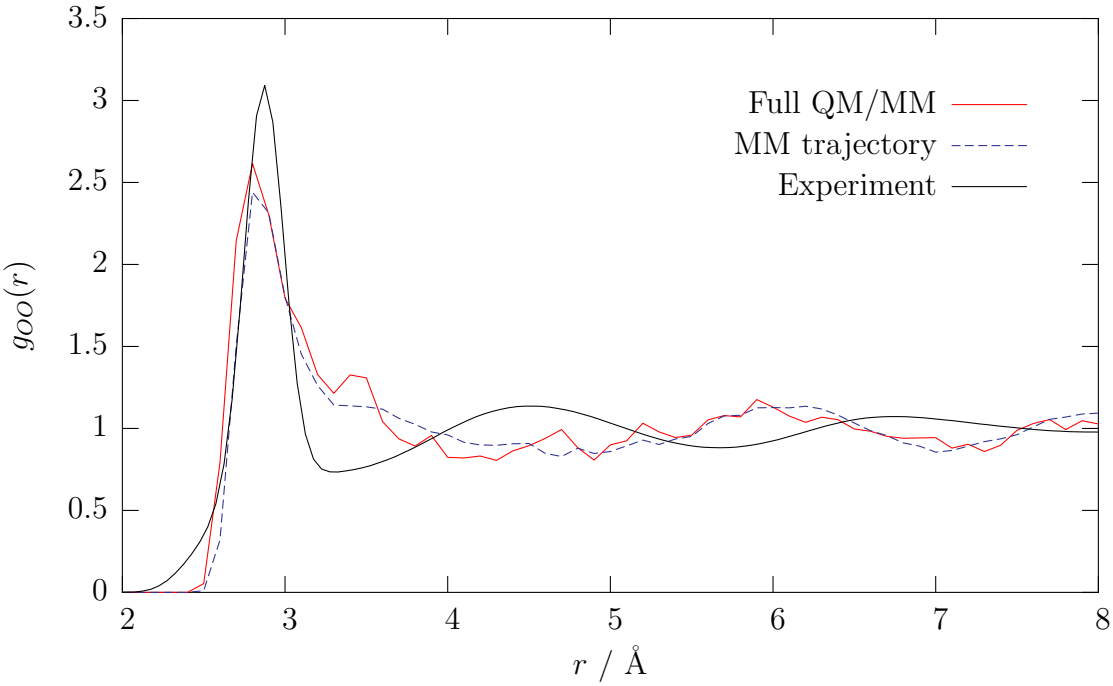

Figure 4.19: Quantum oxygen-solvent oxygen radial distribution functions, for full QM/MM simulation or reevaluation of the energies of an MM trajectory, compared to experimental data. The QM water was solvated by 255 TIP4P-FQ molecules.

interaction, and applied parameters optimized for the water dimer interaction to this system; however, while this gave a slightly more reasonable distribution function, the dipole was underestimated, as 2.49 D.)

Both TIP4P and TIP4P-FQ give much more consistent behaviour between the MM and QM/MM simulations; it can be seen from figures 4.18 and 4.19 that MM and QM/MM give the first O–O peak in very similar places (little information can be gleaned from the RDF at large r, however, as the simulations were too short to adequately sample the second and higher solvation spheres). Again, relaxation of the solvent leads to a higher solute dipole in the QM/MM simulations. TIP4P-FQ gives consistently higher dipoles than TIP4P, due to the higher average solvent dipole — the fixed-charge TIP4P model results in a solvent molecular dipole of 2.18 D, while TIP4P-FQ, as seen in section 4.2.1, averages around 2.6 D. Thus, the FlucQ method is particularly suitable for this system, giving excellent radial distribution functions

and molecular dipoles that are consistent between solvent and solute, and in good agreement with other work.

4.6 Parameterization of the FlucQ model

It was seen in section 4.3.4 that the TIP4P-FQ model, although it is in reasonable agreement with the *total* binding energy as predicted from high-level DFT calculations, overestimates the electrostatic component of this interaction. This was also manifested by the rather large dipole obtained for the water dimer system and the overly large H–H polarization in the donor water. A new model was parameterized by fitting to the QM energy curve obtained in section 4.3.3 and shown in figure 4.15, to explore the possibility that FlucQ could accurately reproduce the CHELPG charges of the *ab initio* treatment.

The FlucQ model was first fitted to the electrostatic component of the interaction. At each point on the dimer-dimer interaction surface, characterised solely by the O–O distance r, the FlucQ charges were determined by dissipative Langevin minimization and used to calculate the electrostatic interaction, $V_{elec}^{FQ}(r)$. The gas-phase dipole, μ_{FQ}, was determined by a similar charge minimization of the experimental geometry water monomer. These data were then compared to the QM electrostatic interaction, $V_{elec}^{QM}(r)$, and the experimental gas phase dipole, μ_{exp} (1.85 D), to evaluate a cost function:

$$\chi^2 = \sum_r [V_{elec}^{FQ}(r) - V_{elec}^{QM}(r)]^2 + W_{gas}(\mu_{exp} - \mu_{FQ})^2 \tag{4.3}$$

The cost function was then minimized by a grid search of the FlucQ parameters $\tilde{\chi}_O - \tilde{\chi}_H$, ζ_O and ζ_H. The weight W_{gas} was chosen as 1000 (kcal/mol/D)^2, such that the resulting parameters reproduced the gas phase dipole adequately. With the best-fit FlucQ electrostatic interaction thus determined, a Lennard-Jones function was fitted to the difference between this and the QM full interaction energy, to determine the van der

	TIP4P-FQ1	TIP4P-FQ
$\tilde{\chi}_O - \tilde{\chi}_H$ (kcal/mol/e)	145.0	68.49
ζ_O (a_0^{-1})	0.52	1.63
ζ_H (a_0^{-1})	1.26	0.90
ϵ (kcal/mol)	0.90	0.2862
σ (Å)	2.94	3.159
M_Q (ps/e)^{2}kcal/mol	9.0×10^{-5}	6.0×10^{-5}

Table 4.11: Determined FlucQ and Lennard-Jones parameters for the TIP4P-FQ1 model, compared to those used for TIP4P-FQ.

	q_{H1} (e)	q_{H2} (e)	ΔE_{bind} (kcal/mol)	d(O$\cdots$R) (Å)
H_2O (R = O)	0.487	0.429	-4.94 (-4.46) [-5.4]	2.96 (2.93) [2.98]
Cl^-	0.601	0.390	-15.10 (-17.55) [-13.4]	3.08
F^{-a}	0.701	0.349	-27.74 (-39.07) [-23.3]	2.51
Na^+	0.518	0.518	-20.75 (-22.94) [-24.0]	2.31

[a]SHAKE constraints were necessary for successful optimization.

Table 4.12: Calculated charges, binding energies, and interatomic distances for the MM water dimer and water/ion complexes, using the TIP4P-FQ1 model. TIP4P-FQ and experimental data (from tables 4.3, 4.6, 4.8 and 4.9) are shown for comparison, TIP4P-FQ in round brackets and experimental in square brackets.

Waals interaction parameters ϵ and σ. The resulting model was termed TIP4P-FQ1, and its parameters are shown in table 4.11, compared to those of the existing TIP4P-FQ model.

The TIP4P-FQ1 model reproduces the total and electrostatic energy profiles of the dimer interaction rather well, as can be seen from figure 4.20. It also gives reasonable MM energies and geometries for the minimum-energy dimer and water-ion complexes, as evidenced by table 4.12. The binding energies of all but the sodium complex are in better agreement than TIP4P-FQ with experimental data, and additionally the induced hydrogen charges, while still larger than those predicted by QM in tables 4.3 and 4.7, show much less H–H polarization and thus generally agree better than the TIP4P-FQ charges for the same systems.

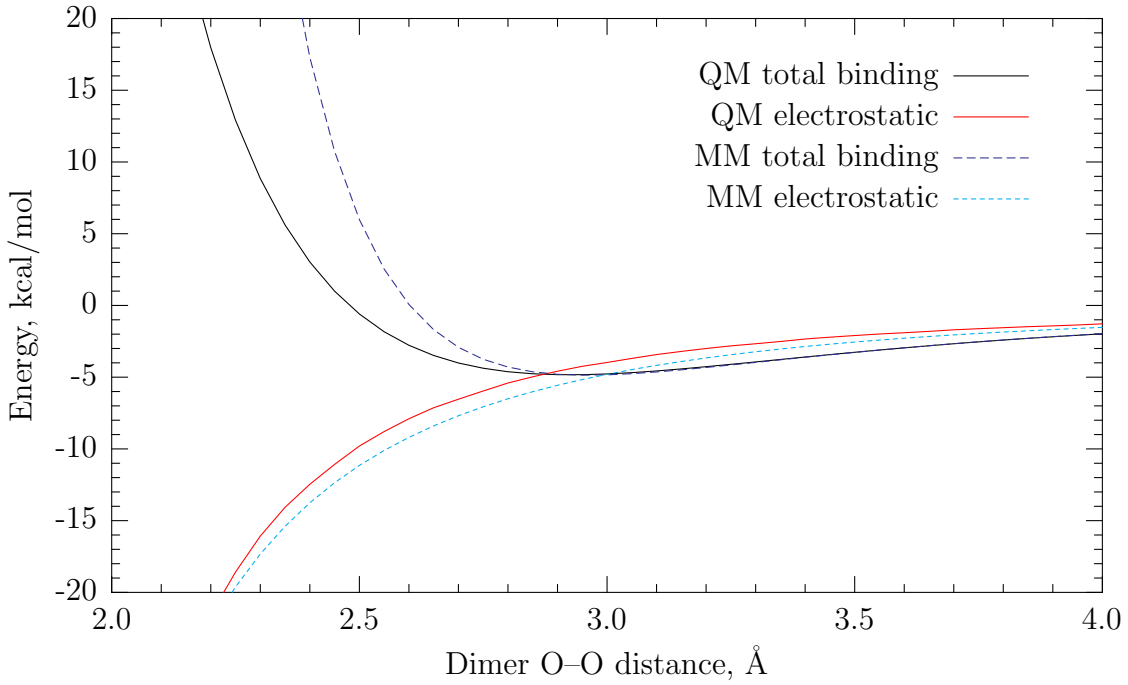

Figure 4.20: Agreement between the water dimer binding energy calculated at B3LYP/TZVP+ level (QM) and with the TIP4P-FQ1 model (MM). The electrostatic component of the interaction is also shown.

The TIP4P-FQ1 model was also tested on bulk solvent, for which a larger charge mass (see table 4.11) than used for TIP4P-FQ, 9.0×10^{-5} $(\mathrm{ps}/e)^2$kcal/mol, was required to prevent too-large charge temperature fluctuations. However, the results were rather disappointing; simulations produced average water dipoles of around 2.1 D, far lower than the 2.6 D obtained from TIP4P-FQ. Thus, it was decided to use the standard TIP4P-FQ model for further studies, despite its problems in reproducing the exact electrostatic interaction. The charges from the QM treatment are in any case method-dependent, as they are not observables, and so cannot be totally relied upon; other methods such as MK [101] may give more reliable charges than CHELPG, while a recent study using dipole-fitted charges [238] predicted a much larger H–H polarization than the QM method used in this work. Other studies [239] suggest that point charges cannot in any case accurately reproduce the electrostatic potential at typical H-bond distances. Thus, further study in this area is warranted.

4.7 Summary

The FlucQ model gives excellent properties for bulk solvent using either MD or MC methods. Additionally, it correctly models the gas phase water monomer, and gives a much better fit to the water dimer interaction energy than static charge treatments. It also performs well for simulating water-ion interactions, in both the conventional MM and the QM/MM arenas. It handles the bulk solvation of a single 'solute' water molecule particularly well, giving a consistent solute and solvent dipole, and reasonable agreement between properties obtained from MM and QM/MM trajectories.

FlucQ is not without its flaws. The standard TIP4P-FQ model over-predicts the water dimer dipole, although it gives an average bulk solvent dipole in reasonable agreement with the accepted value. TIP4P-FQ1, on the other hand, performs much better for the dimer but rather poorly for the bulk. Thus, it seems probable that the FlucQ model in its current form cannot adequately model both bulk solvent and biomolecular complexes. In fact, Liu *et al.* [240] found that it was necessary to introduce a large number of extra parameters to reproduce the *ab initio* binding energies of water trimers in constructing their '*ab initio* FlucQ' model, although no explicit consideration of atomic charges was made in their study. Stern *et al.* [241] further showed that the FlucQ model is not able to accurately model the response to all electric fields, and asserted that the addition of off-nucleus charges or dipoles would be necessary for a full treatment.

Overall, the FlucQ method gives excellent solvent behaviour, and QM/MM interaction energies, for the systems studied, at least as good as with the TIP3P and TIP4P models. This provides strong evidence for the use of FlucQ in QM/MM simulations.

Chapter 5

Application to the Menshutkin reaction

5.1 Introduction

In order to rigorously test the combination of established and new methods, free energy calculations were carried out. These also make large demands on the computational infrastructure developed in Chapter 2, due to the level of theory required to obtain accurate single-point energies, and the large number of configurations needed to achieve reliable free energies. The system studied in this chapter is the Menshutkin reaction between methyl chloride and ammonia, a simple and computationally tractable S_N2 reaction that has been the subject of a number of theoretical studies. A variety of methods are used to study the gas and solution phase reactions, and the FlucQ model employed to investigate the effect of solvent polarization.

5.1.1 S_N2 reactions

Bimolecular nucleophilic substitution, or S_N2, reactions have been the subject of many experimental and theoretical studies [242, 243]. A complete understanding of

their mechanism is of immense value, due to their involvement in many fundamental chemical processes. An accurate treatment of solvent is an essential prerequisite for such investigations, as many S_N2 reactions show a strong dependence on the nature of the solvent [244, 245].

5.1.2 Type I S_N2 reactions

The energy surface for an S_N2 reaction in the gas phase can be obtained by quantum mechanical treatments, and several studies have met with success in this area [246]. A number of researchers have considered the S_N2 reaction between chloride and methyl chloride:

$$Cl^- + CH_3Cl = ClCH_3 + Cl^- \tag{5.1}$$

This is a 'Type I' S_N2 reaction, where a negative nucleophile attacks a neutral group, as classified by Ingold [242]. The symmetry of the reaction halves the computer time required to carry out the simulation, as the departure of the Cl^- leaving group is exactly analogous to the initial chloride attack.

The addition of solvent does, however, increase the difficulty of determining the energetics tremendously; the reacting system may be as small as half a dozen atoms in the gas phase, but to adequately solvate it, up to a radius of 10 Å or so, would require the addition of hundreds of solvent molecules. Obviously, if the system were treated quantum mechanically, this would result in a large increase in the computer time required. Several approaches have been used to alleviate this problem; an early method, used by Morokuma [247], approximated the solvent environment by a hydrated cluster:

$$Cl^-(H_2O)_n + CH_3Cl = ClCH_3 + Cl^-(H_2O)_n \tag{5.2}$$

where n was varied from 0 (for the gas phase reaction) to 2. The energy surface was then determined by *ab initio* methods at the 3-21G level. This followed an experimental

study by Bohme *et al.* [248] which investigated the similar reaction between hydrated hydroxide ions and methyl bromide.

The cluster approximation is not ideal, as the reaction is dominated by the transfer of solvent molecules between the nucleophile and leaving group. In a complete solvent treatment, all of the reacting molecules are solvated throughout the reaction. Additionally, the rearrangements of the bulk solvent beyond the first solvation sphere can contribute to the reaction energy. Later studies, such as that of Chandrasekhar *et al.* [249–251], therefore considered the solvent explicitly. Their calculations were made computationally tractable by using the gas phase energy surface to parameterize a molecular mechanics model for the reacting system; the solution phase surface was then determined by a purely molecular mechanical simulation, with explicit solvent. More recent simulations have used the Molecular Orbital-Valence Bond method [252] to treat the QM region [253].

5.1.3 Type II S_N2: The Menshutkin reaction

A Menshutkin reaction is a Type II S_N2 reaction (meaning that both of the reagents are neutral [242]) of the form

$$XR + NR'_3 = X^- + NR'_3R^+ \tag{5.3}$$

which proceeds via nucleophilic attack of the nitrogen on the R moiety, followed by expulsion of the halide X^-. This follows from Menshutkin's original study [254, 255], where measurements were taken of the rate of the reaction between triethylamine and ethyl iodide,

$$(C_2H_5)_3N + C_2H_5I = (C_2H_5)_4N^+ + I^- \tag{5.4}$$

The reaction was carried out in a variety of solvents, and Menshutkin noted that the identity of the solvent had a large effect on the reaction rate, the more polar solvents

resulting in higher rates. This sensitivity to solvent polarity makes the Menshutkin reaction a useful test case for any solvent model. The reaction is strongly endothermic in the gas phase, and so has been studied experimentally only on sterically hindered amines, with molecular beam techniques [256]. However, a polar solvent can form a solvent cage around each of the ions, thus stabilizing the products; in fact, in water the forward reaction is strongly preferred. This is in contrast to other types of S_N2 reaction, which are slowed by more polar solvents.

As the Menshutkin reaction is asymmetric, unlike the Type I reactions considered earlier, the *position* of the transition state can also be altered by the polarity of the solvent, as suggested by the Hammond postulate [243, 257]. This is in fact observed in theoretical studies of the reaction [258].

The Menshutkin reaction has attracted a good deal of experimental and theoretical interest [259–264]. Computational studies of the reaction often target the reaction of methyl chloride with ammonia [258, 265–267], due to the small number of atoms involved in the reaction, and the availability of semi-empirical parameters for chlorine:

$$CH_3Cl + NH_3 = Cl^- + NH_3CH_3^+ \tag{5.5}$$

(The drawback of such studies is that experimental data are only available for the analogous methyl iodide reaction [268, 269], so direct comparisons with experiment are not possible.) As with the chloride-methyl chloride reaction, a number of studies have approximated the solvent. For example, a recent study of the ammonia/methyl bromide reaction [270] used the Effective Fragment Method [271, 272] in conjunction with an *ab initio* representation of the solute at the DZVP level. An earlier study of the same reaction by Solà *et al.* [273] approximated the solvent by two explicit waters or with the PCM implicit solvent treatment [126], and used a 3-21G basis set, augmented with polarization and diffuse functions to properly describe the bromine atom. PCM has also been used in very recent studies in conjunction with large basis set DFT (6-31+G*

with B3LYP or B3PW91) to look at a variety of Menshutkin reactions in solvent [274] or to study the ammonia/methyl chloride reaction within carbon nanotubes [267]. Similar studies have moved beyond DFT to MCSCF [275], or have replaced the PCM model with the similar GCOSMO [276] method, using DFT and MP2 to evaluate energies [266].

The QM/MM method (section 1.4) is ideally suited to simulation of the Menshutkin reaction; the small reacting region can be treated with QM, and the solvent with MM. A study of the ammonia/methyl chloride reaction was carried out by Gao *et al.* [258,265], using AM1 for the reacting region and the TIP3P model for the solvent. The C–Cl and C–N distances were varied independently to give a two-dimensional map of the free energy surface.

Another approach involves approximation of the solute region, similar to that used by Chandrasekhar *et al.* for the chloride-methyl chloride reaction. For example, one such study [277] used an *ab initio* gas phase energy surface to parameterize a molecular mechanics force field. Alternatively, the Empirical Valence Bond method [278], whereby the reactive region is treated by a combination of resonance structures in a modified molecular mechanical force field, was utilised in a study by Hwang *et al.* [279].

5.2 Adiabatic mapping of the reaction surface

In this work, the ammonia/methyl chloride Menshutkin reaction is examined in water with QM/MM techniques. This enables comparison with the study by Gao *et al.* [258,265] using similar methods. An ideal starting point for examination of the solution-phase reaction is the determination of the gas-phase energy profile. Due to the small size of the system (9 atoms) it was possible to utilise rather rigorous QM techniques for this process, with the adiabatic mapping method.

5.2.1 The adiabatic mapping method

The technique of *adiabatic mapping* [107] aims to determine the energy profile of a chemical reaction. It is, essentially, an application of the minimization methods discussed in section 1.9, with constraints added to the system. First, a reaction coordinate (as described in section 1.10.6) is chosen. The adiabatic map is then obtained by varying the reaction coordinate in steps so as to cover the entire range of the reaction, and at each step minimizing the energy of the system, subject to a constraint that fixes the reaction coordinate. This constraint is usually enforced by imposing a harmonic penalty function on the reaction coordinate, with its minimum at the desired reaction coordinate value.

It should be noted that the profile determined from an adiabatic mapping is of the energy, not of the free energy. The structures obtained are representative of the system at 0 K, and include no entropic contributions.

5.2.2 Choice of reaction coordinate

The Menshutkin reaction is a simple substitution reaction, and thus the obvious choice of reaction coordinate is a function of the distances of the entering and leaving groups from the reaction centre, i.e. the bond lengths r(C–N) and r(C–Cl), as shown in figure 5.1. In this work, the difference between these two bond lengths was used as the coordinate:

$$R_C = r(\text{C–Cl}) - r(\text{C–N}) \tag{5.6}$$

This is in keeping with previous studies [258, 273] and has been shown to be a suitably representative coordinate [266].

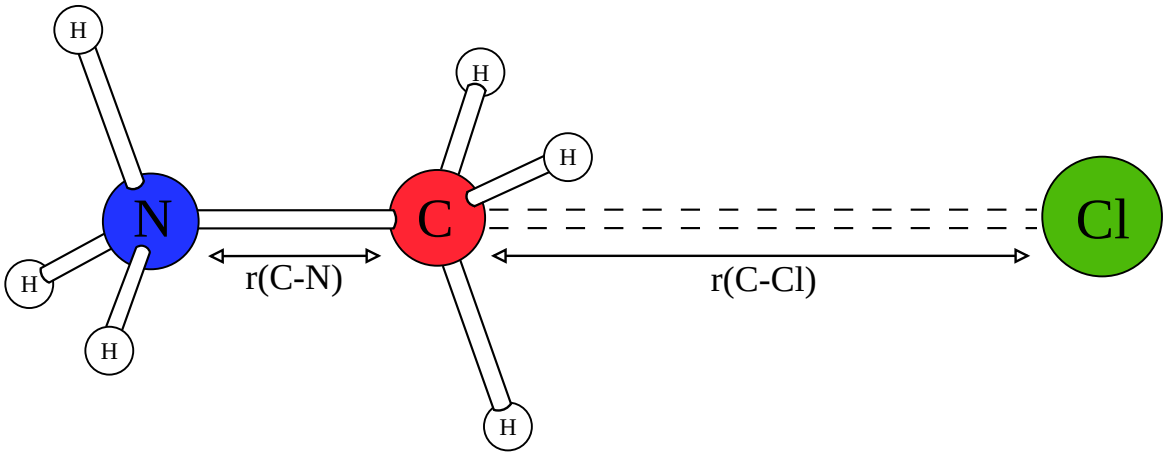

Figure 5.1: The reactive region in the Menshutkin reaction, demonstrating the choice of reaction coordinate, $R_C = r(\text{C–Cl}) - r(\text{C–N})$.

5.2.3 Comparison of methods in the gas phase

Adiabatic maps were obtained for the methyl chloride/ammonia Menshutkin reaction (equation 5.5) by varying the reaction coordinate from -2 to 4 Å in increments of 0.1 Å. At each point, the system energy was minimized by Adopted Basis Newton Raphson minimization until the forces dropped to below 0.005 kcal/mol/Å. The desired value of the reaction coordinate was obtained by imposing a harmonic constraint on R_C, after equation 1.130, of the form

$$E_{harm} = \frac{1}{2}K(R_C - R_{ref})^2 \tag{5.7}$$

where K is a force constant, and R_{ref} the reference value of the coordinate, corresponding to the minimum of the function. For this study, a force constant of 2000 kcal/mol/Å^2 was used. The N–C–Cl angle was maintained at linearity by imposition of a harmonic constraint with force constant 50 kcal/mol/radian^2. The assumption that the reaction proceeds via a perfectly linear approach of the ammonia or chloride nucleophile is used throughout this work, as in other studies [258, 277]. (Although it would be preferable to allow varying angles of attack, particularly in the

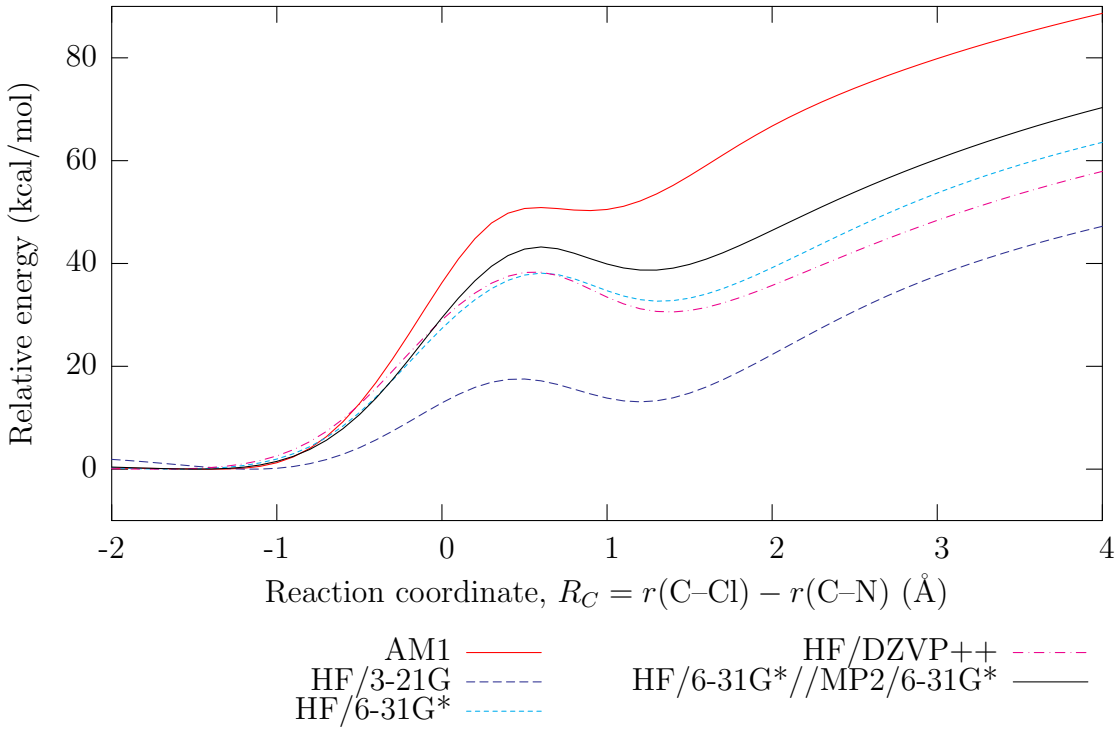

Figure 5.2: Adiabatic gas-phase maps of the methyl chloride-ammonia Menshutkin reaction at different levels of theory.

solution phase studies, the addition of these extra degrees of freedom would require more extensive sampling, and an unjustifiable increase in computer time.)

The resulting adiabatic maps are shown in figure 5.2. The semi-empirical AM1 method is compared to Hartree Fock treatments, ranging from the simple 3-21G basis through a basis enhanced with polarization functions (6-31G*) to one with both polarization and diffuse character, DZVP++. For comparison, the results from a post-Hartree Fock split-level technique are also shown; the energies of the structures minimized with the HF/6-31G* method were reevaluated at MP2/6-31G* level. This approach is denoted HF/6-31G*//MP2/6-31G*, and belongs to a class of techniques which is discussed in more detail in section 6.9.

It can clearly be seen from these adiabatic maps that a relatively high level of theory is required in order to determine accurate energies. The 3-21G basis, for example, grossly underestimates the energy required to reach the $[NH_3 \cdots CH_3 \cdots Cl]^{\ddagger}$

transition state, and predicts a shallow minimum at around R_C = -1 Å. Higher activation energies are obtained as the level of theory is improved. However, all of the *ab initio* methods underestimate the free energy of activation, which is estimated to be around 45 kcal/mol from simulations at MP4 level [266, 277]. (No experimental data are available for this activation energy itself; however, both of the cited simulations also predict free energies of *reaction* in close agreement with the experimental figure of 111 ± 5 kcal/mol [280].) It is likely that part of this discrepancy is due to the entropic contribution to the free energy; however, the AM1 *energy* of activation, of roughly 50 kcal/mol, comes remarkably close to this figure.

5.2.4 Solution-phase energies from continuum treatments

In order to get an idea of the profile of the reaction in solution, PCM [126] calculations were carried out (see section 1.10.5). For each configuration previously optimized at HF/DZVP++ level, the energy was reevaluated at HF/6-311++G** level, using PCM to approximate the surrounding solvent. The resulting adiabatic solution-phase map is shown in figure 5.3. This shows a number of striking differences to the gas-phase Hartree-Fock maps; the reaction is strongly exothermic, with a energy of about -30 kcal/mol, while the activation energy is reduced from around 40 kcal/mol in the gas phase to about 19 kcal/mol. The position of the transition state is also shifted from around 0.5 to 0.1 Å, as seen in previous studies [273] and as suggested by the Hammond postulate. This profile is in reasonable agreement with the available experimental data, coming close to the reaction free energy of -34 ± 10 kcal/mol (as estimated from standard free energies of hydration and formation [277]) and the activation energy of the analogous methyl iodide reaction, 23.5 kcal/mol [268, 269].

