Hybrid Simulation Methods for Systems in Condensed Phase

Dissertation
for the award of the degree
“Doctor rerum naturalium” (Dr.rer.nat.)
of the Georg-August-Universität Göttingen

within the doctoral program of Chemistry
of the Georg-August University School of Science (GAUSS)

submitted by
Jonas Feldt
from Neustadt am Rübenberge

Göttingen, 2018

Thesis Committee

Prof. Dr. Ricardo Mata
Institute of Physical Chemistry, Georg-August-Universität Göttingen
Prof. Dr. Burkhard Geil
Institute of Physical Chemistry, Georg-August-Universität Göttingen

Members of the Examination Board

Reviewer:
Prof. Dr. Ricardo Mata
Institute of Physical Chemistry, Georg-August-Universität Göttingen
Second Reviewer:
Prof. Dr. Burkhard Geil
Institute of Physical Chemistry, Georg-August-Universität Göttingen

Further members of the Examination Board

Prof. Dr. Jörg Behler
Institute of Physical Chemistry, Georg-August-Universität Göttingen
Prof. Dr. Inke Siewert
Institute of Inorganic Chemistry, Georg-August-Universität Göttingen
Prof. Dr. Konrad Koszinowski
Institute of Organic and Biomolecular Chemistry, Georg-August-Universität Göttingen
Priv.-Doz. Dr. Thomas Zeuch
Institute of Physical Chemistry, Georg-August-Universität Göttingen

Date of the oral examination: 08.03.2018

Acknowledgments

First of all, I would like to thank Prof. Dr. Mata for supporting my research career for many years, for allowing me freedom to pursue my own scientific interests and for finding always time for discussions. I believe he prepared me very well for my further path in academia.

I would like to thank Prof. Dr. Geil for agreeing to be the second supervisor for this thesis. I am also thankful to all the members of the examination board: Prof. Behler, Prof. Siewert, Prof. Koszinowski and Priv.-Doz. Zeuch for taking interest in my work and finding time for the examination.

Special thanks goes to Sebastião Miranda who worked during his master thesis together with me on this project. The months that he spent here in Göttingen have been the most productive time of my thesis and I would like to thank him for the excellent collaboration, many fruitful discussions and his willingness to understand my theoretical chemistry ideas.

I would like to thank all former and current members of the Computational Chemistry and Biochemistry group. Special thanks to Johannes Dieterich, an excellent mentor for software development, as well as Thorsten Stolper and Rainer Oswald for many interesting discussions, the good atmosphere and for being fellow espressionists.

I would like to thank my wife Milica who did not only support me in my scientific endeavours but kept my spirit up for all these years and most importantly for showing me the world.

Finally, many thanks to my parents for their unconditional support for all these years.

Abstract

Reactions in solution are important in the chemical and pharmaceutical industry as well as in biochemical contexts. In fact, the solvent effects are stronger than many other factors and can slow down or speed up a reaction by many orders of magnitude. However, the complexity of chemistry in solution poses a serious challenge for the computational chemistry community. Therefore, methods have been developed starting from different approximations which are well suited for some aspects but are forced to neglect others.

Two aspects have to be considered in order to model a system in solution. The first aspect is the description of the potential energy, which have to be described accurately. Methods ranging from a purely classical description to a complete quantum mechanical one are available. The second aspect concerns temperature effects. Molecular simulations are commonly used although continuum solvation models pose an alternative since they include these effects implicitly.

In this work I will present a novel hybrid quantum mechanics/molecular mechanics approach. The solute is described with quantum mechanical methods allowing to account for reactivity and polarisation effects while the solvent is described by molecular mechanics. This strikes a good compromise between accuracy and computational costs for the energetics. The simulations are carried out with the Metropolis Monte Carlo method. High efficiency is achieved by three key approaches: (1) computation of the electrostatic coupling between solute and solvent with 1^{st} order perturbation theory, (2) efficient evaluation of the long-range electrostatics with a shifted force operator, (3) efficient evaluation of the interactions by an numerical integration implemented for graphical processing units.

The influence of the parameters inherent to our approach has been thoroughly investigated on a number of benchmark systems. Empirical guidelines have been established along the way which have been used for subsequent applications to biochemically relevant systems e.g. the mutagenic properties of halogenated uracil bases. Properties like solvent structures and electronic spectra as well as relative free energies can be computed efficiently by the here presented approach.