There are a number of deficiencies with this method. Firstly, because it uses gas-phase optimized geometries for the energy calculations, there is no opportunity for the solute to relax in response to the solvent. This is likely to be the cause of the

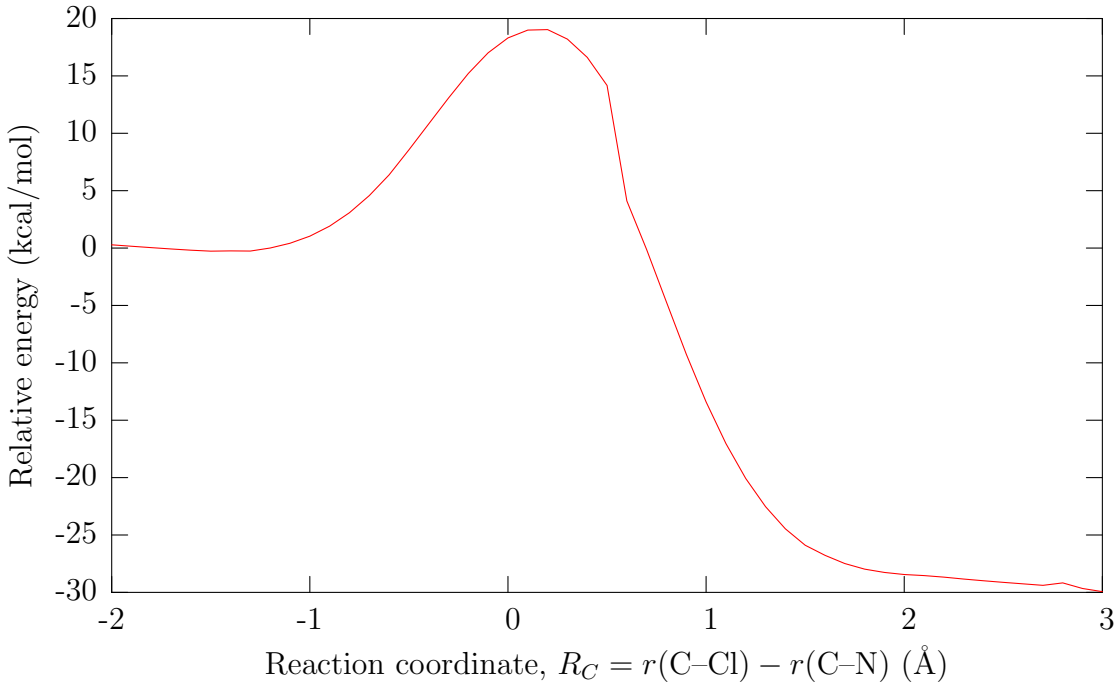

Figure 5.3: Adiabatic PCM map of the methyl chloride-ammonia Menshutkin reaction using DZVP++ optimized geometries. Energies were evaluated at HF/6-311++G** level.

drop in energy at around 0.5 Å, which corresponds to the position of the gas-phase transition state. Secondly, the lack of explicit solvent means that simple comparisons with the FlucQ model cannot be made. Finally, the energies, while they incorporate some approximate entropic contributions from PCM, are not true free energies as the gas-phase geometries correspond to 0 K. Although more reliable energies could be obtained by optimizing the geometries while using PCM, convergence is often difficult to achieve between the solute and cavity degrees of freedom.

The AM1 method has been shown to yield reasonable energies for the gas phase reaction, at a fraction of the computational cost of methods such as HF/6-31G, which actually would appear to yield less reliable energies. Previous studies [277] have shown that these trends transfer reasonably well from the gas to the solution phase, and results in good agreement with experiment are obtained using gas-phase geometries in combination with a solvent representation. For this reason, the AM1 method was

further employed to investigate the behaviour of the Menshutkin reaction in explicit solvent.

5.3 Hybrid QM/MD studies

5.3.1 Overview

The free energy profile of the Menshutkin reaction can be determined by umbrella sampling (see section 1.10.6), using molecular dynamics to generate configurations. Since the reaction involves the formation and breakage of C–Cl and C–N bonds, and changes in the charges of the species, a conventional molecular mechanics force field is unsuitable. Thus, a QM/MM treatment was used in this study, after Gao and Xia [258], using AM1 for the reactive region.

Although both periodic and stochastic boundary techniques could be applied to this system (see sections 1.7.1 and 1.7.3, respectively), the latter was used only for test simulations. In a stochastic boundary simulation, Langevin dynamics is used in the buffer region. While FlucQ will function in conjunction with Langevin dynamics, some analogue of the random and frictional Cartesian forces should also be applied to the charge degrees of freedom to represent the average effect on the charges of solvent in the reservoir region, and this is not present in the current model. Additionally, the periodic boundary method is easily applied to both molecular dynamics and Monte Carlo simulations, and was thus used so that results from this study could be easily compared with the Monte Carlo simulations in Chapter 6.

The nature of the QM/MM interaction when combined with periodic boundaries deserves some attention. The *ab initio* codes that are interfaced with CHARMM are done so in a very simple fashion, with the primary MM atoms appearing as point charges to the QM code; no account is made of cutoffs or periodicity. The semi-empirical interface implements cutoffs, but is still ignorant of any boundary conditions.

Thus, reasonable energies are only obtained while the QM region remains roughly near the centre of the primary simulation box; if it drifts from this region during a simulation then it will begin to see vacuum beyond the edge of the primary cell. Two solutions to this problem present themselves: the first fixes the centre of mass of the QM region at the origin, either explicitly or by redefining the coordinate system, while the second applies simple periodic boundaries to the QM/MM interactions, by taking each Cartesian component of any distance between the QM and MM regions modulo the relevant box length. The second approach is clearly the more rigorous (although only possible for simple types of repeating cell, such as cuboids) and was used for the QM/MM simulations in this chapter. (Incidentally, the same problem exists in stochastic boundary simulations, and is solved by using a force to constrain the QM region near the centre of the reaction region.)

No cutoffs were used for the QM/MM interaction, even for semi-empirical calculations, for consistency between *ab initio* and semi-empirical approaches. In any case, the lack of interactions with MM images ensured that no QM atom interacted with itself. Also, while the FlucQ cutoff test cases would recommend the use of a group-based force switch (section 4.2.1), CHARMM does not support group-based electrostatics for QM/MM simulation; atom-based cutoffs were therefore used for solvent-solvent interactions.

5.3.2 Parameterization of the QM/MM interaction

In many QM/MM simulations, such as those in Chapter 4, the van der Waals parameters used for the QM atoms are taken from the same force field used for the MM region. This is often a reasonable approximation; however, these parameters are designed to reproduce the average interaction between MM atoms, and as such may be unsuitable if the environment is unlike the average. For example, over the course of the Menshutkin reaction, the Cl atom moves from incorporation in a CH_3Cl molecule

to being a Cl^- ion, and it is unlikely that van der Waals parameters optimized for one extreme will be suitable for the other. Several researchers have commented on the need for specially optimized parameters for QM/MM interactions [17, 235] and have even obtained such parameters for some systems [237, 258].

The method used in this work to optimize the parameters used for QM/MM simulation is similar to that used by Gao *et al.* [258]. Three complexes in which a single water molecule is hydrogen-bonded with the reactive region were considered. In complex 'a' the water oxygen interacts with an amine hydrogen, while in complexes 'b' and 'c' it interacts with the chlorine atom, in the open form in 'b' and the bifurcated form in 'c'. These three complexes are depicted in figure 5.4. (Interactions with the methyl hydrogens were not considered in this work, as previous studies [258] suggested that these interactions are weak.) Each of the complexes was considered for reactive region configurations optimized at HF/6-31+G* level, obtained from the adiabatic sampling simulation (section 5.2.3), for reaction coordinate values from -1.6 to 1.6 Å inclusive, spaced by 0.2 Å.

For each complex, the geometry of the reactive region was fixed, while the single water molecule was fixed at experimental geometry, and the system was then optimized at HF/6-31+G* level by conjugate gradient minimization, until the forces dropped to below 0.05 kcal/mol/Å. This served to optimize the hydrogen bond distance (and, for complexes 'b' and 'c', the angle) and to obtain a solvent-solute binding energy, for each configuration and complex. The resultant binding energies are shown in figure 5.5. Note that the interaction with solvent increases for each complex as the reaction moves from the electrically neutral ammonia and methyl chloride starting materials to the charged products, methyl ammonium and chloride ion. Also, the open form (complex 'b') for the water-chlorine interaction is fractionally more stable than the bifurcated, presumably due to weaker chlorine-oxygen repulsion in the former case.

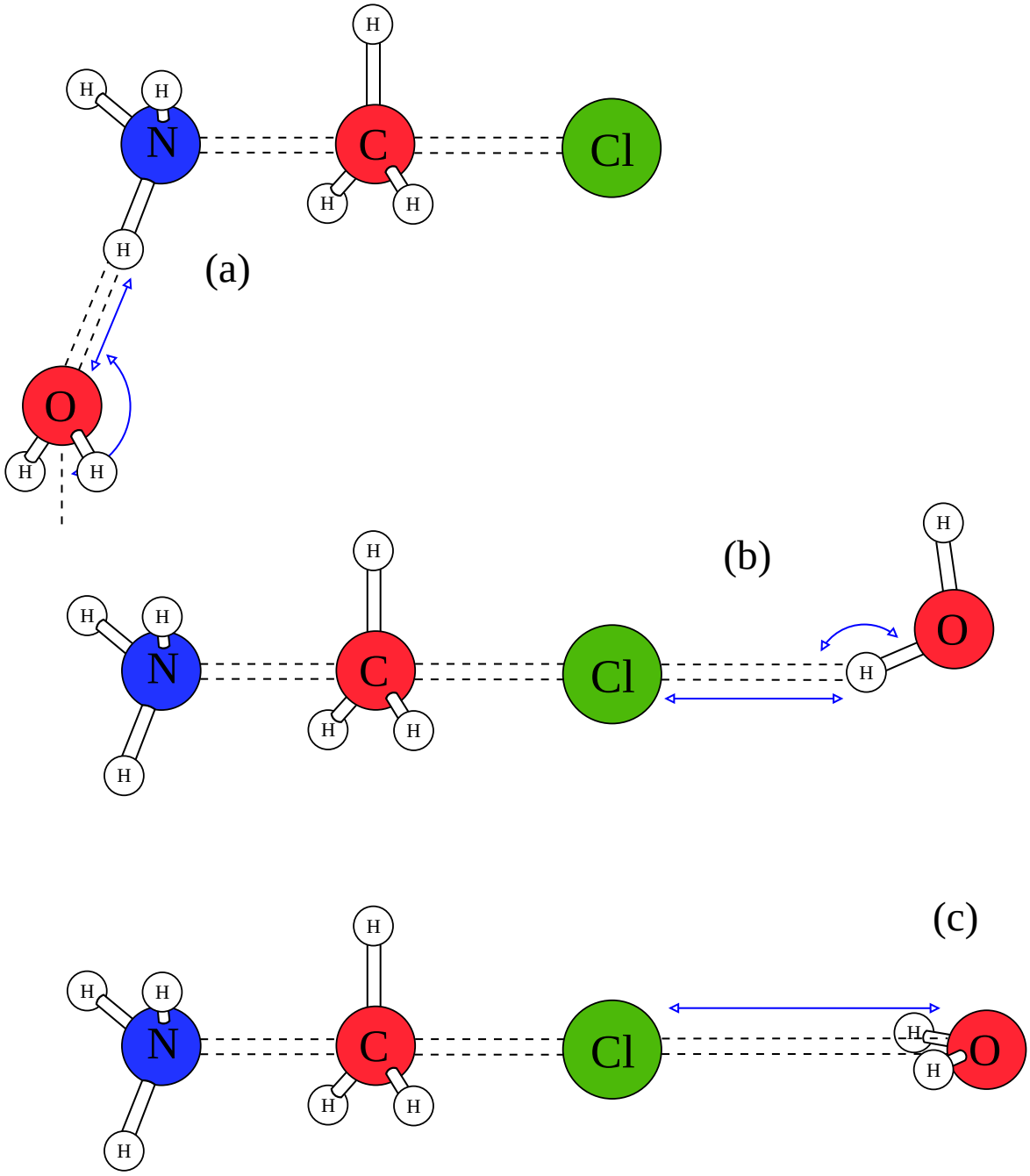

Figure 5.4: Hydrogen-bonded complexes used for optimization of the QM/MM interaction. Interatomic distances and rotation angles that were optimized are shown in blue.

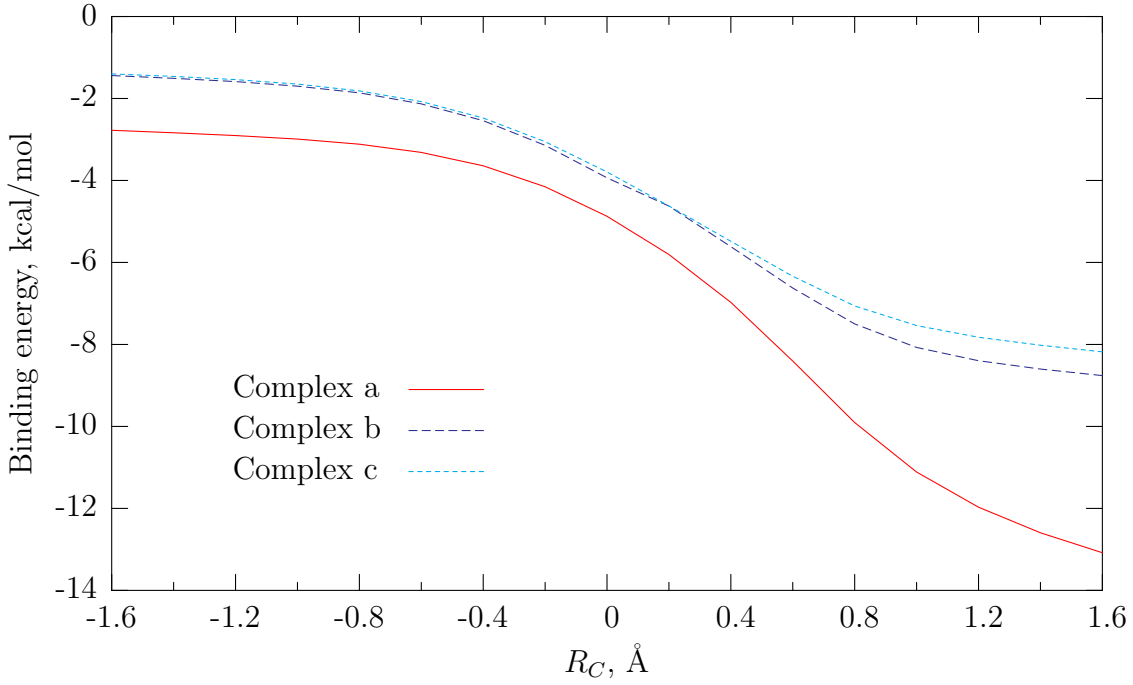

Figure 5.5: Water-solute binding energies for the hydrogen-bonded complexes shown in figure 5.4, optimized at HF/6-31+G* level at various points along the Menshutkin reaction coordinate, R_C.

The optimizations were repeated for the same reactive region configurations and complexes, using the same constraints, but using the QM/MM method, with AM1 for the reactive region and the TIP3P water model for the solvent. Gao's previously determined Lennard-Jones parameters [258] were used for the QM atoms, using the combining rules in section 1.3 to derive the van der Waals part of the QM/MM interaction. This yielded another set of binding energies; in figure 5.6 these QM/MM energies are compared to the corresponding *ab initio* energies. The agreement can be seen to be good, with an RMS error in the energies of 0.49 kcal/mol, and in the distances (N–O and Cl–O) of 0.16 Å.

Ideally, the van der Waals parameters used for QM/MM interactions of the Menshutkin reactive region with TIP3P solvent would be applicable to simulations with other water models as well. In practice, however, because MM force fields do not treat electron density explicitly, the van der Waals parameters include some

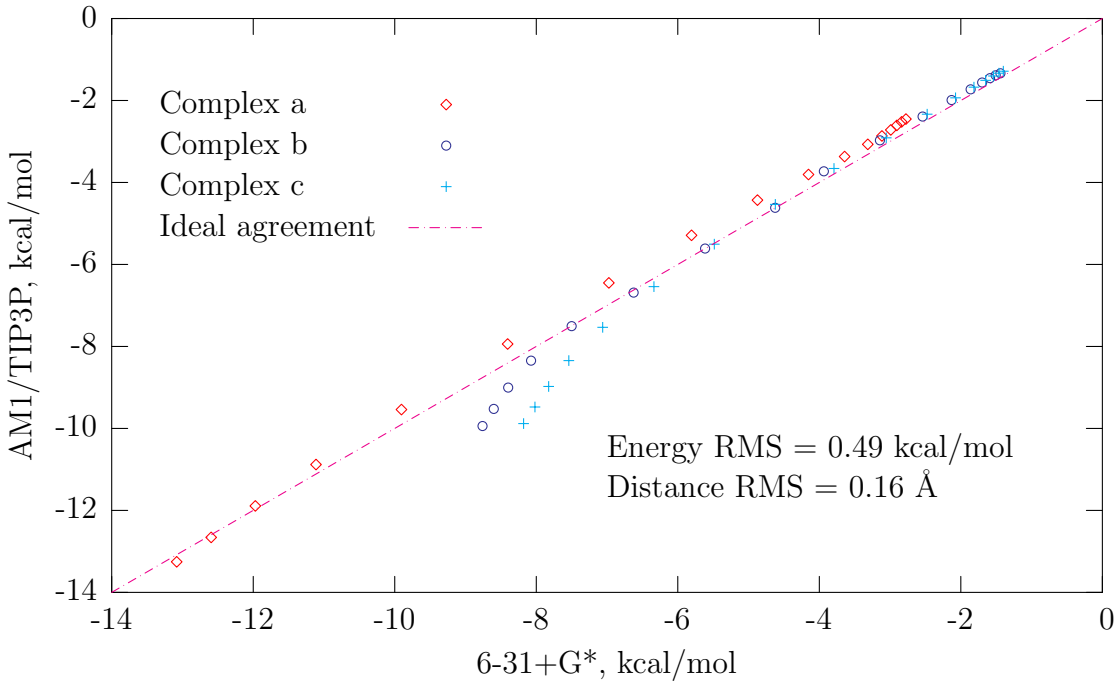

Figure 5.6: Comparison of the binding energy of the Menshutkin reaction region along the reaction coordinate, solvated by a single water molecule. QM/MM results, using AM1 for the QM and TIP3P for the MM region, are compared to structures at HF/6-31+G* level.

implicit handling of average polarization, which differs between solvent models. (If the parameters optimized for TIP3P are used for minimizations with TIP4P and TIP4P-FQ, energy RMS errors of 0.69 and 1.28 kcal/mol respectively are incurred.) Therefore, the interactions must be optimized for each solvent model; this was carried out by minimizing the function

$$\chi = R_E + \zeta R_d \tag{5.8}$$

where R_E is the RMS error in the binding energy between the *ab initio* and QM/MM minimizations, and R_d the RMS error in the distances. ζ is a weighting parameter, taken as 3.0 kcal/mol/Å for this study. The minimization was carried out by a grid search of the N van der Waals parameters, ϵ and σ, using complex 'a', and then a separate search of the Cl parameters, using complexes 'b' and 'c'. As in the original

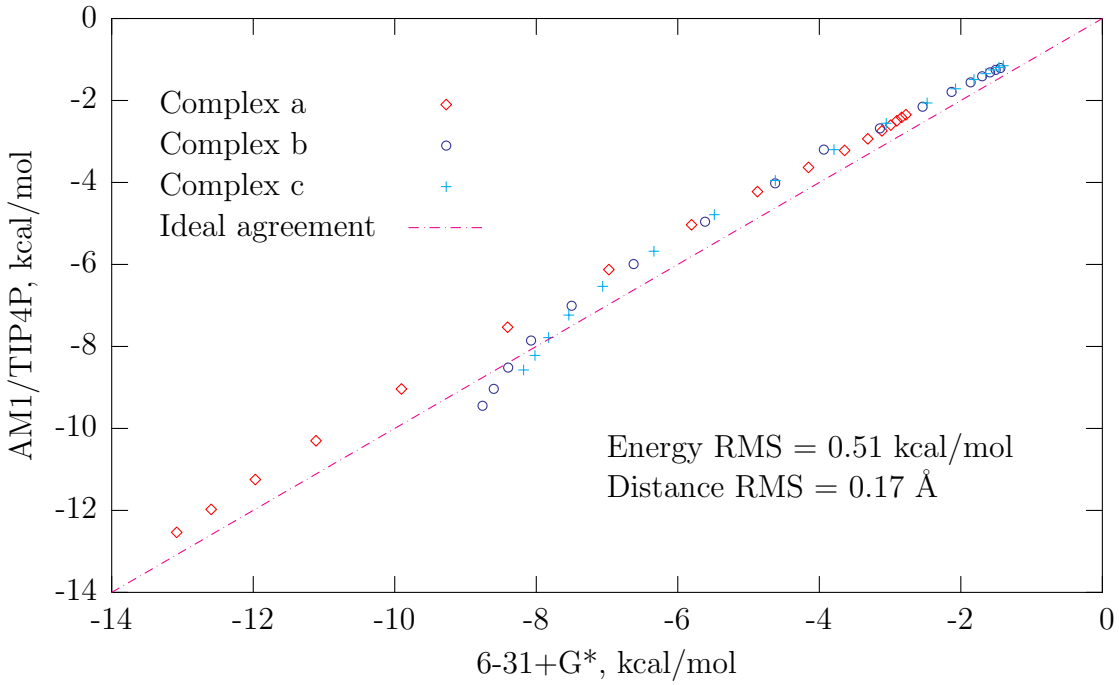

Figure 5.7: Comparison of the binding energy of the Menshutkin reaction region along the reaction coordinate, solvated by a single water molecule. QM/MM results, using AM1 for the QM and TIP4P for the MM region, are compared to structures at HF/6-31+G* level.

fit for the TIP3P interaction, no van der Waals interactions were defined for amine hydrogens; carbon, and methyl hydrogens were taken to have the same parameters as in the TIP3P study.

The resulting fit for the AM1/TIP4P energies is shown in figure 5.7, and that for AM1/TIP4P-FQ in figure 5.8. The determined van der Waals parameters, compared to those for the TIP3P system, are shown in table 5.1. The fit for TIP4P can be seen to be very similar to that for TIP3P; that for TIP4P-FQ is less good, with an RMS error in the energies of 0.79 kcal/mol, but still reasonable. In all cases the fit for complexes 'b' and 'c' shows the greatest deviations from linearity, with AM1 tending to overestimate the binding energy of the highest-energy complexes (those for $R_C > 1.0$ Å) while underestimating that of the low-energy configurations.

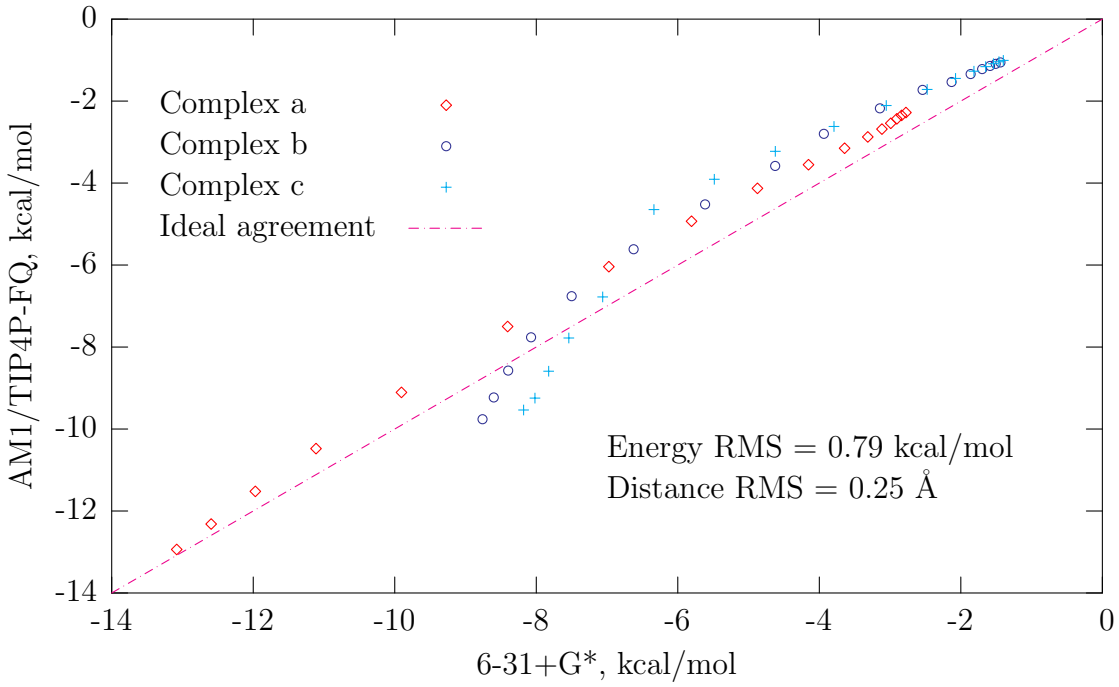

Figure 5.8: Comparison of the binding energy of the Menshutkin reaction region along the reaction coordinate, solvated by a single water molecule. QM/MM results, using AM1 for the QM and TIP4P-FQ for the MM region, are compared to structures at HF/6-31+G* level.

	TIP3P/AM1[a]		TIP4P/AM1		TIP4P-FQ/AM1	
Atom	ϵ	σ	ϵ	σ	ϵ	σ
C	0.1000	3.4000	0.1000	3.4000	0.1000	3.4000
N	0.1615	3.0875	0.2000	2.9043	0.1900	2.5836
Cl	0.1119	4.1964	0.1500	4.0981	0.1500	4.2763
H_C	0.0700	2.0000	0.0700	2.0000	0.0700	2.0000
H_N	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

[a]From Gao *et al.* [258]

Table 5.1: Lennard-Jones parameters (see equation 1.52) determined for the Menshutkin reactive region to optimize the QM/MM interaction between it, treated at AM1 level, and the solvent, treated with TIP3P, TIP4P or TIP4P-FQ models. ϵ values are given in kcal/mol, and σ in angstroms. H_C and H_N are the hydrogen atoms directly bound to the carbon and nitrogen respectively.

5.3.3 Preparation of the system

With the AM1 method chosen for the QM treatment, and the interactions between this region and the first solvation sphere parameterized, the remainder of the solvent was then added and equilibrated. In order to determine properties that are directly comparable with experiment, Gibbs rather than Helmholtz free energies are required. This necessitates simulating the system in the isothermal-isobaric (NPT) ensemble [119, 196, 281], allowing the volume of the system to alter during the simulation. However, such volume fluctuations introduce an extra degree of freedom into the system, which can cause complications during equilibration. Additionally, water is fairly incompressible, so the volume is not likely to change greatly in any event. Thus, the majority of the equilibration phase of the simulation was carried out in the standard NVT ensemble, and only once the system was well equilibrated was the volume allowed to fluctuate in NPT dynamics.

Generation of the solvent box

Water molecules were first generated on a simple cubic lattice; the repeat distance of this lattice was chosen so as to reproduce the experimental density of liquid water, 1 g/cm^3, or 0.0334 molecules/Å^3. The dimensions of the complete generated water box were chosen to be integer multiples of this cubic lattice distance, and periodic boundary conditions were imposed for the duration of the generation process. In the case of the umbrella sampling runs described in this chapter, a cubic box with sides 21.724 Å long, containing 343 water molecules, was constructed to allow adequate solvation of the reaction region.

A regular cubic lattice is far too ordered a structure for liquid water; thus some disorder was introduced by performing molecular dynamics at 50 K for 500 fs, using a 1 fs time step. This served to randomise the orientations of the molecules, and to

break the periodicity of the starting lattice, leaving only the periodicity of the entire solvent box.

A reasonably minimized structure was then obtained by performing Adopted Basis Newton Raphson minimization until the energy gradient dropped to below 0.01 kcal/mol/Å. During the minimization, water molecules adopted the favourable tetrahedral structure and the majority of stabilising hydrogen bonds were formed. However, the likelihood of any minimization of a system containing even a few hundred molecules reaching the global minimum is remote; some parts of the system would likely be stuck in unfavourable configurations. Thus, the minimization was followed by 40 ps of dynamics with a 1 fs time step, to evenly distribute energy throughout the system. Dynamics were initiated at a temperature of 0 K, and energy was gradually introduced by increasing the temperature by 10 K every 500 time steps. The temperature increase was effected by reassigning the velocities of all particles randomly, such that the velocities fitted to a Gaussian distribution characteristic of the new temperature. This served to evenly spread the newly-introduced energy, and to reduce the possibility of any 'hot spots' developing in the system. Once the temperature reached 298 K, it was maintained by the same procedure of reassigning velocities every 500 time steps.