Contents

Acknowledgments **v**

Abstract **vii**

1 Introduction **1**

2 Theory **3**

2.1 Quantum Mechanics . . . 3
2.1.1 General Wave Function Theory . . . 3
2.1.2 Density Functional Theory . . . 5
2.1.3 Numerical Integration in DFT . . . 9
2.2 Molecular Mechanics . . . 13
2.2.1 Bonded Terms . . . 14
2.2.2 Non-Bonded Terms . . . 15
2.2.3 Long-Range Electrostatic Interactions . . . 17
2.3 Solvation Models . . . 21
2.3.1 Continuum Solvation Methods . . . 21
2.3.2 Reference Interaction Site Model . . . 25
2.3.3 Explicite Solvation with Hybrid QM/MM Methods . . . 27
2.4 Molecular Simulations . . . 30
2.4.1 Molecular Dynamics . . . 31
2.4.2 Metropolis Monte Carlo Simulations . . . 33
2.5 Free Energies in Solution . . . 36
2.6 Computation of Electronic Spectra . . . 38

3 The Perturbative Metropolis Monte Carlo Method **41**

3.1 Basic Formulation . . . 42
3.1.1 Limits of the Perturbation Approach . . . 44
3.1.2 Computation of the Electrostatic Perturbation Term . . . 45
3.2 Implementation in Hybrid Architectures . . . 47
3.3 Additional Modules of the PMC Program Suite . . . 50

4 Benchmarks 55

4.1 Numerical Integration . . . 55

4.2 Updates of the Density . . . 59

4.3 Free Energies of Solvation . . . 62

4.4 Free Energy of Torsional Potentials . . . 69

5 Impact of QM/MM on Solvent Structures 71

5.1 Hydrogen Peroxide . . . 72

5.2 Methylchloride . . . 77

6 QM/MM Calculations of Electronic Spectra in Solution 83

6.1 The PMC Methodology - Absorption and Emission Spectra . . . 84

6.2 Applications in Solute-Solvent Systems . . . 85

6.2.1 Formaldehyde . . . 85

6.2.2 Propenal . . . 89

6.3 Applications in Binary Solvent Mixtures . . . 91

6.3.1 Sampling of the Conformational Space . . . 91

6.3.2 Nitroaniline . . . 94

7 Differential Solvation Effect of the Uracil Base 99

7.1 Computational Details . . . 102

7.2 Ionisation Model . . . 103

7.3 Tautomerisation Model . . . 105

7.4 Wobble Model . . . 108

7.5 Deactivation Pathways . . . 109

8 Summary and Future Work 113

Bibliography 116

Curriculum Vitae 139

List of Figures

2.1 Grid for an ethanol structure according to Lebedev for the angular integration and Mura and Knowles for the radial integration. 12
2.2 Schematic representation of the different terms in MM methods. 13
2.3 Rotation potential corresponding to Equation 2.62 with $V_1 = 0.5$, $V_2 = -0.2$ and $V_3 = -0.5$ and all $\phi_i = 0$. 15
2.4 Schematic representation of PBC in two dimensions for a square. The original cell and particles in black and the images in green. For one of the particles a spherical cut-off is shown. 18
2.5 Coulomb interaction of two unit charges with opposing signs for the Coulomb, a shifted and a shifted force potential, both with $\alpha = 0$. 20
2.6 Schematic representation of the solute (red) inside of a cavity (green) in the continuous solvent (black waves). 22

3.1 Schematic representation of the PMC algorithm with the functions computing the Coulomb and vdW interaction of the MM system and the Coulomb and vdW interaction between QM and MM system. 48
3.2 Schmematic representation of the PMC algorithm with multiple Markov chains being generated in parallel. 49
3.3 Mindmap of modules and functionalities included in the PMC program suite. . 51
3.4 Schematic representation of the memory layout of different algorithms storing the distances. Left: full matrix without diagonal, middle: only the upper triangle, right: the upper triangle mapped onto a vector. 53