The water box at 298 K was then equilibrated by performing further molecular dynamics for 80 ps with a 1 fs time step. The final velocities from the heating stage were used as starting velocities for this equilibration. No velocity reassignment was carried out; instead, the temperature was maintained at 298 K by use of a Berendsen weak coupling thermostat, with coupling constant 0.005 ps^{-1}.

Solvation of the reactive region

The solution phase system was built by first taking the optimized gas-phase Menshutkin system at R_C = -0.5 Å and overlaying it with the water box. The two systems were aligned such that the centre of the water box, and the QM carbon atom, both resided

at the origin of the coordinate system. It would be expected that the structure of pure water would be very different from that of water around a solute; no waters should be found within the van der Waals surface of the solute (as their interactions with the QM region would be extremely unfavourable), and the waters should not be randomly oriented, but forming a cage structure around the solute. Thus, such a simple overlaying of two coordinate sets would not be expected to give reasonable solvation.

The first of these problems was approached by removing any water molecules that had their oxygen atoms within 2.8 Å of any of the QM heavy atoms. Periodic boundary conditions were then imposed, and any remaining unfavourable interactions were relaxed by performing 100 steps of steepest descents minimization, using the full QM/MM potential, but only permitting the solvent to move.

Simply removing waters that overlap with the QM region cures the problem of high-energy interactions between the QM and MM regions; however, it can also lead to cavities developing in the system. These are void regions where a water molecule should be present, and its absence can cause large inaccuracies in the solute-solvent interaction energy. Equilibration is not generally sufficient to remove these cavities, as they can persist over extremely long timescales. This problem was addressed by considering points on a simple cubic lattice, centred on the origin, spaced by 0.1 Å, and covering the entire system. Any point that was more than 3.0 Å away from all QM heavy atoms and water oxygens was taken to be a potential cavity, and filled by adding a water molecule to the system, such that its oxygen atom was on the relevant grid point. The orientations of these newly-added waters were then optimized by performing a further 100 steps of steepest descents minimization with a fixed QM region.

Equilibration

The structure of the water was allowed to gradually relax around the solute by means of gradual heating with periodic velocity reassignment for 20 ps and equilibration for

50 ps of the solvent, keeping the solute position fixed. The final configuration was then placed into the NPT ensemble, and simulated for a further 30 ps.

For this and all subsequent NPT simulations, the Langevin piston method of Feller *et al.* [282] was used to maintain constant pressure and temperature. This approach is similar to the weak coupling methods [117] described earlier for maintaining the temperature (equation 1.106) with the following equations of motion of the system with instantaneous volume V and pressure P:

$$\dot{\boldsymbol{r}}_i = \frac{\boldsymbol{p}_i}{m_i} + \frac{1}{3}\frac{\dot{V}}{V}\boldsymbol{r}_i \tag{5.9}$$

$$\dot{\boldsymbol{p}}_i = -\frac{\partial U}{\partial \boldsymbol{r}_i} - \frac{1}{3}\frac{\dot{V}}{V}\boldsymbol{p}_i \tag{5.10}$$

$$\ddot{V} = \frac{1}{W}(P - P_{ext}) - \gamma\dot{V} + R \tag{5.11}$$

where W is the mass of the Langevin piston, taken as 500.0 amu in this study, P_{ext} the reference pressure (1.0 atm), and γ the collision frequency of the piston, 20.0 ps^{-1}. R is a random force, given by a Gaussian distribution with variance proportional to a piston bath temperature, 298 K in this simulation. The coupling to this heat bath is rather weak, giving very slow exchange of energy [282], and so this was augmented by a Nosé-Hoover thermostat with mass 1000.0 ps^2 kcal/mol.

5.3.4 Umbrella sampling

With a reasonably well equilibrated solution phase system, umbrella sampling was then carried out to obtain the potential of mean force of the Menshutkin reaction, as outlined in section 1.10.6. For each window, the reaction coordinate, R_C, as defined by equation 5.6, was constrained by means of a harmonic restraint term (equation 5.7). Trial simulations were carried out with different force constant (K) values, and the resulting reaction coordinate distributions compared. A K value of 800 kcal/mol/Å^2

was chosen such that reasonable overlap was achieved between windows separated by 0.5 Å along the reaction coordinate (see also figure 1.3).

The N–C–Cl angle was constrained near 180° with a harmonic force constant of 50 kcal/mol/radian^2. Simulations were carried out in the NPT ensemble at 1.0 atm and 298 K, using the Langevin piston technique described above.

The first window was run with R_{ref} = -0.50 Å. The starting configuration for this window was taken from the final state of the solvent equilibration, and sampling was carried out for 70 ps. The final coordinates and velocities from this simulation were then used as the initial state for two more windows, centred at -0.55 and -0.45 Å, and again simulated for 70 ps. These windows in turn provided starting states for further windows at -0.60 and -0.40 Å respectively, and so on. The reaction coordinate was covered in this way from -1.60 to 2.20 Å, in 0.05 Å increments. Each window took approximately 3 hours of CPU time on a 1667 MHz Athlon XP machine.

During each simulation, the instantaneous value of the reaction coordinate was recorded at each energy calculation, and then gathered into bins of width 0.005 Å to give biased probability distributions of the coordinate for each window, as in figure 1.3. As can be seen from figure 5.9, the cumulative average of this quantity takes some time during a simulation to reach its final value — in particular, for the initial part of the run, the system is far from equilibrium, as the starting configuration corresponds to a different R_{ref} value. Thus, the first 10 ps of each window were ignored, and treated as equilibration. The remaining 60 ps of sampling from each window were then joined by means of a simple technique. The relative free energy profile for the first window, at -1.60 Å, was first obtained by application of equation 1.129, where the form of the biasing function, $w(\zeta)$, was given by equation 5.7, and the final term was ignored. This process was repeated for the next window, at -1.55 Å. The second free energy curve was then offset on the energy axis to join the two sets; this offset was calculated as the average energy difference between the two curves in the part of the

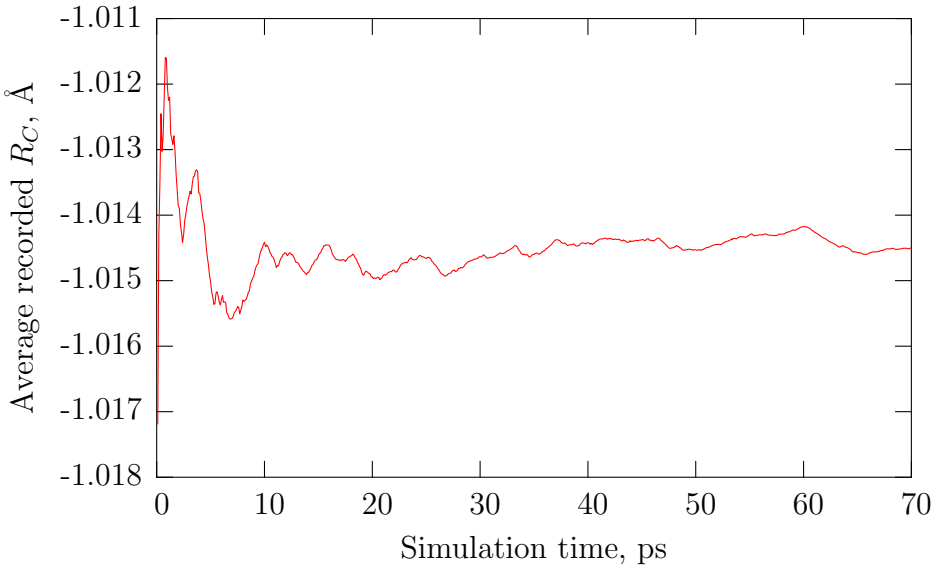

Figure 5.9: Behaviour of the cumulative average of the reaction coordinate during simulation of the solvated Menshutkin system at R_C = -1.0 Å.

reaction coordinate where they overlapped, weighted by the number of samples in both windows. A combined distribution was then generated, where the energy of each point was taken as the average of the two curves, again weighted by the number of samples in each set. (The number of samples of each point in the combined set was simply taken as the sum from both sets.) The third window, at -1.50 Å, was then added to this set to produce a new combined set, and so on, until a complete PMF was generated. This sampling-weighted averaging procedure automatically removes any regions of poor sampling, except in parts of the reaction where the overlap between adjacent windows is particularly poor.

For FlucQ simulations, the initial equilibrated system was modified by adding M sites to each water molecule, to generate TIP4P waters. The FlucQ polarization energy of the system was then reduced to below 10^{-6} K by application of dissipative Langevin dynamics in charge space, with a frictional coefficient of 1600 ps^{-1} and a 1 fs time step. The umbrella sampling simulation was then carried out in exactly the same way as for TIP3P water, except that the optimized van der Waals parameters for the QM/MM interaction with TIP4P-FQ water were used. Charge temperature was maintained at

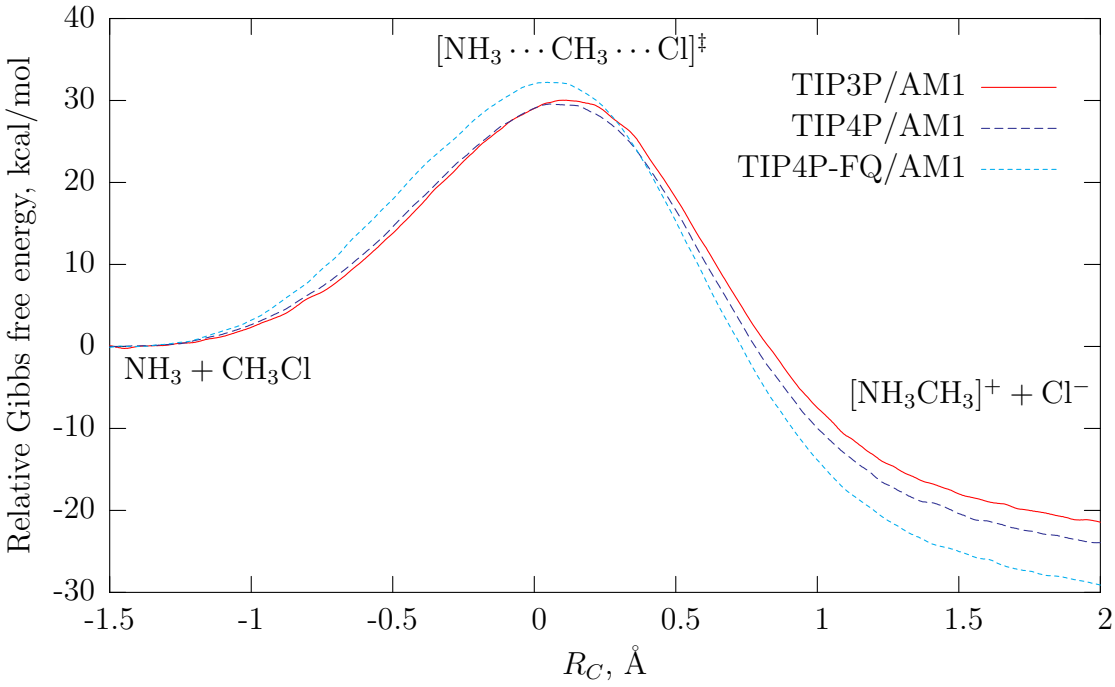

Figure 5.10: Determined solution-phase potential of mean force of the Menshutkin methyl chloride/ammonia reaction, using AM1 for the reactive region and TIP3P, TIP4P or TIP4P-FQ for the solvent.

1.0 K during each window by application of a Nosé-Hoover thermostat with inertia parameter 10.0 ps^2 kcal/mol. Finally, for comparison, the free energy profile was also determined using the TIP4P solvent model and the corresponding optimized van der Waals parameters. The free energy profiles from these three simulations are shown in figure 5.10.

5.3.5 Discussion and analysis

The AM1/TIP3P simulation yields an activation free energy of about 30.0 kcal/mol. This is in reasonable agreement with the study by the previous QM/MC study by Gao and Xia using the same method [258], which predicted 26.3 ± 0.3 kcal/mol. However, it does overestimate the experimentally-determined activation energy of the methyl iodide/ammonia reaction, which is 23.5 kcal/mol [268, 269]. (Unfortunately, experimental data for the analogous methyl chloride reaction are not available.) The

calculated free energy of reaction is about -21 kcal/mol, again agreeing well with Gao's value of -18 kcal/mol, although this figure is very approximate as the free energy curve is not completely flat at $R_C = 2.0$ Å. This is consistent with the behaviour seen in the earlier PCM simulation, although it is more pronounced, probably because the periodic box is insufficiently large to prevent interactions between the solvation spheres of the primary and 'image' QM atoms. (The primary and image atoms themselves obviously do not interact directly, due to the cutoff imposed on the simulation; at any rate, the 'image' QM atoms appear to the MM force field as uncharged particles.) The predicted free energy of reaction is also in very poor agreement with that estimated from a thermodynamic cycle using experimental free energies, -34 ± 10 kcal/mol [277], although this is caused by the AM1 method itself, which predicts a heat of formation 18.2 kcal/mol higher than the experimental value for Cl^- [56].

The discrepancy between this simulation and Gao's could be caused by a number of factors. The earlier simulation covered the reaction in a two-dimensional fashion, varying the C–N and C–Cl bonds independently, and obtained the PMF as the least-energy path over the resulting 2D surface, rather than using a one-dimensional reaction coordinate. The result could also be affected by the size of the simulation box and the cutoff method used; a 9.5 Å cutoff is perhaps too short to adequately treat the solvation of charged species [85]. However, test simulations along part of the reaction coordinate, using a switched cutoff rather than a shift, agreed to within 1 kcal/mol with the simulation presented here, as did simulations using a much larger 31.04 Å box of 1000 water molecules and a longer cutoff (with the exception of the reaction near the product state, where the PMF was predicted to drop much less steeply). It is also possible that differences could be caused by the use of MC by Gao rather than MD, and this is explored in Chapter 6.

The TIP4P simulation is in excellent agreement with TIP3P, although the free energy of reaction is overestimated by about 2.5 kcal/mol. The discrepancy is explained

by the corresponding deviations from the HF/6-31+G* binding energies in figures 5.6 and 5.7, which are largest for the TIP3P complexes b and c. It should also be remembered that only complexes up to $R_C = 1.6$ Å were considered in the fit for the QM/MM interactions in this study, while Gao used complexes only up to 0.75 Å, so the interactions were not optimized at all for the reaction at 2.0 Å.

The TIP4P-FQ model gives higher activation and reaction free energies, of 32.2 and -29 kcal/mol respectively, and a slightly earlier transition state. This is likely due in part to the poorer agreement between HF/6-31+G* and AM1/TIP4P-FQ binding energies than with TIP3P in section 5.3.2. (For comparison, optimized van der Waals parameters for AM1/TIP4P-FQ1 give energy and distance RMS errors of 0.62 kcal/mol and 0.17 Å respectively.) While TIP3P gives very good binding energies around the transition state, TIP4P-FQ consistently predicts the complexes to be one or two kcal/mol less strongly bound. This is probably a stronger effect than that of the bulk solvent which, as commented on in section 4.2.1, has a higher molecular dipole moment on average (2.5–2.6 D) for TIP4P-FQ than for TIP3P (2.35 D), and which might thus be expected to more stabilise the TIP4P-FQ transition state, as the free energy of solvation becomes steadily more negative as the reaction proceeds [277]. (At any rate, the TIP4P model, with a *lower* dipole than TIP3P at 2.18 D, does not show any destabilisation.)

The poor QM/MM energies for TIP4P-FQ are surprising, given its utility for *ab initio* QM/MM demonstrated in section 4.4. However, it should be recalled that the AM1/MM method suffers from a number of deficiencies — for example, it predicts an incorrect C_{2v} structure for the water/chloride complex [17] — while AM1 itself generally gives poorer hydrogen-bond geometries than PM3 [283], so it is entirely possible that AM1 does not well describe the QM electron density. Additionally, the form of the QM/MM interaction in semi-empirical methods, equation 1.61, includes parameters that must be optimized for the MM atoms, and these have previously

been determined for the AM1/MM method for use with non-polarizable water models such as TIP3P [17], rather than for TIP4P-FQ. Finally, note that the QM region is fixed at the geometry optimized at HF/6-31+G* level during the determination of the QM/MM parameters in section 5.3.2. This fails to take into account the difference between the AM1 and *ab initio* geometries, the relaxation of the QM region in response to the approach of a water molecule, or the difference in QM polarization between singly-solvated and multiply-solvated solute, factors which are likely to be important when dealing with polarizable solvent. (The TIP4P-FQ1 model has already shown in section 4.6 that polarization suitable for bimolecular systems is not necessarily transferable to the bulk.)

In order to further analyse the solvent structure, molecular dynamics simulations were run in the NVT ensemble[1] at 298 K. Three runs were carried out, constraining the reaction coordinate to -1.50, 0.10 and 2.00 Å, to study the reactants, transition state, and products, respectively. Each simulation was run for 100 ps, using the coordinates and velocities from the umbrella sampling window at the same R_{ref} value for the initial configuration. Trajectories from these simulations were used to obtain the Cl–O radial distribution function. This function is shown for TIP3P in figure 5.11 and for TIP4P-FQ in figure 5.12.

For both solvent models, it can be seen that the chlorine-oxygen interaction strengthens during the course of the reaction, as charge on the chlorine develops, from the movement of the first peak of each radial distribution function from around 4.0 Å for the reactants down to about 3.2 kcal/mol for the products. The TIP4P-FQ model yields very similar structure to TIP3P, although there are small differences in peak heights; this is probably caused by the differing solute-water distances predicted by the two models in section 5.3.2.

[1]NPT was not used, as the CHARMM solvent analysis module requires a constant periodic box size. At any rate, the box length fluctuated by less than 0.2 Å during NPT simulations, so NVT is a reasonable approximation.

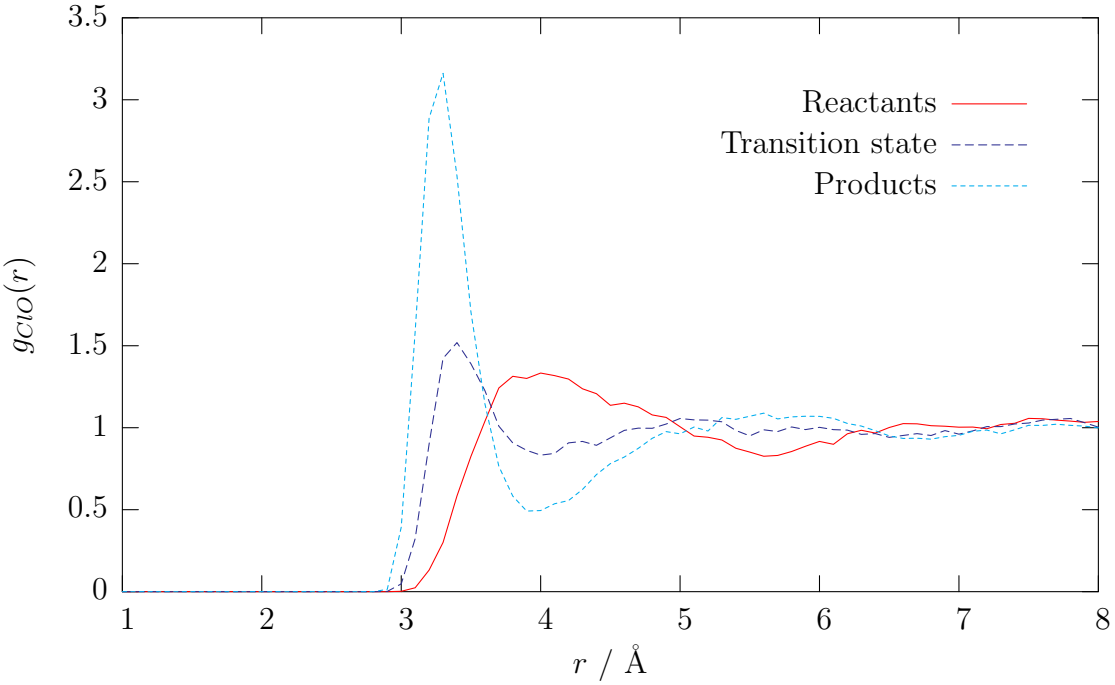

Figure 5.11: Quantum chlorine-classical oxygen radial distribution function, evaluated for AM1/TIP3P simulations at $R_C = -1.50$ (reactants), 0.10 (transition state) and 2.00 Å (products).

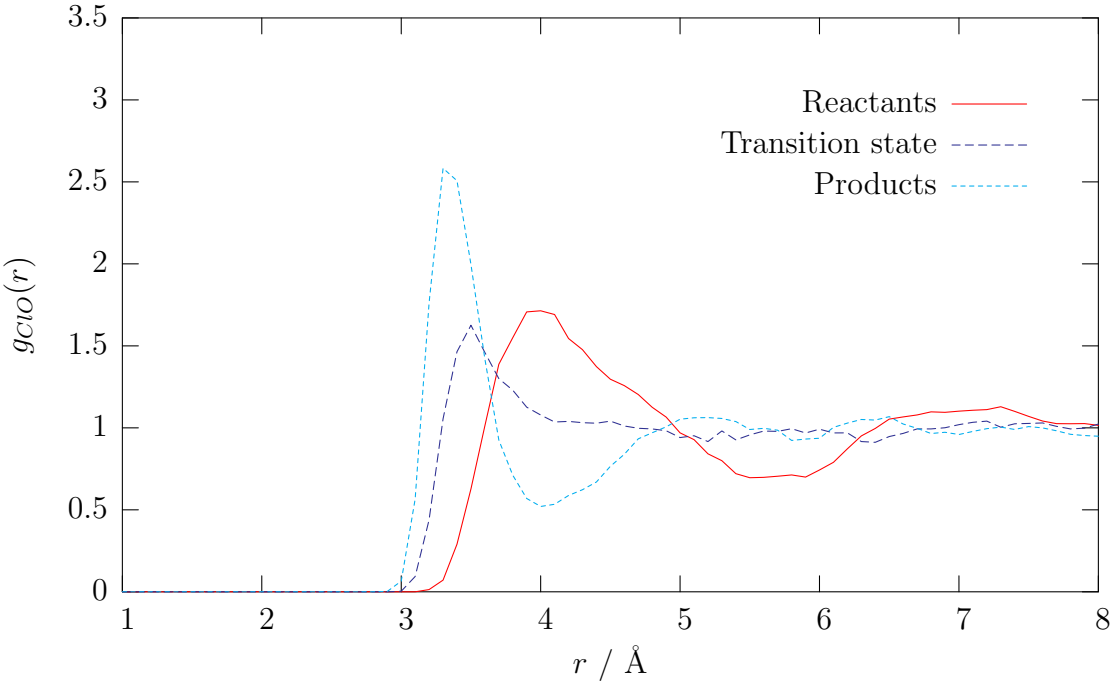

Figure 5.12: Quantum chlorine-classical oxygen radial distribution function, evaluated for AM1/TIP4P-FQ simulations at $R_C = -1.50$ (reactants), 0.10 (transition state) and 2.00 Å (products).

5.4 Summary

The Menshutkin S_N2 reaction has dramatically different energetics in the gas and solution phases; the addition of solvent, whether it be implicit PCM solvent or explicit water molecules, strongly stabilises the charge-separated transition state and products, and converts a strongly endothermic gas reaction into a rather exothermic solution phase process. The umbrella sampling method at constant pressure, using a QM/MM energy treatment, generates Gibbs free energies suitable for comparison with experiment. However, it is vital to determine optimum van der Waals parameters for the QM/MM interaction, as rather different solute-solvent binding energies are predicted for each solvent model. TIP3P, TIP4P and TIP4P-FQ all give results in reasonable agreement with previous QM/MM studies, although the experimental free energy is consistently overestimated, suggesting that the AM1 method itself is at fault. The TIP4P-FQ method in particular gives no real improvement; an early transition state is predicted, but it is also less stable than with TIP3P.

The free energy simulations in this chapter use molecular dynamics in combination with QM/MM. The Menshutkin reactive region, however, being small and highly symmetrical, is also amenable to treatment with Monte Carlo methods, which is the subject of the next chapter.

Chapter 6

Use of hybrid QM/MC methods

6.1 Overview

It has already been demonstrated that molecular dynamics methods can be used in conjunction with a QM/MM representation of the system energy to elucidate average energetic and structural properties, such as the molecular dipole and radial distribution functions, as well as the free energy, vital for comparison with experimental data. However, it is also possible to use Monte Carlo simulation methods in a similar way, using the same hybrid QM/MM energy function. (This is distinct from so-called quantum Monte Carlo [284, 285] methods, which use the Monte Carlo algorithm to optimize a quantum mechanical wavefunction.)

Such hybrid QM/MC methods have been applied to a variety of systems [78, 286]. Monte Carlo methods have several advantages over a corresponding QM/MD simulation:

1. Constraints and symmetry can easily be conserved by suitable choice of a move set; by default molecular dynamics explores all degrees of freedom, and a variety of constraints must be applied if the simulation of highly symmetric systems is desired. This also enables the shape of solvent boxes to be more closely fitted to the solute. For example, a long linear molecule could be simulated within a

cuboidal box, with the molecule aligned with the longest box side; this would only be possible in an MD simulation if suitable restraints were placed on the solute.

2. Forces (energy gradients) are not required, thus saving significant time in a QM energy calculation. This can also reduce code complexity and any errors introduced for new methods such as DFT, where forces may be approximated numerically.

3. Temperature control is straightforward, simply by adjusting the acceptance criterion. Different temperatures can be enforced for different degrees of freedom by using different acceptance criteria.

4. The problem of the QM region drifting away from the centre of the primary box in periodic boundary simulations (see section 5.3.1) is simply solved by a suitable choice of moves.

Unfortunately, there are also a number of disadvantages:

1. As has already been discussed in section 4.2.4, when using FlucQ in combination with Monte Carlo, many electronic moves must be sampled for every nuclear move. This makes fully polarizable simulations, at least using the sequential FQ/MC algorithm, very much more expensive than non-polarizable analogues. Possible solutions to this problem include use of the ANES-MC algorithm, as also mentioned in section 4.2.4.

2. The move set must be chosen with care to sample all relevant degrees of freedom, with a reasonable acceptance rate, and with the minimum number of moves. A number of methods exist to tackle this problem, such as dynamically optimized Monte Carlo [287] (DOMC), which adjusts the size of moves during a simulation in response to the acceptance rate of these moves, and preferential

sampling (described further in section 6.2), which concentrates sampling in more energetically important regions of phase space.

3. Every move, no matter how minor, affects the system Hamiltonian and therefore requires a new SCF calculation. One method for ameliorating this problem is the 'Perturbative QM/MC' method described in section 6.3.

In this chapter, the preferential sampling and perturbative QM/MC methods are developed and tested, and are applied to determine the free energy profile of the Menshutkin reaction, for comparison with the MD methods described in Chapter 5. Approximate methods for increasing the accuracy of the QM representation are also proposed.

6.2 Preferential sampling techniques

The Monte Carlo method was originally developed for the study of bulk liquids [121]. In such a study, a truly random selection of moves, as suggested in section 1.10.4, is perfectly adequate, as the system is homogeneous. However, in a simulation of a solution, properties are largely determined by solute-solvent interactions and solvent-solvent interactions in the first solvation sphere. Thus, the efficiency of a Monte Carlo simulation could be improved by concentrating the sampling around the solute [288, 289]. This is usually referred to as 'preferential sampling'.

The simple preferential sampling method [288] divides the system into two distinct regions — an 'in' region, where sampling is to be concentrated, and an 'out' region comprising the rest of the system. The 'in' region comprises all particles within a certain user-specified distance R of the solute. Then, when a molecule is chosen to be moved, if it is in the 'in' region it is immediately considered, while if in the 'out' region, it is considered with user-specified probability ρ. Thus, the distribution of sampling between the two regions can be controlled.

Note that by choosing states with different probabilities in the two regions, the underlying stochastic matrix α (see section 1.10.4) is no longer symmetric. For example, since 'in' molecules are selected more frequently, $in \rightarrow out$ moves occur more frequently than $out \rightarrow in$ moves, and thus the 'in' region will be gradually depleted. Therefore, the acceptance criterion must be altered in order to ensure correct sampling.

A more refined technique [289] does not arbitrarily partition the system; instead, a weighting function $w(r_i)$ is used to bias the choice of molecules to perturb, where r_i is the distance between the centre of mass of solvent molecule i and the solute. A common form for the weighting function, used in several studies [258, 290], is:

$$w(r_i) = \frac{1}{r_i^2 + C} \tag{6.1}$$

where C is an adjustable parameter.

After choosing a molecule i to perturb in the usual way, a random number ζ is chosen between 0 and 1. The molecule is further considered only if

$$\zeta < \frac{w(r_i)}{w_{\max}} \tag{6.2}$$

where $w_{\max}$ is the value of the weighting function for the solvent molecule closest to the solute. This results in the molecules nearest the solute being sampled with near 100% of normal probability, and the sampling probability dropping smoothly as the distance from the solute is increased. In this work, however, $w_{\max}$ is approximated by the value of $w(r_i)$ when $r_i = 0$, which is an easily-evaluated constant, and which removes the need to keep track of which solvent molecule is closest to the solute. This does of course reduce the efficiency of this part of the algorithm, since the probability of failing the test in equation 6.2 is increased; however, choosing a new molecule is a computationally inexpensive process.