4.1 Arginine cation with NMe cap at the C terminus and acetate cap at the N terminus with a sphere of TIP3P water. 56
4.2 Electrostatic interaction for different grids with the exact results as a reference for snapshots of a PMC simulation of a capped arginine cation in TIP3P water at the PBE/def2-SVP level of theory. 57
4.3 Comparison of relative potential energies of PMC simulations with B3LYP-D3/def2-TZVP and OPLS-AA parameters of the same trajectory with different grid sizes. $-\log(\text{Target Accuracy})$ is given in the legend with the reference being 11. 58

4.4 Decoupling of the electrostatic interaction of an free energy perturbation calculation for ethanol in water with PMC (B3LYP-D3/def2-TZP and OPLS-AA vdW parameters) with different grid sizes. $-\log(\text{Target Accuracy})$ is given in the legend and the energies are relative to 11. The shaded area denotes 5 % of the free energy (≈ 1.8 kJ/mol). 59

4.5 Free solvation energy simulations with PMC (B3LYP-D3/def2-TZVP and OPLS-AA vdW parameters) with increasing frequency of density updates for one QM water in TIP3P solvent. Top: First λ step. Bottom: Distributions of interaction energies between QM and environment. 61

4.6 The potential energy of a simulation with PMC (B3LYP-D3/def2-TZVP and OPLS-AA vdW parameters) with increasing frequency of density updates for one QM water in TIP3P solvent. Left: ensemble average and standard error, right: standard deviation. 62

4.7 Solvation free energies computed with PMC (B3LYP-D3/def2-TZVP and OPLS-AA, OPLS-UA for chloroform) in comparison with experimental results. The solutes and solvents are listed in table 4.1. The colour coding is water→blue, toluene→black, acetonitrile→green and chloroform→red. 65

4.8 Solvation free energies computed with PMC (B3LYP-D3/def2-TZVP and OPLS-AA, OPLS-UA for chloroform) and MM MC (OPLS-AA, OPLS-UA for chloroform) in comparison with experimental results. The solutes and solvents are listed in table 4.1. 65

4.9 Solvation free energies computed with PMC (B3LYP-D3/def2-TZVP and OPLS-AA vdW parameters) in chloroform with standard parameters (OPLS-UA) and chloroform with an increased dipole moment (OPLS-UA(1.3x)). 68

4.10 MC Radial distribution functions for pure chloroform with OPLS-UA and an increased dipole moment (OPLS-UA(1.3x)). Top left: Cl-Cl, top right: Cl-CH, bottom: CH-CH. 68

4.11 The torsional potential of hydrogen peroxide computed in gas phase, with COSMO ($\epsilon = 80$) and in TIP3P water with PMC (B3LYP-D3/def2-TZVP). . . 70

5.1 PMC Radial distribution function $g(r)$ of donating (O···HW) and accepting (H···OW) hydrogen bonds of hydrogen peroxide and the oxygen-oxygen function (O-OW). The full lines have been simulated with B3LYP and the dashed ones with PBE. 73

5.2 PMC (B3LYP-D3) Radial distribution functions for pairs X-Y with the QM atoms X and the MM site Y. HW and OW denote the oxygen and hydrogen atom of the respective water model and LP the site describing the lone pairs of a water molecule (only for TIP5P). 76

5.3 Left: Nucleophilic substitution involving chloride and fluoromethane. Right: Definitions of distance and angle for the analysis of the angular radial distribution functions. 77

5.4 PMC Radial distribution functions of the leaving Cl in the educt state with B3LYP-D3/def2-TZVP and OPLS-AA parameters. Left: Cl to H of water, Right: Cl to O of water with the literature results X=Cl, Y=Cl [148] and X=Br, Y=Cl [136] (at the transition state). 78

5.5 PMC Radial distribution functions of the leaving Cl in the educt state with the water models TIP3P, TIP4P and TIP5P and the QM approach B3LYP-D3/def2-TZVP. 79

5.6 Angular radial distribution function for the educt of the S_N2 reaction with PMC (B3LYP-D3/def2-TZVP and OPLS-AA vdW parameters) using the TIP3P water model and MM (OPLS-AA and TIP3P). 80

5.7 Angular radial distribution function for the educt of the S_N2 reaction with PMC (B3LYP-D3/def2-TZVP and OPLS-AA vdW parameters) and the water models TIP4P (left) and TIP5P (right). 81

5.8 Schematic representation of the sequential QM/MM approach in the case of a strong dependence of the sampling on the level of theory. The low level (LL) is shown in green, the high level (HL) in red and the sampled distribution in gray. The energy corrections are represented by the arrows. 81

6.1 Electronic excitations for formaldehyde in TIP3P water computed for snapshots of MM MC (OPLS-AA) simulations and based on PMC (B3LYP/aug-cc-pVDZ and OPLS-AA vdW parameters) simulations. 89

6.2 Electronic excitations for propenal in TIP3P water computed on-the-fly with PMC (B3LYP/aug-cc-pVDZ and OPLS-AA vdW parameters) simulations. . . . 91

6.3 Simulations with conventional Metropolis MC and swap moves analyzed every 10 k steps. The number of particles in one half of the system $n_{1/2}$ is measured. 92

6.4 Lewis structures of the three isomers of nitroaniline — 2-nitroaniline (2-NA), 3-nitroaniline (3-NA) and 4-nitroaniline (4-NA) — and the mesomeric structures for 2-NA and 4-NA. 94

6.5 The PMC (B3LYP/aug-cc-pVDZ) computed solvent shifts in the binary solvent mixture of cyclohexane and tetrahydrofuran of 2-nitroaniline with experimental results [215] (top left), 3-nitroaniline (top right) and 4-nitroaniline (bottom). The dashed lines denote the experimental shift for x(THF)=1. For 3-NA the linear shift has been plotted as well with the continuous black line. 96

7.1 Lewis structures of the nucleobases guanine (G), adenine (A), thymine (T) or uracil (U) and cytosine (C). 99

7.2 Lewis structures of the Watson-Crick base pairs as occurring in DNA with the backbone indicated by R. Hydrogen bonds are shown in dashed lines. 100

7.3 Schematic representation of the replication of DNA. This work has been created by Mariana Ruiz. [220] 100

7.4 Schematic representation of the assumed mechanism that leads to point mutations during the DNA replication. Each branching represents a DNA replication and the activated enol Uol is one of the intermediates shown in Figure 7.5. 102
7.5 Top: Lewis structures of the intermediates corresponding to the ionisation (U^-), tautomerisation (Uol) as well as U and G with hydrogen bond donor and acceptor capabilities indicated by arrows. Bottom: The resulting base pairs. 103
7.6 Thermodynamic cycle with dissociation and solvation energies. Square brackets denote a system in aqueous solution. 104
7.7 Top: Definition of the dihedral angle with the involved atoms marked in green. Bottom: Potential energy computed with B3LYP-D3/def2-SVP and COSMO ($\epsilon = 80.0$) as a function of the dihedral angle for the U^--G and BrU^--G base pairs. 105
7.8 Schematic representation of the experimental approach to determine the ration of the tautomers of uracil. 106
7.9 Thermodynamic cycle with tautomerisation and solvation free energies. Square brackets denote a system in aqueous solution. 107
7.10 Lewis structure of the deactivation step from G-Uol to Gol-U. 110
7.11 Schematic representation of the proposed deactivation mechanism. Each branching represents a DNA replication. 110
7.12 The potential energy surface computed with B3LYP-D3/def2-SVP and COSMO ($\epsilon = 80$) for the proton transfer between Uol and G (left) and BrUol and G (right). The colour scale is in kcal/mol. 111

List of Tables

3.1 Computation of the integral (eq. 3.21) for different values of T in atomic units. The values $F_k(T^*)$ are precomputed and tabulated. 46

3.2 Kernel execution times of a CPU (i7-4770K) reference implementation and a GPU (GTX 780Ti) implementation. 49

4.1 All combinations of solute and solvent that have been used to compute free energies of solvation with the standard error σ and experimental results (Exp.). [117] All energies are given in kJ/mol. Computations (Comp.) have been carried out with PMC (B3LYP-D3/def2-TZVP and OPLS-AA, OPLS-UA for chloroform). 66

5.1 Comparison of the peak positions of the RDFs shown in Figure 5.1 with results from Cabral [138]: BOMD (B3LYP-D3), Ruiz-Lopéz [125]: QM/MM MD (B3LYP/6-31G*, TIP3P), Coutinho [127]: MM MC (TIPS), Yang [139]: MM MD (ABEMM), Rode [140]: QMCF MD (HF/DZP, BJH-CF2) and Choi [134]: QM/MM-MD (MP2/6-31G*, TIP5P). Not reported distributions are left empty while peaks that are not present are denoted by ”–”. 74