If the chosen molecule passes the test, it is perturbed as per the normal Monte Carlo technique. However, once again the acceptance criterion must be modified to account for the asymmetric underlying matrix α. The original Metropolis acceptance criterion, equation 1.121, can be written concisely as:

$$P(a \rightarrow b) = \min\left[1\,,\ \frac{\pi_b}{\pi_a}\right] \tag{6.3}$$

where π_a is the Boltzmann factor for state a,

$$\pi_a = \exp\left(\frac{-V(\boldsymbol{r}_a)}{k_B T}\right) \tag{6.4}$$

With preferential sampling in effect, this acceptance criterion must be modified to:

$$P(a \rightarrow b) = \min\left[1\,,\ \frac{w(r_b)\pi_b}{w(r_a)\pi_a}\right] \tag{6.5}$$

Clearly, this also means that moves with $\Delta V(a \rightarrow b) \leq 0$, which are always accepted in conventional unbiased Monte Carlo, can be rejected by the preferential sampling algorithm.

6.3 Perturbative QM/MC

Monte Carlo simulations typically require many more moves to simulate a system than the number of time steps required for the same simulation using molecular dynamics. In conventional Monte Carlo, however, this is balanced out, in that each of these moves only requires the evaluation of the energy terms in which the moving atoms are involved. This balance is lost in Monte Carlo simulations with a hybrid QM/MM treatment, because the total Hamiltonian for the system (equation 1.55) includes both the QM and MM total Hamiltonians. Thus, *any* move, of either the QM or MM regions, modifies the total Hamiltonian, necessitating a full QM SCF calculation. This

increases the time taken to perform a QM/MC simulation drastically in comparison to the equivalent QM/MD treatment.

One method for solving this problem is to realise that many moves, for example those of uncharged solvent molecules far away from the QM solute, have very little influence on the electron density in the QM region. Thus, to a reasonable approximation, the electron density can be assumed to be unchanged for such a move, and the SCF calculation can be avoided. This method is known as 'frozen density' [234] or 'perturbative QM/MC' [291, 292].

Following from equations 1.55 and 1.56, moving an MM particle i causes a change in the total Hamiltonian:

$$\Delta\hat{H} = \Delta\hat{H}_{elec} + \Delta\hat{H}_{vdW}^{QM/MM} + \Delta\hat{H}_{MM} \tag{6.6}$$

where

$$\Delta\hat{H}_{elec} = -\sum_{e}\left(\frac{Q_i'}{r_{e,i}'} - \frac{Q_i}{r_{e,i}}\right) + \sum_{m} Z_m\left(\frac{Q_i'}{r_{m,i}'} - \frac{Q_i}{r_{m,i}}\right) \tag{6.7}$$

In conventional QM/MM, this change of the Hamiltonian necessitates a new quantum calculation. However, if the change is small enough, the resultant energy change can be approximated by first order perturbation theory as

$$\Delta E_{pert} = \left\langle \Psi \middle| \Delta\hat{H} \middle| \Psi \right\rangle \tag{6.8}$$

By comparison with equations 1.57 and 1.58, this yields:

$$\Delta E_{pert} = \Delta E_{pert,elec} + \Delta E_{vdW}^{QM/MM} + \Delta E_{MM} \tag{6.9}$$

where

$$\Delta E_{pert,elec} = \sum_{\mu}\sum_{\nu} P_{\mu\nu}(I_{\mu\nu}^{i'} - I_{\mu\nu}^{i}) + \sum_{m} Z_m\left(\frac{Q_i'}{r_{m,i}'} - \frac{Q_i}{r_{m,i}}\right) \tag{6.10}$$

Observe that the change in the van der Waals contribution to the QM/MM interaction, $\Delta E_{vdW}^{QM/MM}$, and the change in the energy of the MM system, ΔE_{MM}, are both independent of the QM electron density, and are thus identical to the non-perturbative energies.

The only time-consuming part of this calculation is the evaluation of the one-electron integrals, $I_{\mu\nu}^{i}$ in equation 6.10. However, they have only to be evaluated for the moving atoms, in their old and new positions; an SCF calculation, on the other hand, requires evaluation of the 1-electron integrals for all MM atoms.

6.3.1 Quantification of error

The energy change calculated by equation 6.9 is an approximation to the true value, as first order perturbation theory assumes that the QM density matrix is not changed by the perturbation. In order for this method to be of use for Monte Carlo simulation, it must be possible to quantify or at least control the error introduced. In the original work by Truong *et al.* [291] a full SCF was carried out every NMOVE steps, and the error was controlled simply by adjusting the value of NMOVE. However, this approach assumes that the error introduced by an MM move is roughly constant. In reality, it might be expected that movement of an MM particle directly bound to the QM system would have a larger effect on the QM density matrix (and thus yield a larger error) than that of a distant particle. This was accounted for by a more rigorous treatment in a later paper [292], in which classical electrostatic theory was used to approximate the error E_{err} introduced by a move of an MM particle at distance r from the QM region as:

$$E_{err} \approx \frac{K_{pert}}{r^7} \tag{6.11}$$

where the proportionality constant K_{pert} was determined by a least-squares fit to exact values of the error from a preliminary test simulation. The error was then controlled by

accumulating the predicted error E_{err} at each step, and performing a new SCF when it exceeded a certain threshold.

Although this new method does allow for the quantification of error, it also requires an initial calibration to determine K_{pert}. Secondly, the value of K_{pert} is obviously dependent on the QM region configuration. After any move that changes the QM region, the predicted errors will be invalid. This was not an issue with the original work, which only used a single ion in the QM region, but is obviously important with this work, where the QM region contains several atoms, and changes drastically throughout the S_N2 reaction.

6.3.2 Error control in the case of a changing QM region

The approach developed in this work, and implemented in CHARMM, is simpler than the second method described above, avoiding the difficulty of a calibration run, while being more sensitive to the approximate error than the original 'NMOVE' technique.

During an MC simulation, if a move of the MM region is picked, then the change in energy is estimated from equation 6.9 — i.e. the electron density of the initial state is assumed to be applicable to the trial state as well. The error introduced by the perturbation approximation is then assumed to scale with the magnitude of $\Delta E_{pert,elec}$, equation 6.10, i.e. the change in the QM/MM interaction electrostatic energy. The error is controlled by testing this energy change against a per-step energy threshold, ΔE_{step}^{max}. If the energy change exceeds this threshold, then the approximated QM/MM electrostatic energy delta is discarded; a full QM SCF calculation is carried out instead to determine the delta. (Note, however, that the delta will still be an approximation, as the energy of the initial state may have been determined by previous perturbations.)

If multiple MC steps all result in energy deltas that lie below the threshold, then the original system could be perturbed far from its starting configuration, such that perturbations become less and less accurate. This is prevented by two mechanisms;

firstly, an upper cap on the number of moves between full SCF calculations, analogous to Truong's 'NMOVE' approach, is enforced. Secondly, the energy deltas are accumulated since the last full SCF calculation. At each step, the magnitude of this accumulated energy change is tested against a cumulative energy threshold, ΔE^{max}_{cumul}.

If a move in the QM region is chosen, then the energy change cannot be reliably determined by perturbation, as the electron density is certain to change. Thus, for each QM region move, a full SCF calculation is forced. (The above comment about the initial state being an approximation still holds in this case.)

Where a full SCF calculation is required for the trial state, the electron density matrix of the initial state is first saved. If the move is accepted, then the new density matrix, determined by the SCF on the trial state, is kept; if it is rejected, then the saved density matrix is restored, for use in equation 6.10.

6.4 Simulation details

All of the Monte Carlo simulations presented here were performed by the Monte Carlo module [293] in CHARMM [10]. This code was originally written for the simulation of protein folding [293] and as such is tailored more towards large, very flexible systems with low symmetry. In comparison, the system of interest in this work is small and highly symmetrical. Thus, CHARMM was modified by the author both to perform preferential sampling and perturbative QM/MC, and to add several moves that did not break the symmetry of the system, which are detailed below. The system used for study of the Menshutkin reaction was the same as that used in Chapter 5, i.e. the reactive region solvated by 336 water molecules, in a 21.724 Å periodic cubic box.

6.4.1 Move set

A standard set of translations and rigid rotations were used to move the solvent molecules. The quantum mechanics region was held in the centre of the primary simulation box simply by not defining any moves that altered the position of the methyl carbon atom from the origin. Similarly, a 180° N–C–Cl angle was maintained at all times by defining the N–C–Cl bond to be collinear with the z axis, and by not defining moves that altered this state of affairs. The moves used are depicted in figure 6.1. Moves (1) and (2) sampled the reaction coordinate by displacing the amine group and chlorine atom respectively along the z axis. Variations of the amine-methyl dihedral angle were allowed for by means of the single torsional move (3) which rotated the methyl group about the N–C–Cl bond.

The individual bonds and angles involving the hydrogen atoms were not varied independently, so as to maintain the threefold symmetry of the amine and methyl groups. However, flexibility of the angles was allowed for by defining two 'symmetric bend' moves, (4) and (5) in figure 6.1, which altered all three H–N–C angles or all three H–C–N angles respectively by equal amounts. This allowed for inversion of the methyl group along the reaction coordinate, as is normal in an S_N2 reaction of this type. Finally, 'symmetric stretch' moves (6) and (7) altered all three H–N bonds or all three H–C bonds respectively by equal amounts.

In order to sample the various degrees of freedom efficiently and evenly, varying weights were assigned to each move type. An acceptance rate of roughly 50% for each move was maintained by choice of maximum move sizes; these move sizes were determined by running several trial simulations, and adjusting their values after each run. The determined parameters are shown in table 6.1. It can be seen that the vast majority of moves affect only the solvent region, due to the large weights for the two H_2O moves. This is to be expected, as there are many more solvent than solute molecules in the system, and in fact the weights are such that, on average, every

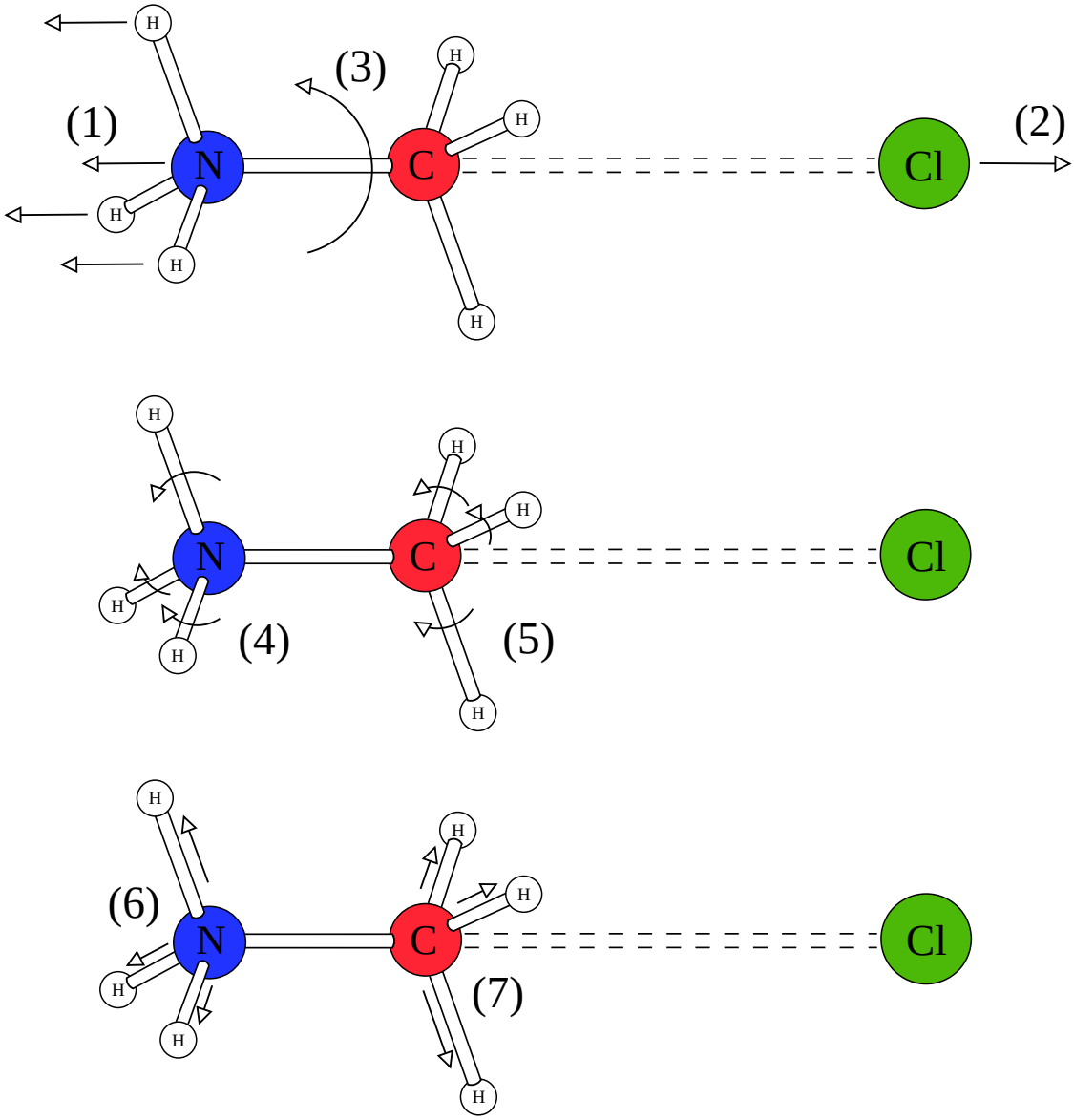

Figure 6.1: The move set used in Monte Carlo simulations of the Menshutkin reactive region, which comprised translations of the amine group (1) and chlorine atom (2), rotations of the methyl group about the N–C–Cl bond (3), symmetric bends (4) and (5), and symmetric stretches (6) and (7).

Move type	Max. move size	Weight
H_2O rigid translation	0.35 Å	800.0
H_2O rigid rotation	30.0°	800.0
NH_3 z axis translation	0.05 Å	1.0
Cl z axis translation	0.05 Å	1.0
CH_3 rotation about z axis	50.0°	0.10
CH_3 symmetric bend	6.0°	0.10
NH_3 symmetric bend	6.0°	0.10
CH_3 symmetric stretch	0.05 Å	0.05
NH_3 symmetric stretch	0.05 Å	0.05

Table 6.1: The weights and maximum move sizes of the move set used in Monte Carlo simulations of the Menshutkin reactive region.

solvent molecule is translated and rotated at least once for each solute move. The two z axis translation moves were given the highest weights of any of the solute moves, as they obviously have a direct influence on the sampling of the reaction coordinate. The remaining QM region moves were given much lower weights, due to their indirect influence on the sampling.

6.5 Perturbative QM/MC test cases

The applicability of the perturbative QM/MC method, as described in section 6.3, was tested by sampling the same system with different maximum per-step energy change ΔE_{step}^{max} and maximum cumulative change ΔE_{cumul}^{max} parameters, and comparing these results to the same system sampled with standard, non-perturbative, QM/MC. The system used was the 9-atom Menshutkin reactive region, solvated by 336 TIP4P waters in a 21.724 Å periodic cubic box. Van der Waals energies, and electrostatic forces, were shifted to zero at 9.5 Å. The reaction coordinate of the Menshutkin reactive region, equation 5.6, was kept at around 2.0 Å by addition to the system energy of a harmonic restraint term, equation 5.7, with force constant 800 kcal/mol/Å^2. Starting

coordinates for the simulation were taken from the final configuration of a previous QM/MC umbrella sampling run at $R_C = 2.0$ Å.

The move set of table 6.1 was used to explore the degrees of freedom of the system at 298 K. The system was sampled for 2×10^6 steps, with the Monte Carlo energies checked every 2000 steps by recalculating the energy of the whole system.

The results are shown in table 6.2, as a function of ΔE_{step}^{max} and ΔE_{cumul}^{max}. A zero per-energy threshold equates to standard non-perturbative QM/MC; i.e. a full QM SCF calculation, including interactions with every MM point charge, is carried out for every MC move. Thus, at least 2×10^6 QM SCF calculations would be required during the MC run; in fact, by reference to table 6.2, it can be seen that actually slightly more than this number of SCF calculations were required for this system. (This excess is caused by additional SCF calculations needed when the energy of the whole system is computed, for example.)

The mean and standard deviation of the absolute energy error are also shown. This error is simply the magnitude of the difference in the total energy, as estimated by adding the energy delta of each accepted MC move to the original system energy, with the energy obtained from a full system energy calculation. It should be seen that this is not zero even when a full SCF calculation is performed at every step, simply due to machine rounding error, minor differences in the convergence of the QM wavefunction, and CHARMM nonbond list[1] updates.

When both of the energy thresholds are raised to 0.1 kcal/mol, the total number of full SCF QM calculations drops dramatically, as the energy changes of many MC steps can be approximated by perturbation. Most of these SCF calculations can be seen to be forced by the perturbative energy delta exceeding the per-step threshold, ΔE_{step}^{max}. A much smaller number are forced by a succession of small energy moves

[1]To avoid the necessity of evaluating every interatomic distance to test against cutoffs, a *nonbond* or *neighbour list* [107, 294] is kept, of all atom pairs separated by a little more than the maximum cutoff; any atom pairs not in this list need not be considered, so increasing the efficiency. This list must be rebuilt periodically, usually by means of applying a heuristic.

ΔE_{step}^{max}, kcal/mol	ΔE_{cumul}^{max}, kcal/mol	Energy deviation, kcal/mol	Relative runtime	Full SCF QM energies		
				Total	Step	Cumul.
0	0	0.00118 ± 0.00236	100.00%	2004001	0	0
0.1	0.1	0.00155 ± 0.00765	18.88%	337949	319826	14116
0.1	1.0	0.00149 ± 0.00648	16.22%	325899	321829	0
0.5	0.5	0.0112 ± 0.0329	7.96%	110313	88958	17359
1.0	1.0	0.0418 ± 0.105	4.60%	54347	37564	12685
1.0	10.0	0.0591 ± 0.102	4.03%	42621	38646	0
2.0	2.0	0.185 ± 0.289	3.56%	22030	10774	7208
5.0	5.0	0.573 ± 0.71	2.66%	6702	878	1819
10.0	10.0	0.869 ± 1.05	2.51%	4310	59	285

Table 6.2: Perturbative QM/MC test cases, with varying values of the maximum allowed per-step and cumulative energy deltas, showing the variations in the number of SCF calculations required, calculated energies, and computational cost.

giving a cumulative energy change that exceeds its threshold. The remainder of the SCF calculations are forced by the energy check calculations every 2000 steps, or by moves in the QM region itself, of which every run had approximately 3000, determined by the weights of the QM region moves. It can be seen also that this simulation took roughly 19% of the CPU time of the non-perturbative QM/MC. In comparison, the number of full SCF calculations was reduced to approximately 17%. Thus, it can be seen that the SCF calculation accounts for a large part of the runtime; however, the time taken for evaluation of the 1-electron integrals, and of the MM system energy, is not negligible. Finally, the energy deviation is still exceedingly small; thus, excellent savings in runtime are accomplished with very little reduction in accuracy.

It can clearly be seen that increasing the energy thresholds ΔE_{step}^{max} and ΔE_{cumul}^{max} further leads to a continued decrease in the runtime and number of required full SCF calculations, but at the expense of increasing the average energy deviation.

6.6 Preferential sampling test cases

The interaction between perturbative QM/MC and preferential sampling was also tested, by running a similar set of simulations where the parameter C in the preferential sampling weighting function, equation 6.1, was varied. The distance r_i in the same equation was taken to be the distance between the centre of mass of the moving molecule and the origin of the coordinate system, which corresponded to the position of the carbon atom in the Menshutkin reactive region. Each simulation was again run for 2×10^6 steps, checking the total energy every 2000 steps, and using perturbative QM/MC with per-step and cumulative energy thresholds of 1.0 kcal/mol. The results of these simulations are shown in table 6.3. The same simulation with preferential sampling disabled, i.e. uniform sampling, is also shown.

It can clearly be seen that the use of preferential sampling increases the number of full SCF calculations, and thus the runtime. These extra SCF calculations are forced by both the per-step and cumulative energy thresholds being exceeded; this in turn is doubtless due to the preferential sampling algorithm favouring movements of the MM system close to the QM solute. As the energy change of an average move is increased, so too is the average error in the energy. As the C parameter is decreased, this energy deviation increases; this is caused by the preferential sampling algorithm becoming more discriminating as C is lessened. (As C approaches zero, molecules are sampled as r^{-2}, while as C approaches infinity, the sampling becomes uniform.)

The effect of preferential sampling and perturbative QM/MC can be seen to be rather similar. The former improves sampling by concentrating it near the solute, i.e. towards the QM/MM boundary where the energetic impact of moves is greatest, while the latter ensures that the 'poor' moves at long distance cost less computer time than those close to the solute. Thus, preferential sampling is not required when perturbative QM/MC is in effect, provided that the cost of making the moves far from the solute is low; the effect is duplicated simply by making more moves, with the additional

Preferential sampling parameter C, Å^2	Energy deviation, kcal/mol	Relative runtime	Full SCF QM energies Total	Step	Cumul.
50	0.0541 ± 0.102	157.43%	100739	75281	21558
150	0.0502 ± 0.098	136.63%	75449	53965	17447
250	0.0476 ± 0.0995	104.95%	68848	48830	16055
(no pref. sampling)	0.0418 ± 0.105	100.00%	54347	37564	12685

Table 6.3: Preferential sampling test cases, with varying values of the weighting parameter C (equation 6.1), showing the variations in the number of SCF calculations required, calculated energies, and computational cost, compared to the same results for uniform sampling.

advantage that moves at long distance, which may still be important where highly charged ionic species are present, are not under sampled.

6.7 A closer examination of the error added by perturbative QM/MC

It was seen in section 6.5 that the error introduced in the perturbative QM/MC method can be controlled by imposition of energy thresholds. The error was determined by comparison with the true QM/MC energy, calculated at intervals during the simulation. For a closer examination of the per-step errors introduced by the approximation, the same system considered in section 6.5 was simulated for 10^6 steps of Monte Carlo. Only the two water moves were used (i.e. the QM region was kept fixed). Preferential sampling, with a weighting parameter C of 100 Å^2, was used to concentrate the sampling towards the origin of the periodic box. At each step, the perturbative approximation for the change in QM electrostatic energy, $E_{pert,elec}$, equation 6.10, was calculated, and then a full SCF was carried out to determine the true energy delta, and thus the error introduced by the perturbative approximation. All moves were rejected, such that the initial state for each move was the same.

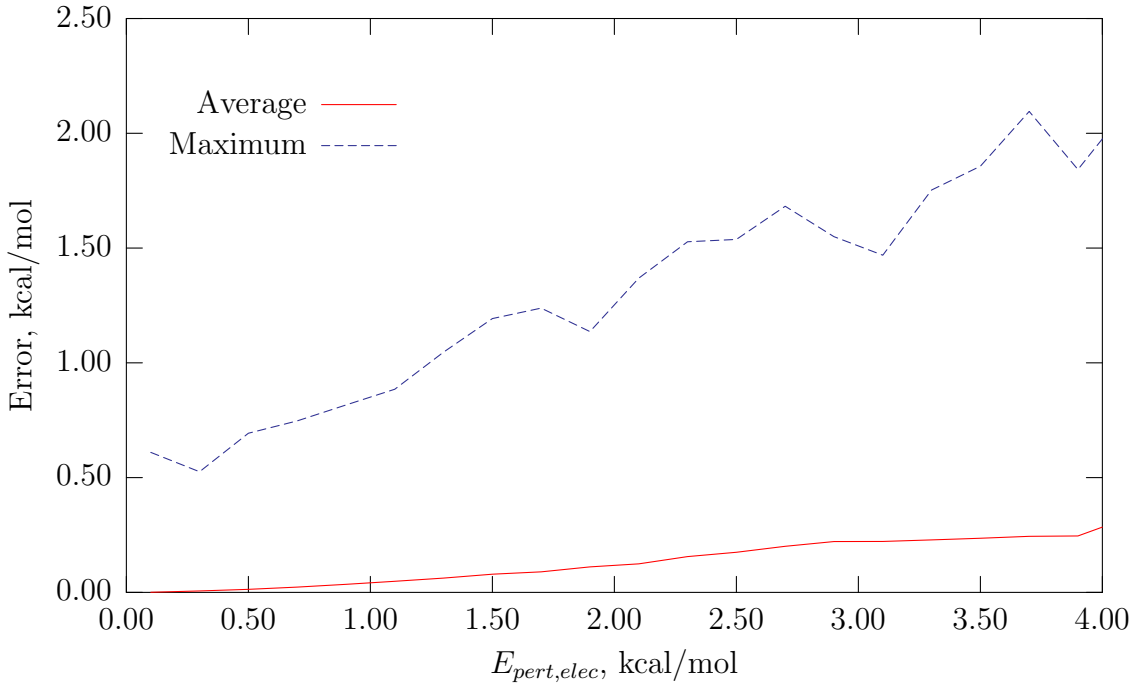

Figure 6.2: The variation in the error introduced by the perturbative QM/MC method, as a function of the perturbative QM energy delta, $E_{pert,elec}$.

The resulting errors were collected, as a function of the perturbative energy $E_{pert,elec}$, into bins 0.2 kcal/mol wide, and averaged. This average error is plotted in figure 6.2. Observe that, consistent with the results discussed in section 6.5, this error increases gradually as the perturbative energy increases. However, the *maximum* error for each bin is also plotted, and this is rather larger. Although that in itself is not surprising, and the number of these poorly-predicted energy deltas must be relatively small, in order for the average error to be so low, the error does not tend to zero as the perturbative energy does, and this suggests that the method introduces a systematic error.

An alternative method for minimization of the error introduced by perturbative QM/MC was suggested by Tuñón *et al.* [234], in which all moves less than a certain distance (5.0 Å, in their study) from any QM atom forced a full SCF calculation. For comparison, the same error data obtained from the simulation in this work were collected as a function of distance into bins 0.4 Å wide, and averaged. This distance was defined as that between the centre of mass of the molecule being moved by Monte

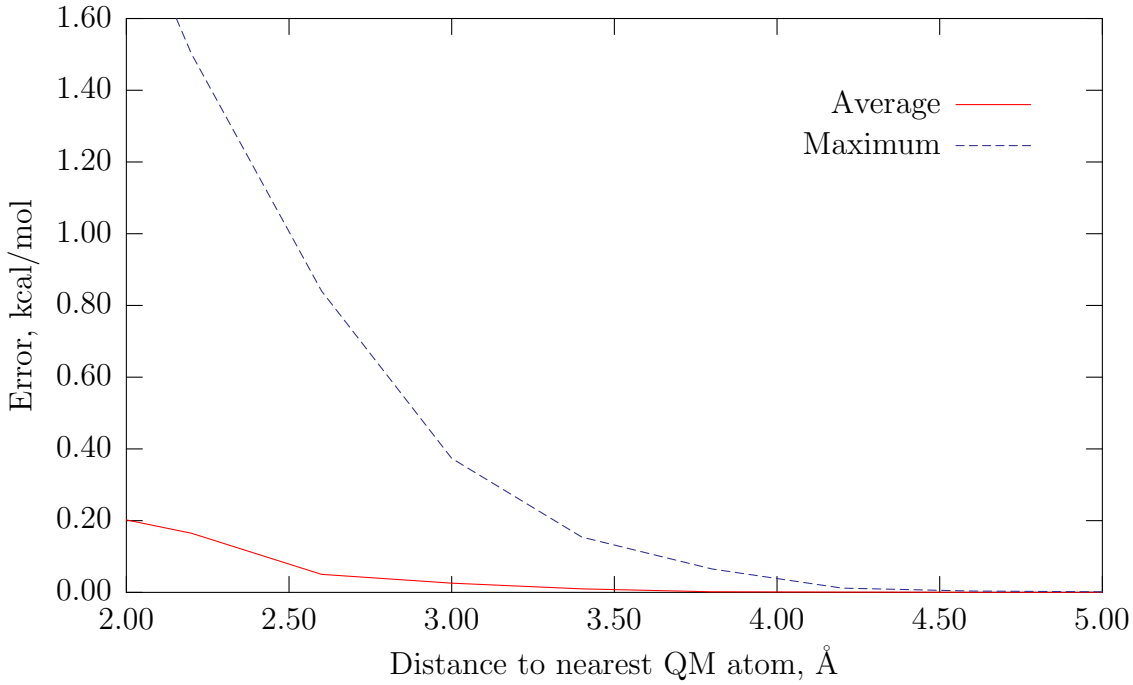

Figure 6.3: The variation in the error introduced by the perturbative QM/MC method, as a function of the distance from the nearest QM atom.

Carlo and the nearest QM atom, in the trial state. The resulting distribution is shown in figure 6.3; the error can be seen to increase dramatically as the distance decreases, behaviour consistent with that seen by Truong *et al.* [291, 292] and Tuñón *et al.* [234]. Again, the maximum error is much larger than the average, but falls off with increasing distance, such that virtually no errors are introduced for distances above 4.00 Å.

An examination of the most poorly-predicted energy deltas from figures 6.2 and 6.3 revealed that those for small perturbative energies in figure 6.2 generally occurred for moves very close to the QM atoms, while those for large distances in figure 6.3 had a rather large predicted energy delta from the perturbative method. Thus, a combination of the two criteria, energy and distance, might be expected to perform more effectively. Such a combination is demonstrated in table 6.4. The data from the last simulation were screened by applying cutoffs to both the perturbative QM energy delta and the distance to the nearest QM atom. Any MC move resulting in a distance larger than the distance cutoff, and a perturbative energy smaller than the

Distance cutoff	Energy cutoff	Average error	Maximum error	% SCF calcs
3.50	0.50	8.76×10^{-5}	0.058358	11.84
3.50	1.00	1.19×10^{-4}	0.097535	9.39
3.50	1.50	1.34×10^{-4}	0.097765	8.74
3.50	2.00	1.41×10^{-4}	0.097765	8.50
3.50	3.00	1.44×10^{-4}	0.097765	8.38
3.50	∞	1.45×10^{-4}	0.097765	8.37
3.00	0.50	2.63×10^{-4}	0.17186	9.76
3.00	1.00	4.17×10^{-4}	0.17186	6.54
3.00	1.50	5.24×10^{-4}	0.19962	5.49
3.00	2.00	5.81×10^{-4}	0.20544	5.02
3.00	3.00	6.16×10^{-4}	0.25385	4.71
3.00	∞	6.39×10^{-4}	0.3154	4.56
2.50	1.00	1.04×10^{-3}	0.31199	4.43
2.50	2.00	1.56×10^{-3}	0.41431	2.44
2.00	1.00	1.42×10^{-3}	0.63091	3.94
2.00	2.00	2.37×10^{-3}	0.93974	1.80
0.00	0.10	2.61×10^{-4}	0.48377	22.94
0.00	0.50	1.09×10^{-3}	0.60985	7.47

Table 6.4: Average and maximum error introduced by the perturbative QM/MC method, for various distance and perturbative QM energy delta cutoffs. All energies and errors are in kcal/mol, and all distances in angstroms.

energy cutoff, was included in the set of moves amenable to the perturbative QM/MC method. (All other moves would require a full SCF calculation for accurate energy determination.) Table 6.4 thus shows the average and maximum error for the set of moves that would be treated in a perturbative fashion in a 'real' simulation. Note that for a large distance cutoff, of 3.50 Å, the resultant error (both average and maximum) is very low. It is decreased slightly when an energy cutoff is also applied, although this comes at a cost; an increased number of moves fail the two criteria, and the percentage of energy calculations that require a full SCF rises concomitantly. For short distance cutoffs, of 2.00 or 2.50 Å, decreasing the energy cutoff from 2.00 to 1.00 kcal/mol improves the simulation by decreasing the maximum error quite dramatically. A more

obvious rise in the number of SCF calculations is also seen, since a large number of moves occurring at distances over 2.50 Å from the QM atoms alter the QM energy by between 1.00 and 2.00 kcal/mol. (Most moves at over 3.50 Å are of low energy, so altering the energy cutoff has less effect.) A reasonable compromise between CPU time and error is achieved with a distance cutoff of 3.00 Å; as the energy cutoff is decreased, a drop in the maximum error is observed for only a moderate increase in the number of SCF calculations.

For comparison, results are also shown for the case where no distance cutoff is applied, and an energy cutoff of 0.10 or 0.50 kcal/mol is imposed. This is similar to the test cases demonstrated in table 6.2. Note that an energy cutoff of 0.10 kcal/mol results in a much larger maximum error than a distance cutoff of 3.50 Å (with no energy cutoff), while requiring more than twice the number of SCF calculations.

6.8 Umbrella sampling

For a comprehensive test of the perturbative QM/MC method, and the use of hybrid QM/MC techniques in general, the free energy profile of the Menshutkin reaction was determined by umbrella sampling and compared with that obtained in section 5.3.4. The same system used in Chapter 5 was employed; the reactive region was treated with the AM1 method, and solvated by 336 TIP3P waters, in a periodic cubic box of side 21.724 Å. MM/MM electrostatic forces were shifted to zero at 9.5 Å, while the QM/MM interactions experienced no cutoff, but were corrected for periodic boundary conditions.

For each window in the umbrella sampling simulation, the reaction coordinate was held near the reference value by application of a harmonic restraint term with force constant 800 kcal/mol/Å^2. Each simulation sampled 4×10^6 configurations, chosen using the move set given in table 6.1. The perturbative QM/MC method was used to reduce the number of SCF calculations required, using a distance threshold of 3.0 Å,

and per-step and cumulative energy thresholds of 1.0 kcal/mol. A full QM calculation was forced after 1000 steps if none of these thresholds had been exceeded.

The umbrella sampling run was started by taking the pre-equilibrated solvated Menshutkin system from section 5.3.4 at R_C = -0.50 Å, and sampling one window with a reference coordinate of -0.50 Å. The final coordinates from this simulation were then used as the initial state for two further windows, centred at -0.55 and -0.45 Å. Two calculations were thus carried out simultaneously, in the same fashion as in section 5.3.4, for reaction coordinate values ranging between -1.60 and 2.50 Å, separated by 0.05 Å. The resulting outputs were then joined by weighted averaging, as in section 5.3.4, ignoring the first 5×10^5 steps from each window and using bins 0.005 Å wide, to give a potential of mean force for the entire reaction. This is shown in figure 6.4; for comparison, the results of the molecular dynamics simulation using the same solvent model (from section 5.3.4) are also shown. It should be remembered, however, that the two simulations are not directly comparable, as the MD simulation yielded Gibbs free energies, while the Monte Carlo simulation, as it was carried out in the NVT ensemble, provides Helmholtz free energies.

The match between the Monte Carlo and molecular dynamics simulations of the Menshutkin reaction is very good. In fact, the shapes of the two curves are almost identical, differing only for $R_C < -0.6$. This suggests that the NVT ensemble is a reasonable approximation for NPT, and is consistent with the observation made in section 5.3.4 that the side lengths of the periodic box in the QM/MD simulation fluctuated by less than 0.2 Å.

The activation free energy predicted by QM/MC is 26.9 kcal/mol, which is in excellent agreement with the previous two-dimensional QM/MC study of this reaction [258], which yielded 26.3 ± 0.3 kcal/mol.

Over the entire reaction coordinate, full QM calculations were required for 3.17% of the configurations considered. (This is significantly less than that seen in table 6.4,

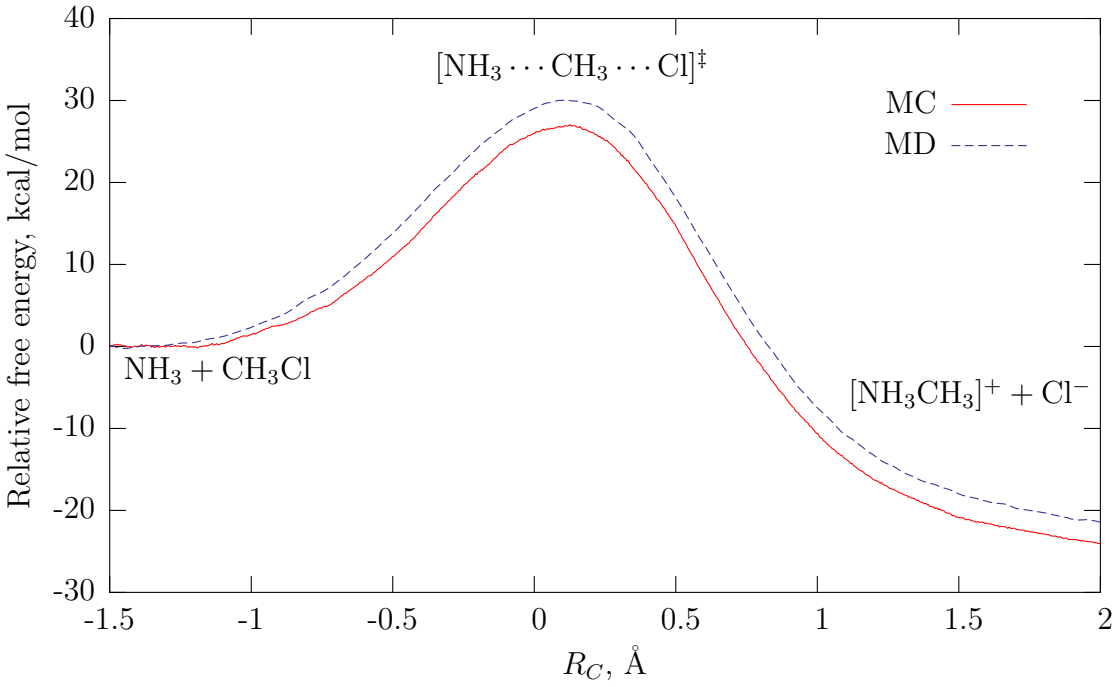

Figure 6.4: Determined solution-phase potential of mean force of the Menshutkin methyl chloride/ammonia reaction, using AM1 for the reactive region and TIP3P for the solvent. The PMF was determined using Monte Carlo simulation at constant volume; for comparison, constant pressure results, from the molecular dynamics simulation originally displayed in figure 5.10, are also shown.

as preferential sampling was not used.) Thus, each window required around 10^5 QM calculations, compared to the 70000 needed for the MD windows, and so the MC and MD simulations took comparable amounts of time to run, the MC being up to 50% more expensive (i.e. an average of about 4 hours of CPU time per window on a 1667 MHz Athlon XP machine). An average acceptance rate of 48.78% was achieved.

MC differs from MD only for description of the reactants — the MC PMF is essentially flat for R_C < -1.1 Å, while the MD surface is still curved even at -1.3 Å. This is probably due to the differing treatment of the NH_3 group in the two methods (in the frame of reference of the CH_3 group); MD allows full rotation of the group, and independent motions of the individual hydrogens, subject only to the constraint that N–C–Cl remains linear, while the symmetry of the MC approach keeps all three hydrogens in a plane perpendicular to the N–C–Cl bond. (Such a difference is not seen

for the products, as the chloride ion is spherically symmetric.) This extra rotational and vibrational freedom would be expected to increase the entropy of the reactants under the MD description relative to the MC, and thus decrease the free energy, leading to a higher activation free energy. It is also possible that part of the discrepancy is caused by insufficient sampling from the MC or MD approaches. Longer simulations, with suitable MC moves to describe the rotation and vibration of the NH_3 group, would be required to better explain the differences between the MC and MD approaches.

6.9 Improvement of the energy function

The results presented thus far are in reasonable agreement with other theoretical studies, and with experimental data for the methyl iodide/ammonia reation; however, the match between QM/MM simulations and experimental data can be improved in several ways. Improvement of the MM representation by use of the FlucQ force field has already been investigated in Chapter 4, as has refinement of the QM/MM interaction energy by parameterization of the van der Waals term (section 5.3.2). However, it must be remembered that an accurate representation of the QM region is vital (if it were not, the region could be modelled with an MM force field, or an approximate treatment of bond breaking such as EVB [278] could be used). Thus, moving to a higher level of theory and/or a larger basis set is desirable. In the case of the Menshutkin reaction study, the semi-empirical AM1 method could be replaced by Hartree Fock or DFT, and a basis such as 6-31G* could be employed. (An *ab initio* technique also confers the advantage that the methyl iodide system can be studied, enabling direct comparison with experiment.) Of course, the obvious downside is that these more rigorous treatments are far more computationally demanding, increasing the time taken to perform an SCF calculation by several orders of magnitude. Even if SCF-avoiding techniques such as perturbative QM/MC (section 6.3) are used, the larger basis sets still greatly increase the time taken to evaluate the 1-electron integrals.

It is well known that even quite a low level of theory such as HF/3-21G* can give reliable geometries for systems of interest, even if the energies are not particularly well reproduced [295]. This is evidenced by the use of split-level methods [36] such as Gaussian-1 [296, 297] and Gaussian-2 [298, 299], where the geometry of the system is optimized at one level of theory, and then the energy corresponding to that geometry is determined at a higher level.

6.9.1 The ONIOM method

One method for improving the accuracy of QM calculations is known as ONIOM [176, 177]. In the ONIOM formalism, the full system is referred to as the 'real' geometry. A subset of the system, termed the 'model' geometry, is taken to be reasonably representative of the 'real' system. Then the energy of the 'real' system at 'high' level (i.e., with more rigorous theory and/or a larger basis set than at 'low' level), $E(high, real)$, is approximated as:

$$E(high, real) = E(low, real) + E(high, model) - E(low, model) \tag{6.12}$$

Clearly, the ONIOM method relies on the following being true:

$$E(high, model) - E(low, model) \simeq E(high, real) - E(low, real) \tag{6.13}$$

Put another way, ONIOM depends on the energy of the part of the 'real' system not covered by the 'model' being largely invariant to the change from low to high level. This requires a careful choice of model system, particularly with the more advanced ONIOM treatments where a three-level separation is employed, introducing an intermediate level of theory and a corresponding intermediate model system.

6.9.2 Consideration of the QM/MM interaction energy

The applicability of methods similar to ONIOM to QM/MM simulations is determined by the degree to which the interaction energy between the QM and MM regions is independent of the level of QM theory and basis set. A recent study by Byun *et al.* [300] suggested that this energy is indeed independent, at least for some systems.

This suggestion was tested for the Menshutkin reaction by evaluating the energy of a number of configurations, all derived from the Menshutkin reaction region at R_C = -0.50 Å, solvated by 336 TIP4P water molecules in a 21.724 Å cubic box and equilibrated for 30 ps. The full solution-phase QM/MM energy for each configuration was evaluated using a variety of QM treatments, and then the solvent molecules were removed and the purely QM energy of the gas-phase 9-atom Menshutkin system was determined using the same treatments. By then calculating the difference between the two energies, the perturbation introduced by the solvent was evaluated. By consideration of equation 1.57, this solvent perturbation energy, $E(solv)$, is given in terms of the solution-phase energy $E(s)$ and the gas-phase energy $E(g)$ as:

$$\begin{aligned} E(solv) &= E(s) - E(g) \\ &= E_{MM}(s) + E_{QM/MM}(s) + E_{QM}(s) - E_{QM}(g) \end{aligned} \tag{6.14}$$

The perturbation thus consists of the solvent MM energy (which is obviously invariant to the QM treatment used), the QM/MM interaction energy, and the difference in the QM energy between the gas and solution (as a result of the altered electron density induced by the solvent).

The solvent perturbation energies as a function of the QM treatment used are displayed in figure 6.5. Configurations 0 through 35 were generated by taking the original solvated Menshutkin system, and rotating the methyl group about the N–C–Cl

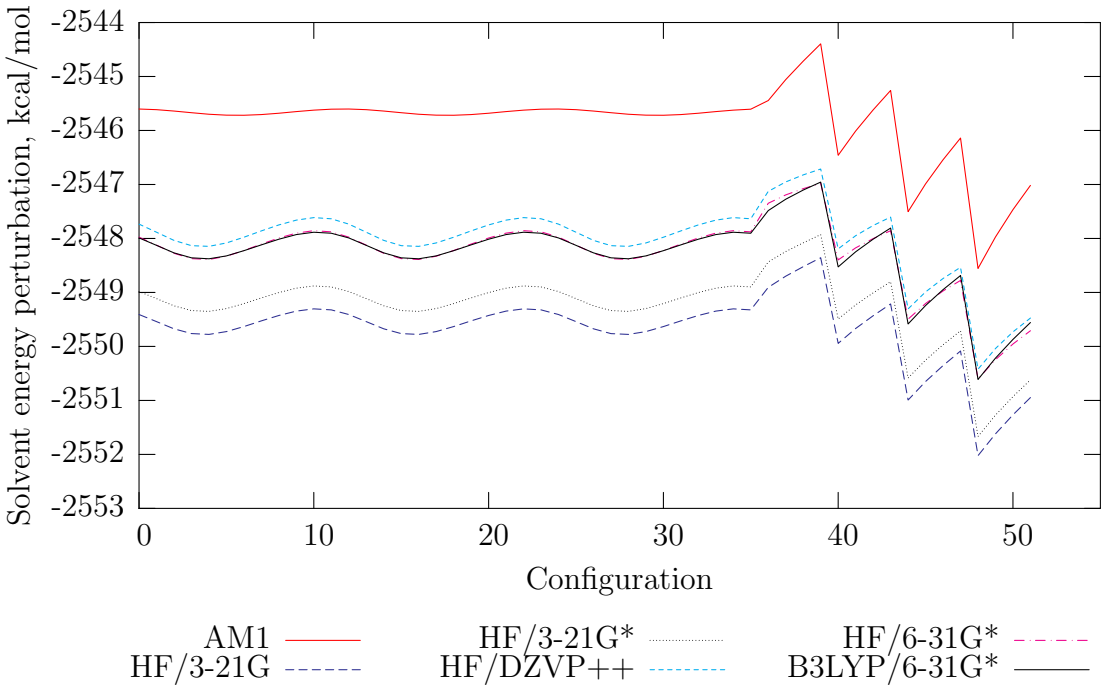

Figure 6.5: Solvent energy perturbation for the Menshutkin system at different levels of QM theory.

bond in 10° increments[2]. (This corresponds to move (3) in figure 6.1.) The remaining configurations were used to map the energy surface determined by changes in the C–Cl and C–N bond distances. Starting from the original configuration, the C–Cl distance was varied from 1.8 to 2.1 Å in 0.1 Å increments; for each C–Cl distance the C–N bond was varied from 2.2 to 2.5 Å, also in 0.1 Å increments. (These two moves correspond to (2) and (1) respectively in figure 6.1.)

It can clearly be seen from figure 6.5 that the solvent energy perturbation is largely independent of the level of QM theory. The very close agreement between B3LYP/6-31G* and HF/6-31G* demonstrates that electron correlation has very little effect on this perturbation energy. In fact, all of the *ab initio* methods considered give energies that agree very closely with each other; although the absolute energies may

[2]If computer time were in very short supply, the number of these configurations could be greatly reduced by taking advantage of the QM system's C_{3v} symmetry.

Method	B3LYP/6-31G* Energy shift, kcal/mol	B3LYP/6-31G* RMS, kcal/mol	HF/DZVP++ Energy shift, kcal/mol	HF/DZVP++ RMS, kcal/mol
AM1	-2.4185	0.1767	-2.1727	0.2010
HF/3-21G	1.4104	0.0080	1.6562	0.0497
HF/3-21G*	0.9916	0.0227	1.2374	0.0311
HF/DZVP++	-0.2458	0.0433	0.0000	0.0000
HF/6-31G*	-0.0083	0.0455	0.2376	0.0147
B3LYP/6-31G*	0.0000	0.0000	0.2458	0.0433

Table 6.5: Energy shifts and RMS differences between the solvent energy perturbation with various QM methods.

differ by one or two kcal/mol, the energy curves[3] are of very similar shapes. AM1, on the other hand, is in comparatively poor agreement, most obviously under-estimating the energy variations as the methyl group is rotated. This is shown a little more rigorously in table 6.5, where the individual energy curves have been shifted to minimize the RMS error between them. It can clearly be seen that the RMS deviation between the AM1 method and higher methods using more rigorous theories (such as B3LYP/6-31G*) or larger basis sets (HF/DZVP++) is much larger than that of small basis set *ab initio* treatments. This difference in the absolute energy perturbation is presumably due to the electron density in the QM region being systematically under- or overestimated at low levels of theory.

It follows from the above discussion that if the correction to the *gas phase* (QM) energy between low and high levels of QM theory is known, then the corrected energy of the whole combined QM/MM system can be determined using the same correction, plus a constant term to account for the shift in the solvent energy perturbation between the two levels of theory. (This constant term is of course not necessary if only relative energies are required.) Thus, ONIOM is a reasonable approximation for this system,

[3]Note, however, that only the relative shape of the curves is important; the shape of each individual curve is purely a function of the set of configurations used, and has no physical significance.

where the 'model' system corresponds to just the QM region of the 'real' QM/MM system. However, the ONIOM approximation is expensive, as three separate QM calculations are required: the full QM/MM system energy at low level, and the energy of the QM system at both low and high levels. This thus only becomes a reasonable method when the QM/MM calculations are much more expensive than the QM, i.e. for very large MM systems where a massive number of 1-electron integrals must be calculated, and many SCF cycles are required to optimize the wavefunction in the presence of so many point charges.

6.9.3 Approximate correction of the total QM/MM energy

The expense of the ONIOM correction at every energy evaluation can be avoided if this correction is known beforehand. Such a precalculation is not as difficult as might be imagined; while, for a full treatment of the 9-atom Menshutkin system, the correction from low to high energy would need to be known at every point in the $(3N - 6)$ or 21-dimensional internal coordinate space, some degrees of freedom are neglected in the move set used in Monte Carlo simulation. As can be seen from consideration of figure 6.1, the configuration of the QM region at any point can be exactly described by seven numbers, denoting the size of each of the seven MC moves required to produce that configuration from some reference configuration. Thus, the problem is reduced from 21 to 7 dimensions. Furthermore, if it is assumed that the symmetric stretch and bend moves (4 through 7, in figure 6.1) have little effect on the system energy (or at least, that the low level of theory can describe them about as well as the high level) then the problem is reduced to 3 dimensions. Finally, if it is assumed that rotation of the methyl group is largely independent of variations in the C–Cl and C–N distances, then two corrections are required: a 2-dimensional correction as a function of these two bond distances, and a 1-dimensional correction as a function of the H–N–C–H dihedral angle. This method could also be applied to larger QM regions, provided that

the number of degrees of freedom is not too large; for example, a similar method has been recently applied to methyl acetate, where only one degree of freedom (dihedral rotation about the ester C–O bond) was used in the correction [300].

CHARMM implementation

For the Menshutkin system in the region $-0.5 < R_C < 0.5$ Å, the configurations whose energies are shown in figure 6.5 were used as a template. A larger number of C–Cl and C–N configurations were generated, however, varying C–Cl from 1.7 to 2.4 Å, and C–N from 1.4 to 2.5 Å, both in 0.1 Å increments. For each of these configurations, the QM region geometry was optimized at AM1 level, keeping the two bond distances fixed, with 200 steps of Adopted Basis Newton Raphson minimization, so that representative hydrogen positions were used. The difference between the energies of the high and low QM regions was then simply tabulated as a function of the dihedral angle (for the first 36 configurations) or of the C–Cl and C–N bond lengths (for the remainder). Each dihedral energy correction was shifted such that it was zero for a 180° angle, as all of the remaining configurations, having been optimized at fixed bond lengths to minimum energy structures, included an implicit correction for a 180° dihedral. For the remainder of the reaction coordinate in the range $-2.0 < R_C < 3.0$ Å, additional C–Cl and C–N configurations were required; $R_C > 0.5$ Å was handled by considering a grid of points spaced by 0.2 Å in the range $1.4 \leq r(\text{C–N}) \leq 2.2$ Å, $2.6 \leq r(\text{C–Cl}) \leq 4.6$ Å, while another 0.2 Å grid of points in the range $2.7 \leq r(\text{C–N}) \leq 3.9$ Å, $1.5 \leq r(\text{C–Cl}) \leq 2.1$ Å covered the reaction for $R_C < -0.5$ Å. It can be seen from inspection of figure 6.6 that these points cover the C–N and C–Cl bond space explored by the minimum energy gas phase path.

CHARMM was then modified to include a suitable correction at each energy evaluation; after determining the total system energy at low level, the H–N–C–H dihedral angle was calculated, and the energy corrected by adding a value obtained by

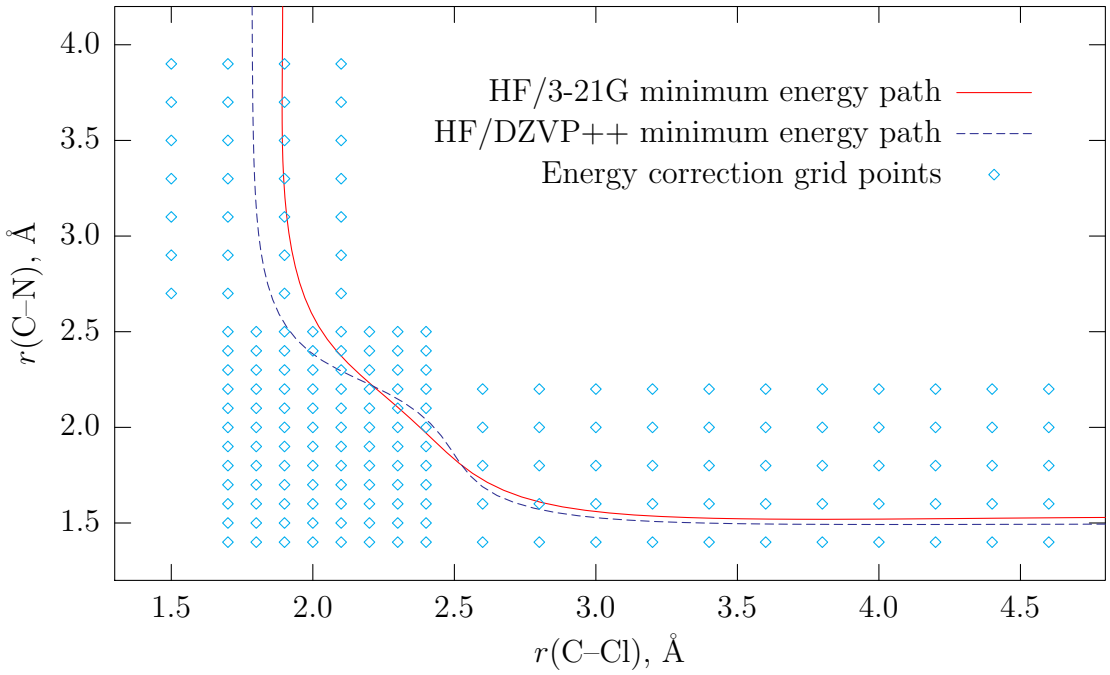

Figure 6.6: The minimum energy path from adiabatic mapping at HF/3-21G and HF/DZVP++ levels. The points at which the bond distance energy correction term was evaluated are also indicated.

a simple linear interpolation of the two closest tabulated dihedral energy corrections. Then the C–Cl and C–N distances were evaluated and used to further correct the energy, by a 2-dimensional interpolation of the 4 closest tabulated distance energy corrections.

6.9.4 Energy test cases

The performance of the correction method was evaluated by using it to calculate the energy of several solution-phase configurations obtained from an earlier QM/MC simulation at AM1 level (they were in fact the configurations obtained at the end of each window in a QM/MC umbrella sampling simulation, and cover the reaction coordinate in the range $-1.0 \leq R_C \leq 0.5$ Å). In figure 6.7 the agreement is shown between the energies of the 9-atom QM region only, determined at HF/DZVP++ level, and those determined at HF/3-21G level and then corrected to HF/DZVP++ level (denoted

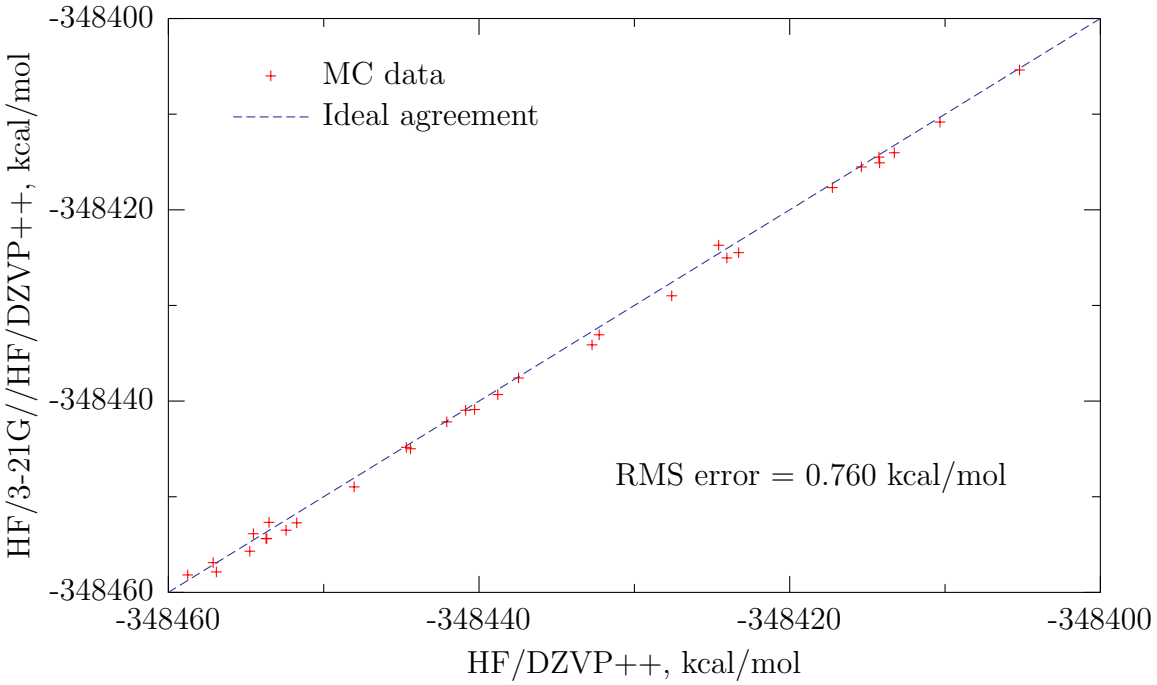

Figure 6.7: Agreement between corrected HF/3-21G energies and true HF/DZVP++ energies for the 9-atom Menshutkin reaction region, in the gas phase.

by HF/3-21G//HF/DZVP++) by application of the technique described above. In figure 6.8 the same correction is applied to the entire solvated system, and a constant term of 1.6562 kcal/mol (from table 6.5) is added to account for the difference in the solvent energy perturbation between the two levels of theory.