6.1 Theoretical studies for the first n-π^* excitation of formaldehyde in gas phase ω^{gp} and the solvent shift $\Delta\omega$ in aqueous solution compared with PMC and sequential QM/MM simulations from this work both with B3LYP/aug-cc-pVDZ denoted by * and experimental results for formaldehyde in gas phase and the solvent shift for acetone as well as for highly concentrated formaldehyde. . . . 86

6.2 The solvent shifts in eV upon changing from cyclohexane to tetrahydrofuran from experiments (2-NA, [215] 3-NA and 4-NA [216]), COSMO (B3LYP/aug-cc-pVDZ) and PMC (B3LYP/aug-cc-pVDZ and OPLS-AA vdW parameters) calculations. 94

7.1 Free tautomerisation energy in gas phase and solution kcal/mol and $K_{\text{taut}} = \frac{\text{[Enol]}}{\text{[Ketone]}}$. The gas phase calculations from the work of Luque and coworkers [234] used the 6-311+G(d,p) basis set and HF/6-311+G(d,p) optimized geometries. For the PMC (B3LYP-D3/def2-SVP and OPLS-AA vdW parameters) simulations the level of theory for the gas phase reaction is specified in the Table. 108

7.2 Interaction energies computed with B3LYP-D3/def2-SVP and COSMO ($\epsilon = 80$) in kcal/mol. 109

List of Acronyms

A	Adenine
ASEP	Averaged Solvent Electrostatic Potential
BAR	Bennett Acceptance Ratio
BrU	5-Bromouracil
CCSD	Coupled Cluster with Singles and Doubles
C	Cytosine
CIS	Configuration Interaction with Single Excitations
COSMO	Conductor-like Screening Model
CPMD	Car-Parrinello Molecular Dynamics
CPU	Central Processing Unit
DFT	Density Functional Theory
DNA	Deoxyribonucleic Acid
EOM-CCSD	Equations of Motion - Coupled Cluster with Singles and Doubles
FEP	Free Energy Perturbation
GGA	Generalised Gradient Approximation
G	Guanine
GMIPp	Generalized Molecular Interaction Potential with Polarization Correction
GPU	Graphics Processing Unit
GTO	Gauss-type Orbital
HF	Hartree-Fock
HNC	Hypernetted Chain Closure

LCAO	Linear Combination of Atomic Orbitals
LJ	Lennard-Jones
LSDA	Local Spin Density Approximation
MC	Metropolis Monte Carlo
MD	Molecular Dynamics
MM	Molecular Mechanics
MSA	Mean Spherical Approximation
OPLS-AA	Optimized Potentials for Liquid Simulations - All Atoms
PBC	Periodic Boundary Conditions
PCM	Polarizable Continuum Model
PMC	Perturbative QM/MM Metropolis Monte Carlo
PRNG	Pseudo Random Number Generator
QM/MM	Quantum Mechanics/Molecular Mechanics
QM	Quantum Mechanics
RDF	Radial Distribution Function
RISM	Reference Interaction Site Model
RMSD	Root Mean Squared Deviation
RNA	Ribonucleic Acid
SCF	Self-Consistent Field
SMD	Solvent Model Density
STO	Slater-type Orbital
TD-DFT	Time-Dependent Density Functional Theory
T	Thymine
U	Uracil
vdW	van der Waals

CHAPTER 1

Introduction

Water is an ubiquitous solvent in biochemical contexts and the most abundant component of living organisms. In fact, it accounts in most organisms for about 70% of their weight. In humans, for example, more than two thirds of this water is found inside of the cells where the many chemical reactions of the metabolism take place. Without any doubt water is the most important solvent for biological processes and it participates itself in many of them. [1,2] As an example, water is involved in the formation of heteropolypeptides and subsequently catalyses the dynamical disorder of the same. These peptides have been suggested as the precursor for the amino acid based life hence water plays a key role for the chemical origin of life. [3,4]