The RMS error in the gas phase results, of 0.760 kcal/mol, probably accounts for the majority of the RMS error in the full system results. This error is obviously introduced by the assumptions made in the method, and could be reduced by increasing the dimensionality of the correction term, for example by including the C–H distances and angles, or by removing the assumption that the dihedral correction is independent of the bond distance correction. A cubic spline [205] or sinusoidal function fitted to the correction table data would also undoubtedly give a better energy correction than the simple linear interpolation used here. The error for both these results and for the whole system results could also be decreased by using a higher level of theory for the low level, for example HF/3-21G*, although obviously this would require more computer time.

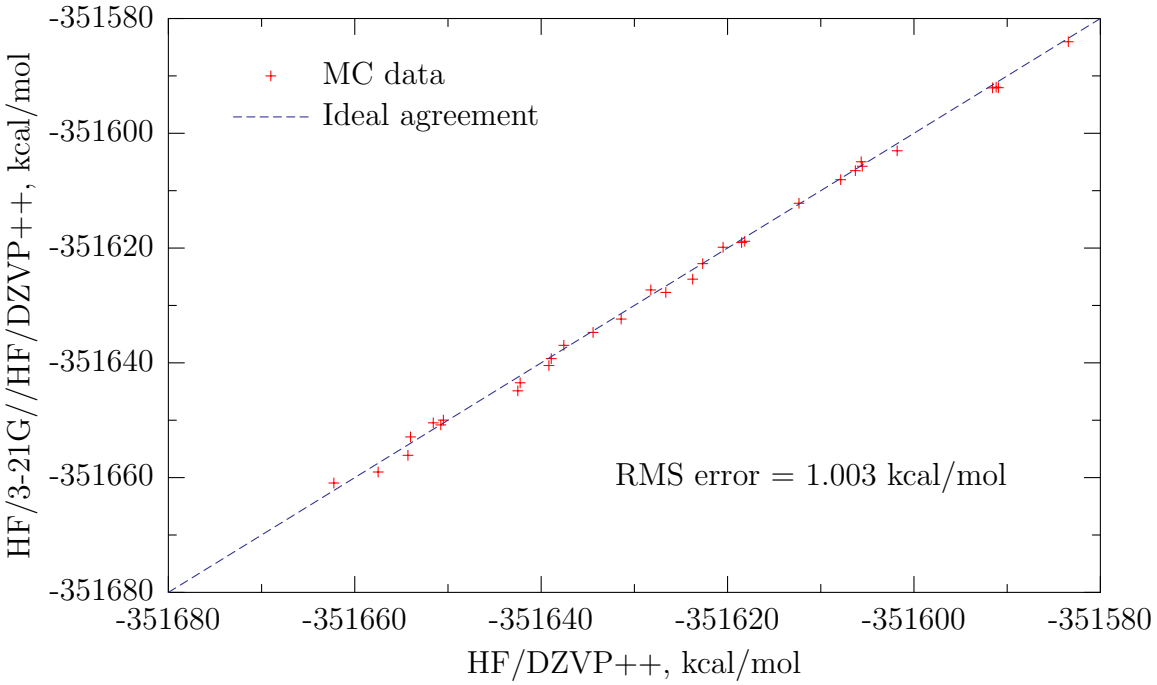

Figure 6.8: Agreement between corrected HF/3-21G QM/MM energies and true HF/DZVP++ QM/MM energies for the solvated Menshutkin system.

The error introduced by the approximation for the QM region can be investigated by evaluating the ONIOM-corrected energy for the entire solvated system. This is shown for the same set of configurations in figure 6.9. As suggested above, the QM/MM energy of the full system was evaluated at low (HF/3-21G) level, and the energy of the 9-atom QM region at both low and high (HF/DZVP++) levels, to make the ONIOM correction, by applying equation 6.12. The resultant ONIOM energy was then further shifted to account for the shift in the solvent energy perturbation between the two levels of theory. It can clearly be seen by comparison with figure 6.8 that the ONIOM method gives more reliable energies than the approximate method, although the errors introduced are still significant, and it must be remembered that this gain in accuracy comes at a large computational cost. The remaining errors are no doubt due to the lower level of theory misrepresenting the electron density; although figure 6.5 would suggest that this error is largely systematic, it must be remembered that only one MM configuration was used to derive these curves. Clearly some of the QM/MM

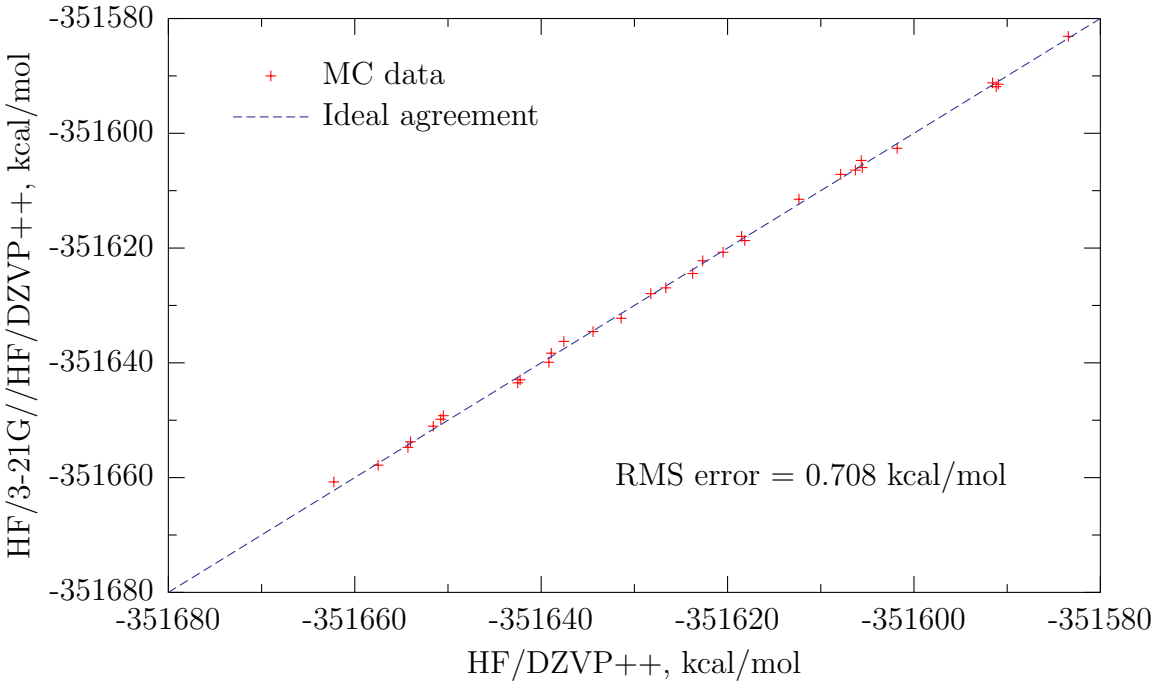

Figure 6.9: Agreement between ONIOM-corrected HF/3-21G QM/MM energies and true HF/DZVP++ QM/MM energies for the solvated Menshutkin system.

interactions, most likely those with MM molecules in the first solvation sphere of the QM region, are highly density-dependent.

For comparison, the uncorrected low level energies for the solvated system are compared with high level energies in figure 6.10. Relative energies are shown; the energy zero for each method was chosen by least squares optimization to minimize the RMS error between the two methods. It can thus be seen that errors in the total QM/MM energies are greatly reduced by application of this correction technique.

6.9.5 Minimization test cases

The utility of the method for minimization was investigated by using it to generate adiabatic maps for the gas-phase reaction, and comparing the results to those obtained in section 5.2.3. As the method corrects only energies, the first-order minimization approach used for the original maps could not be used; thus, the results were obtained by simulated annealing (see section 1.9.1).

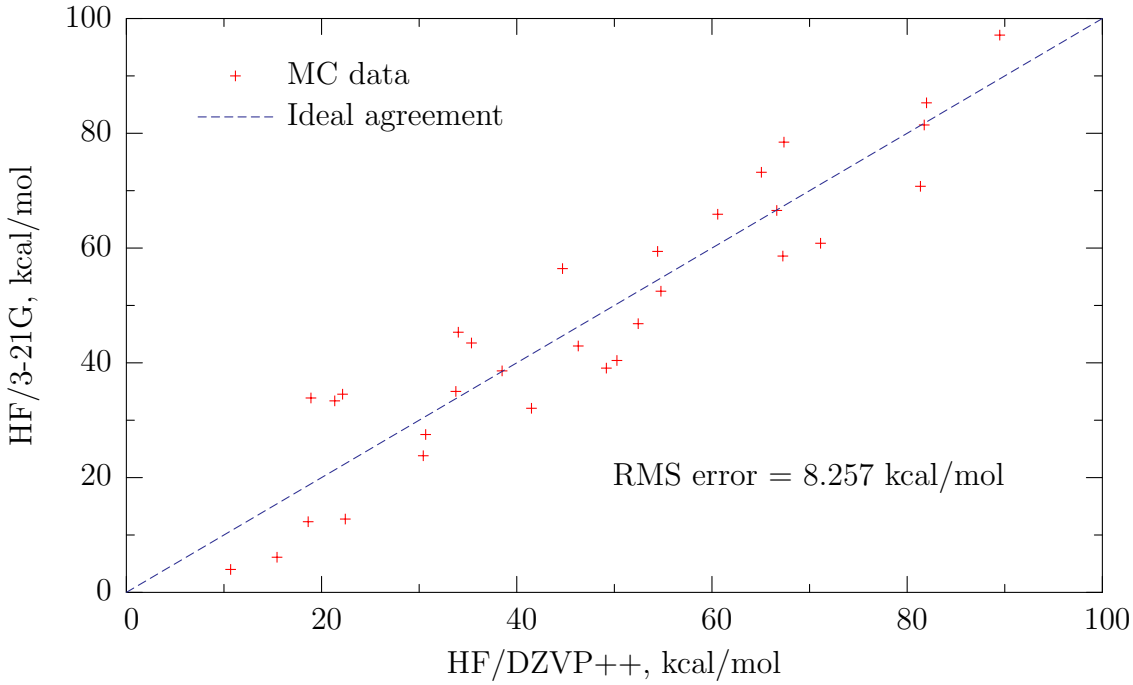

Figure 6.10: Agreement between uncorrected HF/3-21G QM/MM energies and HF/DZVP++ QM/MM energies for the solvated Menshutkin system. The energy zero for each method was chosen to yield the lowest RMS.

The move set used for the simulated annealing minimizations is shown in figure 6.6. This move set can be seen to be very similar to that used for the solvated Menshutkin system in figure 6.1, with the obvious exception of the solvent moves. The maximum move sizes were all greatly reduced for simulated annealing, so that a reasonable fraction of moves would be accepted during the cooling phase. The weights of the rotation, bend, and stretch moves were also increased relative to the z axis translations, as sampling of the reaction coordinate is of less importance in this minimization, and improved sampling of these moves should enable the system to better explore the energy surface.

At each point along the reaction coordinate, the system was first equilibrated for 200 Monte Carlo steps, with a simulation temperature of 298 K. This was then followed by 700 steps of simulated annealing, where the temperature T was varied as a function

Move type	Max. move size	Weight
NH_3 z axis translation	0.02 Å	1.0
Cl z axis translation	0.02 Å	1.0
CH_3 rotation about z axis	10.0°	0.20
CH_3 symmetric bend	2.0°	0.20
NH_3 symmetric bend	2.0°	0.20
CH_3 symmetric stretch	0.01 Å	0.20
NH_3 symmetric stretch	0.01 Å	0.20

Table 6.6: The weights and maximum move sizes of the move set used in Monte Carlo simulated annealing studies of the Menshutkin QM region.

of the MC step number, s:

$$T = T_0 \exp(-\alpha s) \tag{6.15}$$

The annealing run was started at temperature $T_0 = 298$ K. The parameter α, which controlled the rapidity of the cooling, was chosen as 0.01. This resulted in a gradual cooling of the system, to 0.27 K at the end of the run.

Two scans of the reaction were carried out, both covering R_C over the range from -2 to 3 Å. Each scan was started with a previously-minimized HF/3-21G gas phase structure at $R_C = 0.5$ Å, and then minimized the system energy every 0.1 Å along the reaction coordinate. Each scan was carried out in two directions along the reaction coordinate simultaneously. The starting structure for each minimization was taken as the final structure from the previous reaction coordinate point (i.e. that at 0.4 Å was taken from the minimized structure at 0.5 Å, and so on).

The first scan was carried out at HF/3-21G level, and the second at the same level, but using the approximate correction term to correct the energy to HF/DZVP++ quality. The resultant adiabatic maps are shown in figure 6.11. For comparison, the maps obtained in section 5.2.3, at HF/3-21G and HF/DZVP++ levels and by first-order minimization, are shown on the same figure. The HF/3-21G results using the two methods are virtually identical, with an RMS error of 0.033 kcal/mol, strongly

suggesting that the cooling schedule used for the simulated annealing is adequate, and that both methods have found the global minimum. This in turn suggests that deviations between the true HF/DZVP++ adiabatic map and that obtained using the approximate correction term are due largely to deficiencies in the correction. That said, the agreement in this case is also good, with an RMS error of 0.96 kcal/mol, with the trends in the HF/DZVP++ energy being well reproduced by the corrected data; in particular the positions of the extrema, at around 0.5 and 1.4 Å, match closely. The energies of these extrema are less well reproduced, however, with the approximate method overestimating them by about 0.6 and 1.4 kcal/mol, respectively. It can be seen that this is due to a systematic error introduced by the correction method; all the points for $R_C \geq 0$ Å are slightly overestimated in energy. This is because no correction is made for the N–H and C–H angles and distances, and so every minimum energy structure will have these angles and distances at their HF/3-21G predicted values, rather than those at HF/DZVP++ level.

6.9.6 Caveats

The method outlined and tested in this section has been shown to give very reasonable approximations to high-level QM and QM/MM energies, while taking a fraction of the time of the more rigorous ONIOM technique. The accuracy could be increased by using more grid points in making the correction, using a better interpolation method, or moving to a better QM treatment for the low level energy. The method is completely general, and can be applied to correct the QM energy from any method or basis set to any other; the only restriction is that the error introduced depends strongly on the number of degrees of freedom of the system.

It should be noted, however, that the method described here corrects only the total *energy* of the system. It is thus ideal for QM/MC simulations, but without forces it is unsuitable for QM/MD in its current form. Secondly, unconstrained MD explores all

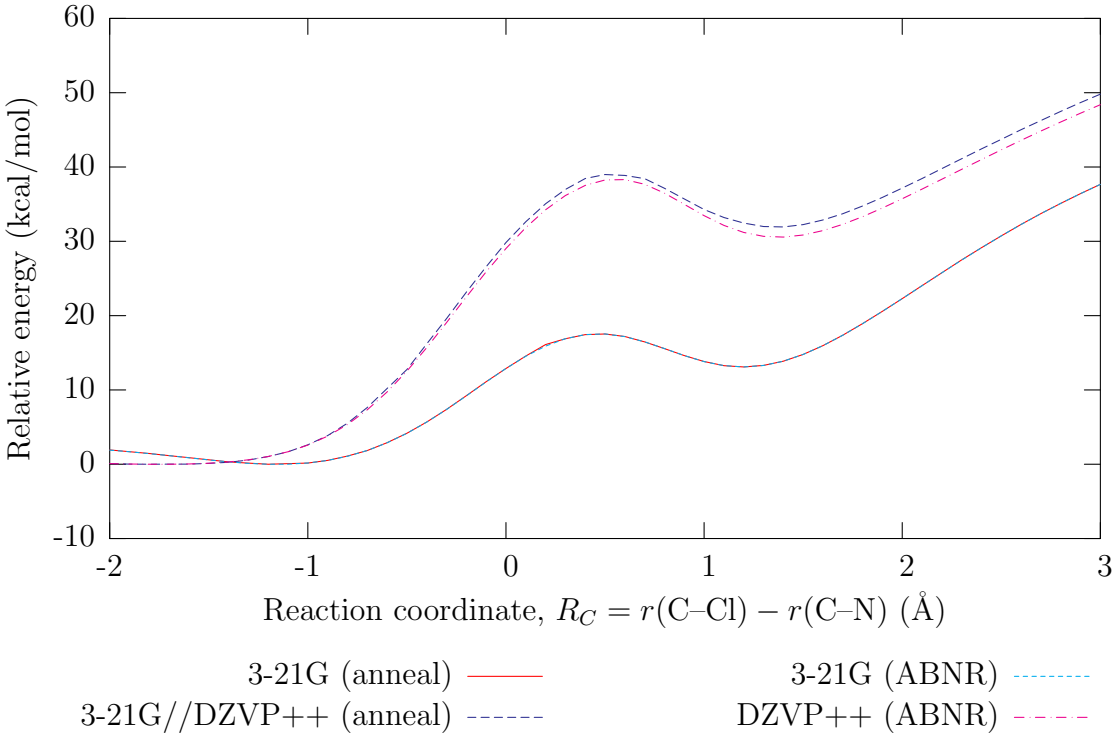

Figure 6.11: Minimum gas phase energies as determined by simulated annealing. The corresponding profiles from ABNR (Adopted Basis Newton Raphson) minimization, originally shown in figure 5.2, are displayed for comparison.

degrees of freedom of the QM region, taking no advantage of any symmetry, and thus does not fit well with the reduced dimensionality correction used here; while the degrees of freedom used in the correction (e.g. the two bond lengths and one dihedral used for the Menshutkin system) are probably still the most important in an MD simulation, the errors introduced for moves along the other degrees of freedom would probably be unacceptable. In principle, however, it should be possible to obtain corrected forces by simply evaluating their differences between the high and low levels of theory, in the same way as the energy is corrected. Correction of forces is also possible in this way with the ONIOM [177] method, where full information on the QM region at both levels of theory is available, and no constraints are required.

6.10 Summary

Monte Carlo simulations in connection with a QM/MM energy description have been shown to give very reasonable free energies for the Menshutkin reaction, in good agreement with previous molecular dynamics studies in this thesis. When combined with the perturbative QM/MC method, the amount of computer time required for the simulations drops enormously (to 3.17% of the original, in the umbrella sampling simulations discussed in section 6.8), with little reduction in the quality of the treatment. This puts QM/MC on an equal footing with QM/MD in terms of the time required to simulate a system; QM/MC, however, has other advantages above QM/MD. It is ideally suited for highly symmetrical systems such as the Menshutkin S_N2 reaction; secondly, its handling of a reduced number of degrees of freedom, and lack of requirement for forces, allows the QM energy to be corrected by means of an approximate technique, which has been shown to give very reasonable agreement for energies in the gas and solution phases, and to reproduce the results of minimizations of the gas phase system. This would allow future *ab initio* simulations of the solution-phase reaction to be carried out at a low and computationally inexpensive level of QM theory, while yielding information comparable to that obtained from extremely expensive high-level calculations.

Chapter 7

Conclusions and future research directions

7.1 Conclusions

The work presented in this thesis provides a critical examination of several important aspects of the simulation of chemical reactions with QM/MM methods. The deficiencies of both the MM and the QM treatments are tackled, and solutions are suggested. A combination of these solutions should enable the QM/MM treatment to be used to obtain higher accuracy information on systems of biochemical interest, such as enzymes and the reactions catalyzed by them.

Computational infrastructure

Modern computational chemistry requires large amounts of computer power, particularly when advanced theories are used to study large systems, and many configurations are required for adequate sampling. It is shown in Chapter 2 that for many applications conventional supercomputers can be replaced with clusters of commodity machines, on which many shared-memory algorithms can be refactored as distributed-memory applications, and run efficiently for a fraction of the price.

There has been a good deal of interest in recent years in techniques for reducing the inter-process communication required by computational chemical methods, and for improving their scalability. These include linear scaling [12, 13] and divide-and-conquer [301,302] approaches, and the increasing use of such techniques will enable the effective use of larger clusters.

Computational chemistry is also an early adopter of Grid technology and similar methods for the creation of virtual supercomputers on demand, with massive parallel processing applications such as the Cancer Screensaver [136, 137] and the Folding@Home project [140]. While it is certainly true that commodity clusters cannot be used to solve all problems, particularly those for which large amounts of fine-grained communication are required, the increasing use of Grid-type technologies suggests that dedicated shared-memory supercomputers are likely to become less and less common, but available as shared resources for specialist applications.

QM/MM methodology

The QM/MM method has been developed and widely used over the last decade. However, it is still difficult to use, and error-prone if not applied correctly; it is also much more expensive than MM treatments, so must be used efficiently. A number of problems and subtleties of the method are considered in this thesis, and either solved or side-stepped. For example, the interaction of the QM region in CHARMM with a periodic MM system fails if the QM system drifts from the centre of the periodic box. It is found in this work that adding periodic boundary conditions to the QM/MM interactions, and using no cutoffs, gives satisfactory results. It is also shown, from calculation of the water dipole, that the choice of QM method has a large effect on the determined properties, and so must be carefully chosen. Finally, while an extensive choice of QM methods is available for use with the method, the relatively poor MM

descriptions are still a limitation, which the Fluctuating Charge technique aims to solve.

The FlucQ model

The FlucQ model, described in Chapter 3, adds efficient polarization to the MM treatment. The original model is implemented in CHARMM as part of this work, with additional modifications based on the work of other researchers to allow Monte Carlo simulation and the evaluation of the QM/MM interaction. Thermostatting is also added to allow long, stable simulations, while the original rather expensive integration method for determining intramolecular interactions is replaced by a much faster approximate technique.

The FlucQ model is shown to perform excellently for bulk solvent, the water dimer, the gas-phase water monomer, and water-ion complexes, with both MM and QM/MM treatments. Energies and forces can be evaluated extremely efficiently using an extended Lagrangian molecular dynamics method, increasing computational cost relative to conventional static charge treatments by only 4% when using standard electrostatics methods. The method gives virtually identical bulk solvent properties when using Monte Carlo techniques, and although these are less efficient, they offer the possibility of simulating phase equilibria in the Gibbs ensemble [201, 208, 303].

The technique is not a solution to all problems; it over-polarizes in the presence of highly-charged compact ions such as fluoride, and does not perfectly reproduce the ESP-fitted charges of QM methods. The TIP4P-FQ1 model is proposed as a first step towards solving these problems, although this model in turn does not produce reasonable bulk solvent behaviour, showing that the FlucQ model does not correctly model all aspects of the electron density. Solute-solvent binding energies are also rather poorly reproduced when using FlucQ in combination with the AM1 method for the Menshutkin reaction, although this is likely to be largely caused by the AM1 method

itself, which treats the QM electron density in such an approximate fashion that any QM/MM interaction must be carefully parameterized.

Free energy simulation

Most of the salient features of the Menshutkin S_N2 reaction can be reproduced by the QM/MM method, even while using the very approximate AM1 treatment for the QM region. It is vital, however, to use appropriate parameters for the QM/MM van der Waals interaction, as standard MM parameters give very poor agreement with *ab initio* binding energies.

The perturbative QM/MC method provides an excellent means for reducing the time taken for QM/MC simulations to the point where they are competitive with QM/MD. The combination of energy- and distance-based methods reduces the error introduced to a negligible level. This enables the QM/MC and QM/MD methods to be compared for the Menshutkin reaction; the Monte Carlo simulation gives an activation free energy in better agreement with previous studies and the experimental value, but it is possible that this is merely fortuitous, as it is probably caused by the neglect of the rotational and vibrational degrees of freedom of the amine group. This again suggests that any further increases in accuracy would require the replacement of the AM1 method with a high basis set *ab initio* QM technique.

Energy correction

The restriction of the number of degrees of freedom of the QM region in a high-symmetry Monte Carlo approach simplifies the process of parameterizing a force field specifically for the system. This is essentially the result of applying the approximate energy correction term developed in Chapter 6 for the Menshutkin reaction. The effectiveness of this method depends on two assumptions — that the low level QM treatment reproduces the high level electron density well enough so that the interaction

energy with solvent is similar at both levels, and that the predicted geometry of the QM region is similar at both levels, so that some degrees of freedom can be neglected. The method performs very well for determining the minimum-energy path for the gas phase Menshutkin reaction, and for single point energies of the solution-phase systems. Although it is slightly less accurate than the ONIOM method, it is much less demanding in terms of computer time. The ONIOM method requires an additional two QM calculations for every energy calculation, and this overhead will only increase as higher QM treatments are used. The method developed in this work, on the other hand, adds essentially no overhead to the QM energy calculation, irrespective of the QM treatment used, although multiple single point energies are required prior to the simulation to initialise the correction term. However, for systems with a reasonable number of relevant degrees of freedom, such that a relatively small number of energy points are required, or if the correction is then required for a large number of simulation points, the method developed in this thesis is far more efficient than ONIOM.

The FlucQ module for CHARMM developed in this work, along with a large number of minor QM/MM improvements, is now part of the standard CHARMM distribution. This, together with developments by other researchers, should enable the QM/MM method to be more widely applied in a flexible fashion. The method is unlikely to ever become a 'black box' approach, requiring as it does some knowledge about the system's electronic structure and dynamics, but the gradual solution of many methodological problems is likely to result in QM/MM becoming a standard technique in the future computational chemist's arsenal.

7.2 Future research directions

The methods developed in this work have a very broad application to hybrid QM/MM studies. The FlucQ method can potentially be applied to any solute-solvent simulation where the solvent (not necessarily water) is treated molecular mechanically; the

implementation is sufficiently general that using the technique for other solvents, or indeed for small molecules, is simply a matter of relevant parameterization, perhaps by use of a parameterization method similar to that used in section 4.6. Hybrid QM/MC techniques can also be used for most of the systems for which QM/MD is applicable, given a suitable choice of move set, while the perturbative QM/MC method as developed in this work should be usable for many solute-solvent systems, with suitable choice of the cutoff parameters.

Perturbative QM/MC

The perturbative QM/MC technique provides a useful intermediate method, far less expensive than conventional QM/MC but potentially far more accurate and flexible than simulations with an MM force field. However, for many simulations adequate results have been obtained simply by using suitable ESP-fitted charges in the region of interest [96, 97], and for such systems the overhead of evaluating many 1-electron integrals is prohibitive. Significant cost savings could thus be achieved by approximating the perturbative energy delta, equation 6.9, by the interaction energy with a set of ESP-fitted charges in the QM region. This would enable the accurate treatment of such systems in a largely automatic manner. (Incidentally, for species such as metal ions where single point charges do not well reproduce the electrostatic potential, some researchers have improved the treatment by adding extra charge sites [304]. It is possible that this procedure could also be automated.) Methods also exist to dynamically optimize the acceptance rate of conventional MC moves [287] and it may be possible to use a similar method to alter the distance and energy cutoffs during a perturbative QM/MC simulation, perhaps by using a formula similar to that developed by Evans and Truong [292]. Such improvements would enable the perturbative QM/MC method to be applied to a wider range of systems without

requiring extensive adjustment of the cutoff parameters, and further cut the computer time required.

The FlucQ model

The FlucQ model does not perform perfectly, and although the TIP4P-FQ1 parameterization appears to give results in better agreement with experiment for several bimolecular systems, further study in this area is pointed to. It may be possible to improve the fit with QM data by adding extra energy terms or by altering the existing parameters, although recent work [240, 241, 305] would suggest that the addition of polarizable dipoles is necessary for a more correct treatment of the electron density.

QM/MM simulation of enzymatic systems requires the introduction of link atoms wherever the QM/MM boundary cuts bonds. The approximate treatment of electron density by the FlucQ method suggests that these link atoms could actually be placed on bonds where significant electron density changes occur, as FlucQ may be able to alter the MM charges to adequately approximate this, in consequence allowing smaller QM regions to be used in such treatments. With a sufficiently accurate FlucQ model, it may even be possible to simulate charge transfer across these bonds, between the QM and MM regions, in an approximate manner.

Free energy and QM/MM simulation

An obvious next step on the application side would be to simulate the Menshutkin solution-phase reaction with *ab initio* methods, such as HF/3-21G*, using the method developed in section 6.9 to increase the accuracy of the QM representation. This should be able to predict activation and reaction free energies in much better agreement with other studies and experiment. Additionally, the improved description of the QM electron density should be able to more conclusively demonstrate the usefulness or otherwise of the FlucQ method. Further MC simulations would benefit from

optimization of the FlucQ moves, perhaps using a technique similar to ANES-MC [200], to reduce the overhead to levels comparable with FQ/MD simulations. For MD studies of enzymatic systems, the stochastic boundary method is popular, and thus a FlucQ analogue of the Langevin random and frictional forces in charge space would be required.