A different aspect is the wide use of solvents in the chemical industry as reaction media and equally important in order to purify chemicals. The size of the global solvent market in 2015 was about USD 20 billions. [5] Consequently, this aspect has received considerable attention in order to devise new techniques, synthetic routes or optimised solvents in order to reduce the solvent consumption and the environmental impact. [6,7]

Interestingly enough, the biochemical and industrial paths coincide when considering the activity of enzymes in organic solvents. The highest enzymatic activity can be found when the conformational mobility and structure of the enzyme is close to its native state. [8] Theoretical chemistry has a key role in understanding these systems because it allows to study such effects at an atomistic level. Warshel already realised nearly four decades ago that the understanding of enzyme reactivity requires the understanding of chemistry in solution and that large macromolecules like enzymes can be seen as special solvents themselves. [9,10] For these efforts to theoretically study complex chemical systems he has been honoured with the Nobel prize in chemistry in the year 2013.

Understanding chemistry in solution has been a driving force for the development of new models, theories and computational approaches from the very beginning in the field of theoretical chemistry as well as related disciplines. And this interest never ceased as it can be seen for example in the cluster of excellence RESOLV [11] which aims not only at an understanding of solvent controlled processes but one step further at designing them.

At the molecular level one of the most important aspects is the structure of solvent molecules around a solute. This arrangement in shells requires a detailed understanding of all involved interactions and the underlying physics. Due to this high complexity available methods focus on different aspects of solvation and are forced in turn to make approximations for others which leads to many different approaches. Currently a selection of methods includes classical i.e. force field-based molecular dynamics or Monte Carlo simulations, hybrid approaches which combine quantum and classical mechanics, empirical methods based on macroscopic properties to screen electrostatic interactions, continuum descriptions of solvents or the statistical mechanics based reference interaction site models. [1]

Quantum mechanics can in principle compute chemical reactions and effects accurately. However, even the considerable advances in the computer hardware and the associated increase in the computing power do not allow to describe these complex systems completely with quantum mechanics. On the other hand, pure molecular mechanics lacks the accuracy and transferability that is required for a quantitative understanding. Hybrid quantum mechanical/molecular mechanical methods pose a viable alternative but nevertheless require large amounts of computational resources. Studying chemistry in solute with these methods on a routinely basis with commodity hardware has been the motivation for this work. By developing an efficient hybrid method for simulations which harnesses recent technological advancements of graphic cards an important step has been made in that direction. In this hybrid scheme perturbation theory has been applied to quantum mechanics/molecular mechanics calculations and combined with Monte Carlo simulations.

This thesis has been partitioned in the following way: The second Chapter describes first the basics of quantum mechanics including the developments of density functional theory which has been used for the studies in this work. Second, classical force fields are introduced with a special focus on the treatment of long-range electrostatic interactions. Then the coupling of these methods and further ways to model solvent effects are presented. Finally, molecular simulation methods and possibilities to compute properties in solution are described.

A description of the method developed in this work is given in Chapter three. First the theoretical basis and limitations of the approach are discussed. Then the focus is laid on the implementation of this method in the context of graphics cards. This Chapter concludes with a description of the infrastructure and suite of modules that has grown over the years around the core of the hybrid simulation algorithm.

Any method development is closely intertwined with a continuous process which reveals errors, limits of the approximations employed and finally the correctness of the here developed approach. This has been investigated in Chapter four. Simple systems have been studied and the results are compared to experimental and theoretical findings.

The Chapters five, six and seven are devoted to different aspects of solvation in the context of computational chemistry. First, the ability of hybrid methods to describe the solvent structure is assessed in Chapter five. Second, the computation of electronic spectra is investigated in Chapter six. Last, in Chapter seven the newly established methodology is used to study the differential solvent effects of the uracil base which are of biochemical interest.

At the end in Chapter eight this thesis finishes with a summary and discusses paths for future work and further improvements of this approach to model solvent effects.