A similar energy correction technique was used recently [172] in a study of the serine protease catalysed deacylation reaction. In that study, however, the *free* energy was corrected, after QM/MM umbrella sampling was completed. (The solution-phase QM/MM PMF, using B3LYP/3-21G for the QM region, was corrected to B3LYP/6-31G* level using the gas-phase free energy at both levels.) The disadvantage of such an approach, however, is that it assumes the geometries and the minimum energy solution-phase path to be the same at both levels of theory, such that the low level simulation reproduces all aspects of the high level treatment except for the energy. The correction technique used in this work, however, and the ONIOM method, both adjust the energy *during* the simulation, such that the path of the reaction can be altered. A detailed comparison of the three methods would thus be instructive, and enormously useful in guiding future QM/MC studies. Such studies using QM/MM techniques are likely to become increasingly common, assuming 'Moore's Law' holds, as the computer power available to the computational chemist continues to increase.

Bibliography

[1] G. Schaftenaar and J. H. Noordik, Molden: a pre- and post-processing program for molecular and electronic structures. J. Comp.-Aided Mol. Des. **14**, 123 (2000).

[2] W. Humphrey, A. Dalke, and K. Schulten, VMD — Visual Molecular Dynamics. J. Mol. Graph. **14**, 33 (1996).

[3] OSWALD, the Oxford Supercomputing Wellcome ALternative Distributed memory system, consisting of 128 dual-processor 1.26GHz PentiumIII machines, connected by Myrinet interlinks, *http://www.osc.ox.ac.uk/*.

[4] G. E. Moore, Cramming more components onto integrated circuits. Electronics **38**, 114 (1965).

[5] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry; Introduction to Advanced Electronic Structure Theory, Revised First Edition* (McGraw-Hill, New York, 1989).

[6] D. R. Hartree, The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods. Proc. Camb. Phil. Soc. **24**, 89 (1928).

[7] V. Fock, Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. Z. Phys. **61**, 126 (1930).

[8] C. Møller and M. S. Plesset, Note on an Approximate Treatment for Many-Electron Systems. Phys. Rev. **46**, 618 (1934).

[9] R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules* (Oxford University Press, Oxford, 1989).

[10] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. J. Comp. Chem. **4**, 187 (1983).

[11] J. Chandrasekhar, D. C. Spellmeyer, and W. L. Jorgensen, Energy Component Analysis for Dilute Aqueous Solutions of Li^+, Na^+, F^- and Cl^- Ions. J. Am. Chem. Soc. **106**, 903 (1984).

[12] S. Goedecker, Linear scaling electronic structure methods. Rev. Mod. Phys. **71**, 1085 (1999).

[13] A. van der Vaart, V. Gogonea, S. L. Dixon, and K. M. Merz, Jr., Linear scaling molecular orbital calculations of biological systems using the semiempirical divide and conquer method. J. Comp. Chem. **21**, 1494 (2000).

[14] F. S. Lee, Z.-T. Chu, M. B. Bolger, and A. Warshel, Calculations of antibody-antigen interactions: microscopic and semi-microscopic evaluation of the free energies of binding of phosphorylcholine analogs to McPC603. Prot. Eng. **5**, 215 (1992).

[15] J. Åqvist, C. Medina, and J.-E. Samuelsson, A new method for predicting binding affinity in computer-aided drug design. Prot. Eng. **7**, 385 (1994).

[16] A. Warshel and M. Levitt, Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. J. Mol. Biol. **103**, 227 (1976).

[17] M. J. Field, P. A. Bash, and M. Karplus, A Combined Quantum Mechanical and Molecular Mechanical Potential for Molecular Dynamics Simulations. J. Comp. Chem. **11**, 700 (1990).

[18] Q. Cui and M. Karplus, Molecular Properties from Combined QM/MM Methods. 2. Chemical Shifts in Large Molecules. J. Phys. Chem. B **104**, 3721 (2000).

[19] V. Gogonea, The QM/MM Method. An Overview. Internet Electron. J. Mol. Des. **1**, 173 (2002).

[20] A. R. Dinner, G. M. Blackburn, and M. Karplus, Uracil-DNA glycosylase acts by substrate autocatalysis. Nature **413**, 752 (2001).

[21] M. J. Loferer, B. M. Webb, G. H. Grant, and K. R. Liedl, Energetic and Stereochemical Effects of the Protein Environment on Substrate: A Theoretical Study of Methylmalonyl-CoA Mutase. J. Am. Chem. Soc. **125**, 1072 (2003).

[22] N. Reuter, A. Dejaegere, B. Maigret, and M. Karplus, Frontier Bonds in QM/MM Methods: A Comparison of Different Approaches. J. Phys. Chem. A **104**, 1720 (2000).

[23] R. B. Murphy, D. M. Philipp, and R. A. Friesner, Frozen orbital QM/MM methods for density functional theory. Chem. Phys. Lett. **321**, 113 (2000).

[24] P. D. Lyne, M. Hodoscek, and M. Karplus, A Hybrid QM-MM Potential Employing Hartree-Fock or Density Functional Methods in the Quantum Region. J. Phys. Chem. A **103**, 3462 (1999).

[25] F. H. Stillinger and C. W. David, Polarization model for water and its ionic dissociation products. J. Chem. Phys. **69**, 1473 (1978).

[26] M. Sprik and M. L. Klein, A polarizable model for water using distributed charge sites. J. Chem. Phys. **89**, 7556 (1988).

[27] P. Ahlström, A. Wallqvist, S. Engström, and B. Jönsson, A molecular dynamics study of polarizable water. Mol. Phys. **68**, 563 (1989).

[28] A. Wallqvist and B. J. Berne, Effective potentials for liquid water using polarizable and nonpolarizable models. J. Phys. Chem. **97**, 13841 (1993).

[29] L. X. Dang, The nonadditive intermolecular potential for water revised. J. Chem. Phys. **97**, 2659 (1992).

[30] S. W. Rick, S. J. Stuart, and B. J. Berne, Dynamical fluctuating charge force fields: Application to liquid water. J. Chem. Phys. **101**, 6141 (1994).

[31] R. T. Sanderson, An Interpretation of Bond Lengths and a Classification of Bonds. Science **114**, 670 (1951).

[32] R. T. Sanderson, *Chemical bonds and bond energy, 2nd ed.* (Academic Press, New York, 1976).

[33] S. W. Rick and B. J. Berne, Dynamical Fluctuating Charge Force Fields: The Aqueous Solvation of Amides. J. Am. Chem. Soc. **118**, 672 (1996).

[34] J. S. Bader, C. M. Cortis, and B. J. Berne, Solvation and reorganization energies in polarizable molecular and continuum solvents. J. Chem. Phys. **106**, 2372 (1997).

[35] S. W. Rick and B. J. Berne, Free Energy of the Hydrophobic Interaction from Molecular Dynamics Simulations: The Effects of Solute and Solvent Polarizability. J. Phys. Chem. B **101**, 10488 (1997).

[36] J. C. Corchado and D. G. Truhlar, in *Combined Quantum Mechanical and Molecular Mechanical Methods, ACS Symposium Series Vol. 712*, edited by J. Gao and M. A. Thompson (American Chemical Society, Washington, 1998), pp. 106–127.

[37] M. Born and R. Oppenheimer, Zur Quantentheorie der Molekeln. Ann. Phys. **84**, 457 (1927).

[38] J. C. Slater, The Theory of Complex Spectra. Phys. Rev. **34**, 1293 (1929).

[39] J. C. Slater, Note on Hartree's Method. Phys. Rev. **35**, 210 (1930).

[40] W. Pauli, Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren. Z. Phys. **31**, 765 (1925).

[41] C. M. Reeves and R. Fletcher, Use of Gaussian Functions in the Calculation of Wavefunctions for Small Molecules. III. The Orbital Basis and Its Effect on Valence. J. Chem. Phys. **42**, 4073 (1965).

[42] W. J. Hehre, R. F. Stewart, and J. A. Pople, Self-Consistent Molecular-Orbital Methods. I. Use of Gaussian Expansions of Slater-Type Atomic Orbitals. J. Chem. Phys. **51**, 2657 (1969).

[43] J. S. Binkley, J. A. Pople, and W. J. Hehre, Self-Consistent Molecular Orbital Methods. 21. Small Split-Valence Basis Sets for First-Row Elements. J. Am. Chem. Soc. **102**, 939 (1980).

[44] P. C. Hariharan and J. A. Pople, The Influence of Polarization Functions on Molecular Orbital Hydrogenation Energies. Theor. Chim. Acta. **28**, 213 (1973).

[45] J. B. Collins, P. von R. Schleyer, J. S. Binkley, and J. A. Pople, Self-consistent molecular orbital methods. XVII. Geometries and binding energies of second-row molecules. A comparison of three basis sets. J. Chem. Phys. **64**, 5142 (1976).

[46] R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople, Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. J. Chem. Phys. **72**, 650 (1980).

[47] H. Kollmar, On the Stability of Alkyl Anions. A Molecular Orbital Theoretical Study. J. Am. Chem. Soc. **100**, 2665 (1978).

[48] T. Clark, J. Chandrasekhar, G. W. Spitznagel, and P. von R. Schleyer, Efficient Diffuse Function-Augmented Basis Sets for Anion Calculations. III. The 3-21+G Basis Set for First-Row Elements, Li–F. J. Comp. Chem. **4**, 294 (1983).

[49] C. C. J. Roothaan, New Developments in Molecular Orbital Theory. Rev. Mod. Phys. **23**, 69 (1951).

[50] G. G. Hall, The Molecular Orbital Theory of Chemical Valency VIII. A Method for Calculating Ionisation Potentials. Proc. Roy. Soc. London **A205**, 541 (1951).

[51] P. Pulay, Convergence Acceleration of Iterative Sequences. The Case of SCF Iteration. Chem. Phys. Lett. **73**, 393 (1980).

[52] P.-O. Löwdin, Correlation Problem in Many-Electron Quantum Mechanics. Adv. Chem. Phys. **2**, 207 (1959).

[53] J. A. Pople, D. P. Santry, and G. A. Segal, Approximate Self-Consistent Molecular Orbital Theory. I. Invariant Procedures. J. Chem. Phys. **43**, S129 (1965).

[54] J. A. Pople and G. A. Segal, Approximate Self-Consistent Molecular Orbital Theory. II. Calculations with Complete Neglect of Differential Overlap. J. Chem. Phys. **43**, S136 (1965).

[55] M. J. S. Dewar and W. Thiel, Ground States of Molecules. 38. The MNDO Method. Approximations and Parameters. J. Am. Chem. Soc. **99**, 4899 (1977).

[56] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart, AM1: A New General Purpose Quantum Mechanical Molecular Model. J. Am. Chem. Soc. **107**, 3902 (1985).

[57] J. J. P. Stewart, Optimization of Parameters for Semiempirical Methods I. Method. J. Comp. Chem. **10**, 209 (1989).

[58] J. J. P. Stewart, Optimization of Parameters for Semiempirical Methods II. Applications. J. Comp. Chem. **10**, 221 (1989).

[59] A. A. Voityuk and N. Rösch, AM1/d Parameters for Molybdenum. J. Phys. Chem. A **104**, 4089 (2000).

[60] W. Thiel and A. A. Voityuk, Extension of the MNDO Formalism to d orbitals: Integral Approximations and Preliminary Numerical Results. Theor. Chim. Acta. **81**, 391 (1992).

[61] W. Thiel and A. A. Voityuk, Extension of MNDO to d orbitals: parameters and results for silicon. J. Mol. Struct. (Theochem) **313**, 141 (1994).

[62] W. Thiel and A. A. Voityuk, Extension of MNDO to d Orbitals: Parameters and Results for the Second-Row Elements and for the Zinc Group. J. Phys. Chem. **100**, 616 (1996).

[63] L. H. Thomas, The calculation of atomic fields. Proc. Camb. Phil. Soc. **23**, 542 (1927).

[64] P. Hohenberg and W. Kohn, Inhomogeneous Electron Gas. Phys. Rev. **136**, 864 (1964).

[65] W. Kohn and L. J. Sham, Self-Consistent Equations Including Exchange and Correlation Effects. Phys. Rev. **140**, 1133 (1965).

[66] J. C. Slater, A Simplification of the Hartree-Fock Method. Phys. Rev. **81**, 385 (1951).

[67] S. H. Vosko, L. Wilk, and M. Nusair, Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. Can. J. Phys. **58**, 1200 (1980).

[68] A. J. Cohen and N. C. Handy, Assessment of exchange correlation functionals. Chem. Phys. Lett. **316**, 160 (2000).

[69] A. D. Becke, Density-functional exchange-energy approximation with correct asymptotic behavior. Phys. Rev. A **38**, 3098 (1988).

[70] C. Lee, W. Yang, and R. G. Parr, Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. Phys. Rev. B **37**, 785 (1988).

[71] A. D. Becke, Density-Functional themochemistry. III. The role of exact exchange. J. Chem. Phys. **98**, 5648 (1993).

[72] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. Phys. Rev. B **46**, 6671 (1992).

[73] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, *Ab Initio* Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. J. Phys. Chem. **98**, 11623 (1994).

[74] A. D. Becke, A new mixing of Hartee-Fock and local density-functional theories. J. Chem. Phys. **98**, 1372 (1993).

[75] A. D. Becke, A new inhomogeneity parameter in density-functional theory. J. Chem. Phys. **109**, 2092 (1998).

[76] M. J. Field, *A practical introduction to the simulation of molecular systems* (Cambridge University Press, Cambridge, UK, 1999).

[77] J. E. Jones, On the determination of molecular fields — II. From the equation of state of a gas. Proc. Roy. Soc. London **A106**, 463 (1924).

[78] J. Gao, in *Reviews in Computational Chemistry*, edited by K. B. Lipkowitz and D. B. Boyd (VCH, New York, 1995), Vol. 7, pp. 119–185.

[79] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, General Atomic and Molecular Electronic Structure System. J. Comp. Chem. **14**, 1347 (1993).

[80] M. F. Guest, J. H. van Lenthe, J. Kendrick, P. Sherwood, R. D. Amos, R. J. Buenker, H. van Dam, M. Dupuis, N. C. Handy, I. H. Hillier, P. J. Knowles, V. Bonacic-Koutecky, W. von Niessen, R. J. Harrison, A. P. Rendell, V. R. Saunders, K. Schoffel, A. J. Stone, and D. Tozer, GAMESS-UK, a package of ab initio programs.

[81] R. D. Amos, I. L. Albert, J. S. Andrews, S. M. Colwell, N. C. Handy, D. Jayatilaka, P. J. Knowles, R. Kobayashi, K. E. Laidig, G. Laming, A. M. Lee, P. E. Maslen, C. W. Murray, J. E. Rice, E. D. Simandiras, A. J. Stone, M. D. Su, and D. J. Tozer, CADPAC 6.0, 1995.

[82] J. J. P. Stewart, MOPAC, A Semi-Empirical Molecular Orbital Program. QCPE **455**, (1983).

[83] J. Gao, P. Amara, C. Alhambra, and M. J. Field, A Generalized Hybrid Orbital (GHO) Method for the Treatment of Boundary Atoms in Combined QM/MM Calculations. J. Phys. Chem. A **102**, 4714 (1998).

[84] C. G. Gray and K. E. Gubbins, *Theory of Molecular Fluids. 1. Fundamentals* (Clarendon, Oxford, UK, 1984).

[85] P. J. Steinbach and B. R. Brooks, New Spherical-Cutoff Methods for Long-Range Forces in Macromolecular Simulation. J. Comp. Chem. **15**, 667 (1994).

[86] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, Oxford, 1987).

[87] P. Ewald, Due Berechnung optischer und elektrostatischer Gitterpotentiale. Ann. Phys. **64**, 253 (1921).

[88] D. Frenkel and B. Smit, *Understanding Molecular Simulation, From Algorithms to Applications, 2nd ed.* (Academic Press, San Diego, California, 2002).

[89] T. A. Darden, D. York, and L. Pedersen, Particle mesh Ewald: An $N\log(N)$ method for Ewald sums in large systems. J. Chem. Phys. **98**, 10089 (1993).

[90] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, A smooth particle mesh Ewald method. J. Chem. Phys. **103**, 8577 (1995).

[91] L. Onsager, Electric Moments of Molecules in Liquids. J. Am. Chem. Soc. **58**, 1486 (1936).

[92] I. G. Tironi, R. Sperb, P. E. Smith, and W. F. van Gunsteren, A generalized reaction field method for molecular dynamics simulations. J. Chem. Phys. **102**, 5451 (1995).

[93] D. van der Spoel, P. J. van Maaren, and H. J. C. Berendsen, A systematic study of water models for molecular simulation: Derivation of water models optimized for use with a reaction field. J. Chem. Phys. **108**, 10220 (1998).

[94] C. L. Brooks III and M. Karplus, Deformable stochastic boundaries in molecular dynamics. J. Chem. Phys. **79**, 6312 (1983).

[95] C. L. Brooks III, A. Brünger, and M. Karplus, Active Site Dynamics in Protein Molecules: A Stochastic Boundary Molecular-Dynamics Approach. Biopolymers **24**, 843 (1985).

[96] K. M. Merz, Jr. and P. A. Kollman, Free Energy Perturbation Simulations of the Inhibition of Thermolysin: Prediction of the Free Energy of Binding of a New Inhibitor. J. Am. Chem. Soc. **111**, 5649 (1989).

[97] K. M. Merz, Jr., CO_2 Binding to Human Carbonic Anhydrase II. J. Am. Chem. Soc. **113**, 406 (1991).

[98] R. S. Mulliken, Electronic Population Analysis on LCAO-MO Molecular Wave Functions. J. Chem. Phys. **23**, 1833 (1955).

[99] L. E. Chirlian and M. M. Francl, Atomic Charges Derived from Electrostatic Potentials: A Detailed Study. J. Comp. Chem. **8**, 894 (1987).

[100] 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. J. Comp. Chem. **11**, 361 (1990).

[101] B. H. Besler, K. M. Merz, Jr., and P. A. Kollman, Atomic Charges Derived from Semiempirical Methods. J. Comp. Chem. **11**, 431 (1990).

[102] M. L. Connolly, Analytical Molecular Surface Calculation. J. Appl. Cryst. **16**, 548 (1983).

[103] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, Optimization by Simulated Annealing. Science **220**, 671 (1983).

[104] E. Aarts and J. Korst, *Simulated Annealing and Boltzmann Machines. A Stochastic Approach to Combinatorial Optimization and Neural Computing* (Wiley, New York, 1990).

[105] I. Andricioaei and J. E. Straub, Finding the needle in the haystack: algorithms for conformational optimization. Comp. Phys. **10**, 449 (1996).

[106] J. A. Nelder and R. Mead, A simplex method for function minimization. Computer J. **7**, 308 (1965).

[107] A. R. Leach, *Molecular Modelling. Principles and Applications* (Addison Wesley Longman, Harlow, Essex, UK, 1996).

[108] R. Fletcher and C. M. Reeves, Function minimization by conjugate gradients. Computer J. **7**, 149 (1964).

[109] E. Polak, *Computational methods in optimization: a unified approach* (Academic Press, London, 1971).

[110] M. J. D. Powell, Restart procedures for the conjugate gradient method. Mathematical Programming **12**, 241 (1977).

[111] M. P. Allen and D. J. Tildesley, in *Computer Simulation of Liquids* (Oxford University Press, Oxford, 1987), p. 78.

[112] W. C. Gear, *Numerical initial value problems in ordinary differential equations* (Prentice-Hall, Englewood Cliffs, NJ, 1971).

[113] R. Car and M. Parrinello, Unified Approach for Molecular Dynamics and Density-Functional Theory. Phys. Rev. Lett. **55**, 2471 (1985).

[114] J. P. Ryckaert, G. Cicotti, and H. J. C. Berendsen, Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of *n*-Alkanes. J. Comp. Phys. **23**, 327 (1977).

[115] A. Münster, *Statistical thermodynamics* (Springer-Verlag, Berlin, 1969).

[116] J. Q. Broughton, G. H. Gilmer, and J. D. Weeks, Constant pressure molecular dynamics simulations of the $2D$ r^{-12} system: Comparison with isochromes and isotherms. J. Chem. Phys. **75**, 5128 (1981).

[117] 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. J. Chem. Phys. **81**, 3684 (1984).

[118] S. Nosé, A molecular dynamics method for simulations in the canonical ensemble. Mol. Phys. **52**, 255 (1984).

[119] W. G. Hoover, Canonical dynamics: Equilibrium phase-space distributions. Phys. Rev. A **31**, 1695 (1985).

[120] S. Nosé, An extension of the canonical ensemble molecular dynamics method. Mol. Phys. **57**, 187 (1986).

[121] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equation of State Calculations by Fast Computing Machines. J. Chem. Phys. **21**, 1087 (1953).

[122] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, in *Intermolecular Forces* (Reidel, Dordrecht, Holland, 1981), p. 331.

[123] A. K. Soper and M. G. Phillips, A new determination of the structure of water at 25 $^\circ$C. Chem. Phys. **107**, 47 (1986).

[124] 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. **79**, 926 (1983).

[125] M. Born, Volumen und Hydratationswärme der Ionen. Z. Phys. **1**, 45 (1920).

[126] S. Miertus, E. Scrocco, and J. Tomasi, Electrostatic Interaction of a Solute with a Continuum — A Direct Utilization of *Ab Initio* Molecular Potentials for the Provision of Solvent Effects. Chem. Phys. **55**, 117 (1981).

[127] P. Kollman, Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. Chem. Rev. **93**, 2395 (1993).

[128] R. W. Zwanzig, High-temperature equation of state by a perturbation method. I. Nonpolar Gases. J. Chem. Phys. **22**, 1426 (1954).

[129] G. M. Torrie and J. P. Valleau, Monte Carlo Energy Estimates Using Non-Boltzmann Sampling: Application to the Sub-Critical Lennard-Jones Fluid. Chem. Phys. Lett. **28**, 578 (1974).

[130] G. M. Torrie and J. P. Valleau, Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. J. Comp. Phys. **23**, 187 (1977).

[131] C. L. Brooks III, M. Karplus, and B. M. Pettitt, *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics* (Wiley, New York, 1988).

[132] A. M. Ferrenberg and R. H. Swendsen, Optimized Monte Carlo Data Analysis. Phys. Rev. Lett. **63**, 1195 (1989).

[133] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg, The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. J. Comp. Chem. **13**, 1011 (1992).

[134] K. Burrage, J. Belward, F. Duncalfe, L. Lau, M. Rezny, and R. Young, in *Proceedings of the Sixth SIAM Conference on Parallel Processing for Scientific Computing* (SIAM Publications, Norfolk, Virginia, 1993), pp. 117–125.

[135] D. P. Anderson, J. Cobb, E. Korpela, M. Lebofsky, and D. Werthimer, SETI@home: an experiment in public-resource computing. CACM **45**, 56 (2002).

[136] W. G. Richards, Virtual Screening Using Grid Computing: The Screensaver Project. Nature Rev. Drug. Disc. **1**, 551 (2002).

[137] M. Glick, G. H. Grant, and W. G. Richards, Docking of Flexible Molecules Using Multiscale Ligand Representations. J. Med. Chem. **45**, 4639 (2002).

[138] M. Shirts and V. S. Pande, Screen Savers of the World Unite!. Science **290**, 1903 (2001).

[139] M. R. Shirts and V. S. Pande, Mathematical Analysis of Coupled Parallel Simulations. Phys. Rev. Lett. **86**, 4983 (2001).

[140] B. Zagrovic, E. J. Sorin, and V. Pande, β-Hairpin Folding Simulations in Atomistic Detail Using an Implicit Solvent Model. J. Mol. Biol. **313**, 151 (2001).

[141] F. Allen, G. Almasi, W. Andreoni, D. Beece, B. J. Berne, A. Bright, J. Brunheroto, C. Cascaval, J. Castanos, P. Coteus, P. Crumley, A. Curioni, M. Denneau, W. Donath, M. Eleftheriou, B. Fitch, B. Fleischer, C. J. Georgiou, R. Germain, M. Giampapa, D. Gresh, M. Gupta, R. Haring, H. Ho, P. Hochschild, S. Hummel, T. Jonas, D. Lieber, G. Martyna, K. Maturu, J. Moreira, D. Newns, M. Newton, R. Philhower, T. Picunko, J. Pitera, M. Pitman, R. Rand, A. Royyuru, V. Salapura, A. Sanomiya, R. Shah, Y. Sham, S. Singh, M. Snir, F. Suits, R. Swetz, W. C. Swope, N. Vishnumurthy, T. J. C. Ward, H. Warren, and R. Zhou, Blue Gene: A vision for protein science using a petaflop supercomputer. IBM Systems Journal **40**, 310 (2001).

[142] Y. Komeiji, M. Uebayasi, R. Takata, A. Shimizu, K. Itsukashi, and M. Taiji, Fast and accurate molecular dynamics simulation of a protein using a special-purpose computer. J. Comp. Chem. **18**, 1546 (1997).

[143] G. M. Amdahl, Validity of single-processor approach to achieving large-scale computing capability. Proceedings of AFIPS Conference, Reston, VA. 483 (1967).

[144] J. L. Gustafson, Reevaluating Amdahl's Law. CACM **31**, 532 (1988).

[145] R. Metcalfe and D. Boggs, Ethernet: Distributed Packet Switching for Local Computer Networks. CACM **19**, 395 (1976).

[146] N. Carriero and D. Gelernter, Linda in context. CACM **32**, 444 (1989).

[147] N. Carriero and D. Gelernter, *How to Write Parallel Programs* (MIT Press, Cambridge, MA, 1990).

[148] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, J. L. Andres, M. Head-Gordon, E. S. Replogle, and J. A. Pople, *Gaussian 98*, Gaussian, Inc., Pittsburgh PA, 1998.

[149] A. A. Granovsky, PC-GAMESS, *http://classic.chem.msu.su/gran/gamess/*.

[150] L. Dagum and R. Menon, OpenMP: An Industry-Standard API for Shared-Memory Programming. IEEE Comp. Sci. Eng. **5**, 46 (1998).

[151] Message Passing Interface Forum, in *Proceedings of Supercomputing '93* (IEEE Computer Society Press, Los Alamitos, USA, 1993), pp. 878–883.

[152] T. Sterling, D. J. Becker, D. Savarese, J. E. Dorband, U. A. Ranawake, and C. V. Packer, in *Proceedings of the International Conference on Parallel Processing* (CRC Press, Boca Raton, FL, 1995), pp. 11–14.

[153] J. J. Dongarra, Performance of various computers using standard linear equations software. Technical Report CS-89-85, University of Tennessee, *http://www.netlib.org/benchmark/performance.ps* (1989).

[154] The Linux 'operating system' consists of the Linux kernel (a Minix derivative written by Linus Torvalds in 1991) and software tools provided by the GNU Project.

[155] RedHat, Inc., RedHat Linux distribution, *http://www.redhat.com/*.

[156] LoBoS, 'Lots of Boxes on Shelves', *http://www.lobos.nih.gov/*.

[157] Google, *http://www.google.com/*.

[158] I. Foster and C. Kesselman, *The Grid: Blueprint For A New Computing Infrastructure* (Morgan Kaufmann, San Francisco, 1999).

[159] TeraGrid — a multi-year effort to build and deploy the world's largest, fastest, distributed infrastructure for open scientific research. *http://www.teragrid.org/*.

[160] Intel Corporation, Preboot Execution Environment (PXE) Version 2.1, *http://www.intel.com/labs/manage/wfm/wfmspecs.htm*.

[161] Compaq Computer Corp., Intel Corp. and Microsoft Corporation, VIA (Virtual Interface Architecture) Specification, Version 1.0, *http://developer.intel.com/design/servers/vi/the_spec/specification.htm*, 1997.

[162] Myricom, Inc., Myrinet scalable cluster interconnect. *http://www.myri.com/*.

[163] D. Turner, Scalable Computing Laboratory, Ames Laboratory, Iowa State University, Iowa, MP_Lite — a lightweight message-passing library. *http://www.scl.ameslab.gov/Projects/MP_Lite/MP_Lite.html*.

[164] Veridian Information Solutions, Inc., OpenPBS (Portable Batch System) v2.3, *http://www.openpbs.org/*.

[165] I. Foster and C. Kesselman, Globus: A Metacomputing Infrastructure Toolkit. Intl. J. Supercomputer Applications **11**, 115 (1997).

[166] W. Gropp, E. Lusk, N. Doss, and A. Skjellum, A high-performance, portable implementation of the MPI message passing interface standard. Parallel Computing **22**, 789 (1996).

[167] W. D. Gropp and E. Lusk, *User's Guide for* `mpich`*, a Portable Implementation of MPI*, Mathematics and Computer Science Division, Argonne National Laboratory, 1996.

[168] G. Burns, R. Daoud, and J. Vaigl, in *Proceedings of Supercomputing Symposium '94*, edited by J. W. Ross (University of Toronto, Toronto, Canada, 1994), pp. 379–386.

[169] P. Wyckoff, Ohio Supercomputer Center, Columbus, Ohio, Mpiexec — a parallel job launcher for PBS, *http://www.osc.edu/∼pw/mpiexec/*.

[170] Q. O. Snell, A. R. Mikler, and J. L. Gustafson, NetPIPE, A Network Protocol Independent Performance Evaluator. *http://www.scl.ameslab.gov/netpipe/*.