CHAPTER 2

Theory

2.1 Quantum Mechanics

Quantum Mechanics (QM) builds the underlying theory to describe the physics of atomic and subatomic particles. In the field of chemistry, which usually focuses on the description of valence electrons, one then refers to quantum chemistry. The laws of classical physics are generally recovered from QM for the limit of large length scales. Newton's second law $F = ma$ describes the evolution of particles in classical physics, while the time-dependent Schrödinger equation is the analogue in QM. Especially of interest are stationary states which occur when the wave functions described by the time-dependent Schrödinger equation form standing waves. These stationary states can be described by the simpler time-independent Schrödinger equation

$$\hat{H}\Psi = E\Psi \tag{2.1}$$

with the Hamilton operator $\hat{H}$, the wave function Ψ and the energy E. Two different approaches to solve this equation will be discussed in the remainder of this Section. There will be one further assumption. This is, that the motion of the electrons and the nuclei can be separated because the electrons are orders of magnitudes lighter than the nuclei. As a result the Schrödinger equation depends only parametrically on the position of the nuclei and solving this electronic Schrödinger equation leads to the potential energy surface. This constitutes the Born-Oppenheimer approximation.

2.1.1 General Wave Function Theory

The Hartree-Fock (HF) theory is the starting point for what is dubbed *ab initio* electronic structure methods. Every electron is described by an orbital which is a function of the coordinates of the latter and its spin. The wave function is constructed as a Slater determinant which respects the anti-symmetry requirement. Namely, that only the sign of the wave function changes upon exchange of two indistinguishable electrons. [12]

The electronic Hamilton operator $\mathbf{H}$ in atomic units for N_{elec} electrons and N_{nuc} nuclei is given by

$$\mathbf{H} = \mathbf{T}_{\text{e}} + \mathbf{V}_{\text{ne}} + \mathbf{V}_{\text{ee}} + \mathbf{V}_{\text{nn}} \tag{2.2}$$

with the kinetic energy operator

$$\mathbf{T}_{\mathrm{e}} = -\frac{1}{2}\sum_{i}^{N_{\mathrm{elec}}} \nabla_i^2, \tag{2.3}$$

the potential energy between nuclei and electrons

$$\mathbf{V}_{\mathrm{ne}} = -\sum_{a}^{N_{\mathrm{nuc}}}\sum_{i}^{N_{\mathrm{elec}}} \frac{Z_a}{|\boldsymbol{R}_a - \boldsymbol{r}_i|}, \tag{2.4}$$

the potential energy between electrons and electrons

$$\mathbf{V}_{\mathrm{ee}} = \frac{1}{2}\sum_{i}^{N_{\mathrm{elec}}}\sum_{j}^{N_{\mathrm{elec}}} \frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} \tag{2.5}$$

and the potential energy between nuclei and nuclei

$$\mathbf{V}_{\mathrm{nn}} = \frac{1}{2}\sum_{a}^{N_{\mathrm{nuc}}}\sum_{b}^{N_{\mathrm{nuc}}} \frac{Z_a Z_b}{|\boldsymbol{R}_a - \boldsymbol{R}_b|}. \tag{2.6}$$

The last term is constant for a given nuclear configuration. These equations can be solved by means of the variational principle which states that the exact wave function is a lower boundary to any trial wave function. Consequently, the parameters of a trial wave function can be optimised by minimising the energy.

The resulting non-linear equations lead to the solution of an effective one-electron operator. The motion of an electron is solved under the influence of the average field of the other electrons. This is denoted as mean-field theory. However, this results in an iterative procedure what is commonly referred to as Self-Consistent Field (SCF) theory. Applying this procedure yields the electronic energy of the ground state as well as a set of occupied molecular orbitals with associated orbital energies.

Realistic systems can only be solved if the molecular orbitals are expanded in terms of known basis functions, the so-called atomic orbitals, which together make up the basis set. This approach is also known as the Linear Combination of Atomic Orbitals (LCAO) approximation. In periodic systems the wave function can be conveniently expanded in plane-waves. On the other hand, in non-periodic systems Gaussian functions are most commonly used as atomic orbitals which are known as Gauss-type Orbitals (GTOs). While the exact functions for a single electron system would be Slater functions (Slater-type Orbitals (STOs)) the GTOs have considerable computational advantages. According to the Gaussian product rule the product of two Gaussian functions is again a Gaussian.