[171] R. C. Wilmouth, I. J. Clifton, C. V. Robinson, P. L. Roach, R. T. Aplin, N. J. Westwood, J. Hajdu, and C. J. Schofield, Structure of a specific acyl-enzyme complex formed between beta-casomorphin-7 and porcine pancreatic elastase. Nat. Struct. Biol. **4**, 456 (1997).

[172] M. Topf, *Computer Simulation of the Deacylation Step of the Serine Protease Reaction Mechanism*, Ph.D. thesis, Oxford University, Oxford, UK, 2002.

[173] M. Topf, P. Várnai, and W. G. Richards, Ab Initio QM/MM Dynamics Simulation of the Tetrahedral Intermediate of Serine Proteases: Insights into the Active Site Hydrogen-Bonding Network. J. Am. Chem. Soc. **124**, 14780 (2002).

[174] S. A. Adcock, *Computer Simulation of Membrane Bound Molecules*, Ph.D. thesis, Oxford University, Oxford, UK, 2001.

[175] O. A. Walsh, *Computer Simulation of Phosphate Ester Hydrolysis Reactions*, Ph.D. thesis, Oxford University, Oxford, UK, 2001.

[176] M. Svensson, S. Humbel, R. Froese, T. Matsubara, S. Sieber, and K. Morokuma, ONIOM: A multilayered integrated MO+MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder reactions and $Pt(P(t-Bu)_3)_2 + H_2$ oxidative addition. J. Phys. Chem. **100**, 19357 (1996).

[177] S. Dapprich, I. Komáromi, K. S. Byun, K. Morokuma, and M. J. Frisch, A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives. J. Mol. Struct. (Theochem) **461–462**, 1 (1999).

[178] V. Gogonea, L. M. Westerhoff, and K. M. Merz, Jr., Quantum mechanical/quantum mechanical methods. I. A divide and conquer strategy for solving the Schrödinger equation for large molecular systems using a composite density functional-semiempirical Hamiltonian. J. Chem. Phys. **113**, 5604 (2000).

[179] G. Hong, M. Strajbl, T. A. Wesolowski, and A. Warshel, Constraining the electron densities in DFT method as an effective way for *ab initio* studies of metal-catalyzed reactions. J. Comp. Chem. **21**, 1554 (2000).

[180] S. A. Clough, Y. Beers, G. P. Klein, and L. S. Rothman, Dipole moment of water from Stark measurements of H_2O, HDO and D_2O. J. Chem. Phys. **59**, 2254 (1973).

[181] C. A. Coulson and D. Eisenberg, Interactions of H_2O molecules in ice. I. The dipole moment of an H_2O molecule in ice. Proc. Roy. Soc. London **A291**, 445 (1966).

[182] E. Whalley, A relation between the strengths of the orientation polarization and the infrared absorption of the O-H stretching vibrations of ice. Chem. Phys. Lett. **53**, 449 (1978).

[183] S. L. Carnie and G. N. Patey, Fluids of polarizable hard spheres with dipoles and tetrahedral quadrupoles. Integral equation results with application to liquid water. Mol. Phys. **47**, 1129 (1982).

[184] P. G. Kusalik and G. N. Patey, On the molecular theory of aqueous electrolyte solutions. I. The solution of the RHNC approximation for models at finite concentration. J. Chem. Phys. **88**, 7715 (1988).

[185] K. Watanabe and M. L. Klein, Effective Pair Potentials and the Properties of Water. Chem. Phys. **131**, 157 (1989).

[186] M. Sprik, Computer Simulation of the Dynamics of Induced Polarization Fluctuations in Water. J. Phys. Chem. **95**, 2283 (1991).

[187] K. Laasonen, M. Sprik, M. Parrinello, and R. Car, *"Ab initio"* liquid water. J. Chem. Phys. **99**, 9080 (1993).

[188] D. N. Bernardo, Y. Ding, K. Krogh-Jespersen, and R. M. Levy, An Anisotropic Polarizable Water Model: Incorporation of All-Atom Polarizabilities into Molecular Mechanics Force Fields. J. Phys. Chem. **98**, 4180 (1994).

[189] R. S. Mulliken, A New Electroaffinity Scale; Together with Data on Valence States and on Valence Ionization Potentials and Electron Affinities. J. Chem. Phys. **2**, 782 (1934).

[190] R. S. Mulliken, Electronic Structures of Molecules. XI. Electroaffinity, Molecular Orbitals and Dipole Moments. J. Chem. Phys. **3**, 573 (1935).

[191] W. J. Mortier, K. V. Genechten, and J. Gasteiger, Electronegativity Equalization: Application and Parametrization. J. Am. Chem. Soc. **107**, 829 (1985).

[192] J. Gasteiger and M. Marsili, Iterative Partial Equalization Of Orbital Electronegativity — A Rapid Access To Atomic Charges. Tetrahedron **36**, 3219 (1980).

[193] J. Gasteiger and H. Saller, Calculation of the Charge Distribution in Conjugated Systems by a Quantification of the Resonance Concept. Angew. Chem. Int. Ed. Engl. **24**, 687 (1985).

[194] A. K. Rappé and W. A. Goddard III, Charge Equilibration for Molecular Dynamics Simulations. J. Phys. Chem. **95**, 3358 (1991).

[195] H. Goldstein, in *Classical mechanics, 2nd ed.* (Addison Wesley, London, 1980), p. 16.

[196] H. C. Andersen, Molecular dynamics simulations at constant pressure and/or temperature. J. Chem. Phys. **72**, 2384 (1980).

[197] D. L. Ermak and H. Buckholtz, Numerical integration of the Langevin equation: Monte Carlo simulation. J. Comput. Phys. **35**, 169 (1980).

[198] S. Chandrasekhar, Stochastic problems in physics and astronomy. Rev. Mod. Phys. **15**, 1 (1943).

[199] M. G. Martin, B. Chen, and J. I. Siepmann, A novel Monte Carlo algorithm for polarizable force fields: Application to a fluctuating charge model for water. J. Chem. Phys. **108**, 3383 (1998).

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

[201] B. Chen, J. J. Potoff, and J. I. Siepmann, Adiabatic Nuclear and Electronic Sampling Monte Carlo Simulations in the Gibbs Ensemble: Application to Polarizable Force Fields for Water. J. Phys. Chem. B **104**, 2378 (2000).

[202] R. A. Bryce, M. A. Vincent, N. O. J. Malcolm, I. H. Hillier, and N. A. Burton, Cooperative effects in the structuring of fluoride water clusters: Ab initio hybrid quantum mechanical/molecular mechanical model incorporating polarizable fluctuating charge solvent. J. Chem. Phys. **109**, 3077 (1998).

[203] R. A. Bryce, R. Buesnel, I. H. Hillier, and N. A. Burton, A solvation model using a hybrid quantum mechanical/molecular mechanical potential with fluctuating solvent charges. Chem. Phys. Lett. **279**, 367 (1997).

[204] GSL, The GNU Scientific Library, *http://www.gnu.org/software/gsl/.*

[205] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, in *Numerical Recipes in FORTRAN, Second Edition* (Cambridge University Press, Cambridge, UK, 1992), p. 107.

[206] C. L. Brooks III, B. M. Pettitt, and M. Karplus, Structural and Energetic Effects of Truncating Long Ranged Interactions in Ionic and Polar Fluids. J. Chem. Phys. **83**, 5897 (1985).

[207] M. Prevost, D. van Belle, G. Lippens, and S. Wodak, Computer-Simulations of Liquid Water — Treatment of Long-Range Interactions. Mol. Phys. **71**, 587 (1990).

[208] B. Chen, J. Xing, and J. I. Siepmann, Development of Polarizable Water Force Fields for Phase Equilibrium Calculations. J. Phys. Chem. B **104**, 2391 (2000).

[209] J. A. Odutola and T. R. Dyke, Partially deuterated water dimers: Microwave spectra and structure. J. Chem. Phys. **72**, 5062 (1980).

[210] L. A. Curtiss, D. J. Frurip, and M. Blander, Studies of molecular association in H_2O and D_2O vapors by measurement of thermal conductivity. J. Chem. Phys. **71**, 2703 (1979).

[211] G. H. F. Diercksen, W. P. Kraemer, and B. O. Roos, SCF-CI Studies of Correlation Effects on Hydrogen Bonding and Ion Hydration. Theor. Chim. Acta. **36**, 249 (1975).

[212] O. Matsuoko, E. Clementi, and M. Yoshimine, CI study of the water dimer potential surface. J. Chem. Phys. **64**, 1351 (1976).

[213] C. J. Marsden, B. J. Smith, J. A. Pople, H. F. Schaefer III, and L. Radom, Characterization of the bifurcated structure of the water dimer. J. Chem. Phys. **95**, 1825 (1991).

[214] M. Schütz, S. Brdarski, P.-O. Widmark, R. Lindh, and G. Karlström, The water dimer interaction energy: Convergence to the basis set limit at the correlated level. J. Chem. Phys. **107**, 4597 (1997).

[215] F. Sim, D. R. Salahub, and S. Chin, The Accurate Calculation of Dipole Moments and Dipole Polarizabilities Using Gaussian-Based Density Functional Methods. Intern. J. Quant. Chem. **43**, 463 (1992).

[216] N. Godbout and D. R. Salahub, Optimization of Gaussian-type basis sets for local spin density functional calculations. Part I. Boron through neon, optimization technique and validation. Can. J. Chem. **70**, 560 (1992).

[217] T. H. Dunning, Jr., Gaussian Basis Functions for Use in Molecular Calculations. III. Contraction of (10s6p) Atomic Basis Sets for the First-Row Atoms. J. Chem. Phys. **55**, 716 (1971).

[218] B. Roos and P. Siegbahn, Polarization Functions for First and Second Row Atoms in Gaussian Type MO-SCF Calculations. Theoret. Chim. Acta. **17**, 199 (1970).

[219] G. D. Zeiss, W. R. Scott, N. Suzuki, D. P. Chong, and S. R. Langhoff, Finite-field calculations of molecular polarizabilities using field-induced polarization functions: second- and third-order perturbation correlation corrections to the coupled Hartree-Fock polarizability of H_2O. Mol. Phys. **37**, 1543 (1979).

[220] D. Wei and D. R. Salahub, A combined density functional and molecular dynamics simulation of a quantum water molecule in aqueous solution. Chem. Phys. Lett. **224**, 291 (1994).

[221] R. V. Stanton, D. S. Hartsough, and K. M. Merz, Jr., An Examination of a Density Functional / Molecular Mechanical Coupled Potential. J. Comp. Chem. **16**, 113 (1995).

[222] R. V. Stanton, L. R. Little, and K. M. Merz, Jr., An Examination of a Hartree-Fock / Molecular Mechanical Coupled Potential. J. Phys. Chem. **99**, 17344 (1995).

[223] S. J. Stuart and B. J. Berne, Effects of Polarizability of the Hydration of the Chloride Ion. J. Phys. Chem. **100**, 11934 (1996).

[224] D. J. Tobias, P. Jungwirth, and M. Parrinello, Surface solvation of halogen anions in water clusters: An *ab initio* molecular dynamics study of the $Cl^-(H_2O)_6$ complex. J. Chem. Phys. **114**, 7036 (2001).

[225] D. Beglov and B. Roux, Finite representation of an infinite bulk system: Solvent boundary potential for computer simulations. J. Chem. Phys. **100**, 9050 (1994).

[226] R. Yamdagni and P. Kebarle, Hydrogen-Bonding Energies to Negative Ions from Gas-Phase Measurements of Ionic Equilibria. J. Am. Chem. Soc. **93**, 7139 (1971).

[227] J. B. Foresman and C. L. Brooks III, An *ab initio* study of hydrated chloride ion complexes: Evidence of polarization effects and nonadditivity. J. Chem. Phys. **87**, 5892 (1987).

[228] R. A. Bryce, M. A. Vincent, and I. H. Hillier, Binding Energy of $F(H_2O)^-$ and the Simulation of Fluoride Water Clusters Using a Hybrid QM/MM (Fluctuating Charge) Potential. J. Phys. Chem. A **103**, 4094 (1999).

[229] B. T. Thole, Molecular polarizabilities calculated with a modified dipole interaction. Chem. Phys. **59**, 341 (1981).

[230] M. Arshadi, R. Yamdagni, and P. Kebarle, Hydration of the Halide Negative Ions in the Gas Phase. II. Comparison of Hydration Energies for the Alkali Positive and Halide Negative Ions. J. Phys. Chem. **74**, 1475 (1970).

[231] S. S. Xantheas and L. X. Dang, Critical Study of Fluoride-Water Interactions. J. Phys. Chem. **100**, 3989 (1996).

[232] I. Džidić and P. Kebarle, Hydration of the Alkali Ions in the Gas Phase. Enthalpies and Entropies of Reactions $M^+(H_2O)_{n-1} + H_2O = M^+(H_2O)_n$. J. Phys. Chem. **74**, 1466 (1970).

[233] J. P. Perdew, Density-functional approximation for the correlation energy of the inhomogeneous electron gas. Phys. Rev. B **33**, 8822 (1986).

[234] I. Tuñón, M. T. C. Martins-Costa, C. Millot, M. F. Ruiz-López, and J. L. Rivail, A Coupled Density Functional-Molecular Mechanics Monte Carlo Simulation Method: The Water Molecule in Liquid Water. J. Comp. Chem. **17**, 19 (1996).

[235] Y. Tu and A. Laaksonen, On the effect of Lennard-Jones parameters on the quantum mechanical and molecular mechanical coupling in a hybrid molecular dynamics simulation of liquid water. J. Chem. Phys. **111**, 7519 (1999).

[236] W. R. Rocha, K. Coutinho, W. B. de Almeida, and S. Canuto, An efficient quantum mechanical/molecular mechanics Monte Carlo simulation of liquid water. Chem. Phys. Lett. **335**, 127 (2001).

[237] M. Freindorf and J. Gao, Optimization of the Lennard-Jones Parameters for a Combined *Ab Initio* Quantum Mechanical and Molecular Mechanical Potential Using the 3-21G Basis Set. J. Comp. Chem. **17**, 386 (1996).

[238] P.-O. Åstrand, K. Ruud, K. V. Mikkelsen, and T. Helgaker, Atomic Charges of the Water Molecule and the Water Dimer. J. Phys. Chem. A **102**, 7686 (1998).

[239] J. P. Ritchie and A. S. Copenhaver, Comparison of Potential-Derived Charge and Atomic Multipole Models in Calculating Electrostatic Potentials and Energies of Some Nucleic Acid Bases and Pairs. J. Comp. Chem. **16**, 777 (1995).

[240] Y.-P. Liu, K. Kim, B. J. Berne, R. A. Friesner, and S. W. Rick, Constructing ab initio force fields for molecular dynamics simulations. J. Chem. Phys. **108**, 4739 (1998).

[241] H. A. Stern, G. A. Kaminski, J. L. Banks, R. Zhou, B. J. Berne, and R. A. Friesner, Fluctuating Charge, Polarizable Dipole, and Combined Models: Parametrization from ab Initio Quantum Chemistry. J. Phys. Chem. B **103**, 4730 (1999).

[242] C. K. Ingold, *Structure and Mechanism in Organic Chemistry, 2nd ed.* (Cornell University Press, Ithaca, NY, 1969).

[243] N. S. Isaacs, *Physical Organic Chemistry* (Wiley, New York, 1987).

[244] H. C. Raine and C. N. Hinshelwood, The Functional Relation between the Constants of the Arrhenius Equation: The Formation of a Series of Quaternary Ammonium Salts in Benzene-Nitrobenzene Mixtures. J. Chem. Soc. 1378 (1939).

[245] N. J. T. Pickles and C. N. Hinshelwood, The Influence of Solvents on Reaction Velocity. The Interaction of Pyridine and Methyl Iodide and the Benzoylation of *m*-Nitroaniline. J. Chem. Soc. 1353 (1936).

[246] A. Dedieu and A. Veillard, in *Quantum theory of chemical reactions*, edited by R. Daudel, A. Pullman, L. Salem, and A. Veillard (D. Reidel Pub. Co., Boston, 1979), Vol. 1, p. 69.

[247] K. Morokuma, Potential-Energy Surface of the S_N2 Reaction in Hydrated Clusters. J. Am. Chem. Soc. **104**, 3732 (1982).

[248] D. K. Bohme and G. I. Mackay, Bridging the Gap between the Gas Phase and Solution: Transition in the Kinetics of Nucleophilic Displacement Reactions. J. Am. Chem. Soc. **103**, 978 (1981).

[249] J. Chandrasekhar, S. F. Smith, and W. L. Jorgensen, S_N2 Reaction Profiles in the Gas Phase and Aqueous Solution. J. Am. Chem. Soc. **106**, 3049 (1984).

[250] J. Chandrasekhar, S. F. Smith, and W. L. Jorgensen, Theoretical Examination of the S_N2 Reaction Involving Chloride Ion and Methyl Chloride in the Gas Phase and Aqueous Solution. J. Am. Chem. Soc. **107**, 154 (1985).

[251] J. Chandrasekhar and W. L. Jorgensen, Energy Profile for a Nonconcerted S_N2 Reaction in Solution. J. Am. Chem. Soc. **107**, 2974 (1985).

[252] Y. Mo and J. Gao, An Ab Initio Molecular Orbital-Valence Bond (MOVB) Method for Simulating Chemical Reactions in Solution. J. Phys. Chem. A **104**, 3012 (2000).

[253] Y. Mo and J. Gao, *Ab initio* QM/MM simulations with a molecular orbital-valence bond (MOVB) method: application to an S_N2 reaction in water. J. Comp. Chem. **21**, 1458 (2000).

[254] N. Menschutkin, Beiträge zur Kenntnis der Affinitätskoeffizienten der Alkylhaloide und der organischen Amine. Erster Teil. Affinitätskoeffizienten der aliphatischen Alkyljodide und Bromide. Z. Phys. Chem. **5**, 589 (1890).

[255] N. Menschutkin, Über die Affinitätskoeffizienten der Alkylhaloide und der Amine. Zweiter Teil. Über den Einfluss des chemisch indifferenten flüssigen Mediums auf die Geschwindigkeit der Verbindung des Triäthylamins mit den Alkyljodiden. Z. Phys. Chem. **6**, 41 (1890).

[256] Y. F. Yen, R. J. Cross, and M. Saunders, Crossed-Beam Study of the S_N2 Reaction. J. Am. Chem. Soc. **113**, 5563 (1991).

[257] G. S. Hammond, A Correlation of Reaction Rates. J. Am. Chem. Soc. **77**, 334 (1955).

[258] J. Gao and X. Xia, A Two-Dimensional Energy Surface for a Type II S_N2 Reaction in Aqueous Solution. J. Am. Chem. Soc. **115**, 9667 (1993).

[259] J.-L. M. Abboud, R. Notario, J. Bertràn, and M. Solà, One Century of Physical Organic Chemistry: The Menshutkin Reaction. Prog. Phys. Org. Chem. **19**, 1 (1993).

[260] V. Dillet, D. Rinaldi, J. Bertrán, and J.-L. Rivail, Analytical energy derivatives for a realistic continuum model of solvation: Application to the analysis of solvent effects on reaction paths. J. Chem. Phys. **104**, 9437 (1996).

[261] U. Maran, T. A. Pakkanen, and M. Karelson, Semiempirical Study of the Solvent Effect on the Menshutkin Reaction. J. Chem. Soc. Perkin Trans. 2 2445 (1994).

[262] S. Shaik, A. Ioffe, A. C. Reddy, and A. Pross, Is the Avoided Crossing State a Good Approximation for the Transition State of a Chemical Reaction? An Analysis of Menschutkin and Ionic S_N2 Reactions. J. Am. Chem. Soc. **116**, 262 (1994).

[263] J. W. Viers, J. C. Schug, M. D. Stovall, and J. I. Seeman, MNDO Study of Reaction Pathways for S_N2 Reactions. Menschutkin Reaction Potential Energy Surfaces. J. Comp. Chem. **5**, 598 (1984).

[264] X. Fradera, Ll. Amat, M. Torrent, J. Mestres, P. Constans, E. Besalí, J. Martí, S. Simon, M. Lobato, J. M. Oliva, J. M. Luis, J. L. Andrés, M. Solà, R. Carbó, and M. Duran, Analysis of the changes on the potential energy surface of Menshutkin reactions induced by external perturbations. J. Mol. Struct. (Theochem) **371**, 171 (1996).

[265] J. Gao, Hybrid Quantum and Molecular Mechanical Simulations: An Alternative Avenue to Solvent Effects in Organic Chemistry. Acc. Chem. Res. **29**, 298 (1996).

[266] T. N. Truong, T.-T. T. Truong, and E. V. Stefanovich, A general methodology for quantum modeling of free-energy profile of reactions in solution: An application to the Menshutkin $NH_3 + CH_3Cl$ reaction in water. J. Chem. Phys. **107**, 1881 (1997).

[267] M. D. Halls and H. B. Schlegel, Chemistry Inside Carbon Nanotubes: the Menshutkin S_N2 Reaction. J. Phys. Chem. B **106**, 1921 (2002).

[268] K. Okamoto, S. Fukui, and H. Shingu, Kinetic Studies of Bimolecular Nucleophilic Substitution. VI. Rates of the Menschtkin Reaction of Methyl Iodide with Methylamines and Ammonia in Aqueous Solutions. Bull. Chem. Soc. Jpn. **40**, 1920 (1967).

[269] K. Okamoto, S. Fukui, I. Nitta, and H. Shingu, Kinetic Studies of Bimolecular Nucleophilic Substitution. VIII. The Effect of Hydroxylic Solvents on the Nucleophilicity of Aliphatic Amines in the Menschutkin Reaction. Bull. Chem. Soc. Jpn. **40**, 2354 (1967).

[270] S. P. Webb and M. S. Gordon, Solvation of the Menshutkin Reaction: A Rigorous Test of the Effective Fragment Method. J. Phys. Chem. **103**, 1265 (1999).

[271] J. H. Jensen, P. N. Day, M. S. Gordon, H. Basch, D. Cohen, D. R. Garmer, M. Kraus, and W. J. Stevens, in *Modeling the Hydrogen Bond*, edited by D. A. Smith (American Chemical Society, Washington, 1994), pp. 139–151.

[272] P. N. Day, J. H. Jensen, M. S. Gordon, S. P. Webb, W. J. Stevens, M. Krauss, D. Garmer, H. Basch, and D. Cohen, An effective fragment method for modeling solvent effects in quantum mechanical calculations. J. Chem. Phys. **105**, 1968 (1996).

[273] M. Solà, A. Lledós, M. Duran, J. Bertrán, and J.-L. M. Abboud, Analysis of Solvent Effects on the Menshutkin Reaction. J. Am. Chem. Soc. **113**, 2873 (1991).

[274] H. Castejon and K. B. Wiberg, Solvent Effects on Methyl Transfer Reactions. 1. The Menshutkin Reaction. J. Am. Chem. Soc. **121**, 2139 (1999).

[275] C. Amovilli, B. Mennucci, and F. M. Floris, MCSCF Study of the S_N2 Menshutkin Reaction in Aqueous Solution within the Polarizable Continuum Model. J. Phys. Chem. B **102**, 3023 (1998).

[276] T. N. Truong and E. V. Stefanovich, A new method for incorporating solvent effect into the classical, ab initio molecular orbital and density functional theory frameworks for arbitrary shape cavity. Chem. Phys. Lett. **240**, 253 (1995).

[277] J. Gao, A Priori Computation of a Solvent-Enhanced S_N2 Reaction Profile in Water: The Menshutkin Reaction. J. Am. Chem. Soc. **113**, 7796 (1991).

[278] A. Warshel and R. M. Weiss, An Empirical Valence Bond Approach for Comparing Reactions in Solutions and in Enzymes. J. Am. Chem. Soc. **102**, 6218 (1980).

[279] J.-K. Hwang, G. King, S. Creighton, and A. Warshel, Simulation of Free Energy Relationships and Dynamics of S_N2 Reactions in Aqueous Soluion. J. Am. Chem. Soc. **110**, 5297 (1988).

[280] J. B. Pedley, *Thermochemical Data and Structures of Organic Compounds* (Thermodynamics Research Center, College Station, TX, 1994), Vol. 1.

[281] S. Nosé and M. L. Klein, Constant pressure molecular dynamics for molecular systems. Mol. Phys. **50**, 1055 (1983).

[282] S. E. Feller, Y. Zhang, R. W. Pastor, and B. R. Brooks, Constant pressure molecular dynamics simulation: The Langevin piston method. J. Chem. Phys. **103**, 4613 (1995).

[283] S. Schröder, V. Daggett, and P. Kollman, A Comparison of the AM1 and PM3 Semiempirical Models for Evaluating Model Compounds Relevant to Catalysis by Serine Proteases. J. Am. Chem. Soc. **113**, 8922 (1991).

[284] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Quantum Monte Carlo simulations of solids. Rev. Mod. Phys. **73**, 33 (2001).

[285] R. Bahnsen, H. Eckstein, W. Schattke, N. Fitzer, and R. Redmer, Variational quantum Monte Carlo calculations for solid surfaces. Phys. Rev. B **63**, 235415 (2001).

[286] J. Gao and X. Xia, A Priori Evaluation of Aqueous Polarization Effects Through Monte Carlo QM-MM Simulations. Science **258**, 631 (1992).

[287] D. Bouzida, S. Kumar, and R. H. Swendsen, Efficient Monte-Carlo methods for the computer simulation of biological molecules. Phys. Rev. A **45**, 8894 (1992).

[288] J. C. Owicki and H. A. Scheraga, Preferential Sampling Near Solutes in Monte Carlo Calculations on Dilute Solutions. Chem. Phys. Lett. **47**, 600 (1977).

[289] J. C. Owicki, in *Computer Modeling of Matter, ACS Symposium Series Vol. 86*, edited by P. Lykos (American Chemical Society, Washington, 1978), pp. 159–171.

[290] W. L. Jorgensen, B. Bigot, and J. Chandrasekhar, Nature of Dilute Solutions of Sodium and Methoxide Ions in Methanol. J. Am. Chem. Soc. **104**, 4584 (1982).

[291] T. N. Truong and E. V. Stefanovich, Development of a perturbative approach for Monte Carlo simulations using a hybrid ab initio QM/MM method. Chem. Phys. Lett. **256**, 348 (1996).

[292] T. J. Evans and T. N. Truong, Optimizing Efficiency of Perturbative Monte Carlo Method. J. Comp. Chem. **19**, 1632 (1998).

[293] A. R. Dinner, *Monte Carlo Simulations of Protein Folding*, Ph.D. thesis, Harvard University, Cambridge, MA, 1999.

[294] L. Verlet, Computer 'Experiments' on Classical Fluids. II. Equilibrium Correlation Functions. Phys. Rev. **165**, 201 (1967).

[295] W. J. Hehre, L. Radom, P. von R. Schleyer, and J. A. Pople, *Ab Initio Molecular Orbital Theory* (Wiley, New York, 1986).

[296] J. A. Pople, M. Head-Gordon, D. J. Fox, K. Raghavachari, and L. A. Curtiss, Gaussian-1 theory: A general procedure for prediction of molecular energies. J. Chem. Phys. **90**, 5622 (1989).

[297] L. A. Curtiss, C. Jones, G. W. Trucks, K. Raghavachari, and J. A. Pople, Gaussian-1 theory of molecular energies for second-row compounds. J. Chem. Phys. **93**, 2537 (1990).

[298] L. A. Curtiss, K. Raghavachari, G. W. Trucks, and J. A. Pople, Gaussian-2 theory for molecular energies of first- and second-row compounds. J. Chem. Phys. **94**, 7221 (1991).

[299] L. A. Curtiss, K. Raghavachari, and J. A. Pople, Gaussian-2 theory using reduced Møller-Plesset orders. J. Chem. Phys. **98**, 1293 (1993).

[300] K. Byun, Y. Mo, and J. Gao, New Insight on the Origin of the Unusual Acidity of Meldrum's Acid from ab Initio and Combined QM/MM Simulation Study. J. Am. Chem. Soc. **123**, 3974 (2001).

[301] W. Yang and T.-S. Lee, A density-matrix divide-and-conquer approach for electronic structure calculations of large molecules. J. Chem. Phys. **103**, 5674 (1995).

[302] S. K. Goh, C. P. Sosa, and A. St-Amant, A scalable divide-and-conquer algorithm combining coarse and fine-grain parallelization. Theor. Chem. Acc. **99**, 197 (1998).

[303] A. Z. Panagiotopoulos, Direct determination of phase coexistence properties of fluids by Monte Carlo simulation in a new ensemble. Mol. Phys. **61**, 813 (1987).

[304] J. Åqvist and A. Warshel, Free Energy Relationships in Metalloenzyme-Catalyzed Reactions. Calculations of the Effects of Metal Ion Substitutions in Staphylococcal Nuclease. J. Am. Chem. Soc. **112**, 2860 (1990).

[305] M. A. Thompson, QM/MMpol: A Consistent Model for Solute/Solvent Polarization. Application to the Aqueous Solvation and Spectroscopy of Formaldehyde, Acetaldehyde, and Acetone. J. Phys. Chem. **100**, 14492 (1996).