The HF method determines the best possible solution for a single Slater determinant for a given basis set. The remaining difference to the exact energy is defined as the correlation energy:

$$E_{\mathrm{corr}} = E_{\mathrm{exact}} - E_{\mathrm{HF}}. \tag{2.7}$$

The correlation energy arises from the instantaneous interaction between electrons which is always negative and cannot be captured due to the mean-field approach. Mainly three approaches are used to capture correlation effects: Configuration Interaction (CI), Many-Body Perturbation Theory (MBPT) and Coupled Cluster (CC) Theory.

2.1.2 Density Functional Theory

The beginning of Density Functional Theory (DFT) is marked by the proof of Hohenberg and Kohn [13] that the electron density uniquely defines the ground state energy. They showed that a one-to-one correspondence between the density and the ground state energy and consequently also between the density and the wave function exists. However, the exact form of this functional is not known. Nevertheless, DFT is of high interest because it promises high computational savings. Wave function based approaches depend on $4N$ coordinates — the position and spin of N electrons — while the density is always defined by exactly three spatial coordinates after integrating out the spin, independent of the system size. The developments in DFT have focused foremost on finding increasingly accurate approximations towards the unknown functional. These will be outlined in the following including examples of functionals that have been used in this work. [12]

The electronic energy functional can be partitioned into several terms analogue to the terms that have been introduced for the general wave function theory (Equation 2.2):

$$E_{\mathrm{DFT}}[\rho] = T[\rho] + E_{\mathrm{ne}}[\rho] + J[\rho] + K[\rho]. \tag{2.8}$$

These terms are the kinetic energy $T[\rho]$, the interaction between the nuclei and the electrons $E_{\mathrm{ne}}[\rho]$ and the electron-electron interactions described by the Coulomb term $J[\rho]$ and the exchange term $K[\rho]$ as a function of the density ρ. The nucleus-electron interaction as well as the Coulomb term are given by the classical expressions

$$E_{\mathrm{ne}}[\rho] = -\sum_{a}^{N_{\mathrm{nuclei}}} \int \frac{Z_a(\boldsymbol{R}_a)\rho(\boldsymbol{r})}{|\boldsymbol{R}_a - \boldsymbol{r}|} \mathrm{d}\boldsymbol{r} \tag{2.9}$$

$$J[\rho] = \frac{1}{2} \iint \frac{\rho(\boldsymbol{r})\rho(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} \mathrm{d}\boldsymbol{r}\mathrm{d}\boldsymbol{r}' \tag{2.10}$$

with the nuclear charges Z_α and positions $\boldsymbol{R}_a$.

The first formulation including kinetic and exchange energy was derived for the uniform electron gas and is known as the Thomas-Fermi-Dirac model [14, 15]:

$$T_{\mathrm{TF}}[\rho] = \frac{3}{10} \left(3\pi^2\right)^{\frac{2}{3}} \int \rho^{\frac{5}{3}}(\boldsymbol{r}) \mathrm{d}\boldsymbol{r} \tag{2.11}$$

$$K_{\mathrm{D}}[\rho] = -\frac{3}{4} \left(\frac{3}{\pi}\right)^{\frac{1}{3}} \int \rho^{\frac{4}{3}}(\boldsymbol{r}) \mathrm{d}\boldsymbol{r} \tag{2.12}$$

However, the assumption of a uniform electron gas works only for a few physical systems, e.g. valence electrons of classical metallic systems. Especially covalently bonded systems are described qualitatively wrong by this model because molecules are not stable under these assumptions. The main reason is the inaccurate description of the kinetic energy which is typically underestimated by about 10%.

A major development that made DFT methods applicable to molecular systems was a different approach to the kinetic energy introduced by Kohn and Sham. [16] The kinetic energy can be computed exactly for a system of non-interacting electrons if these are described by orbitals analogue to HF. The density and the kinetic energy are given then as

$$\rho = \sum_{i=1}^{N_{\mathrm{elec}}} |\phi_i|^2 \tag{2.13}$$

$$T_{\mathrm{S}}[\rho] = \sum_{i=1}^{N_{\mathrm{elec}}} \left\langle \phi_i \middle| -\frac{1}{2}\nabla^2 \middle| \phi_i \right\rangle . \tag{2.14}$$

The obvious disadvantage is the introduction of orbitals, which increases the number of spatial coordinates from 3 to $3N$. Furthermore, the description of a non-interacting systems as in HF means that correlation effects of the kinetic energy are missing. These have to be described by an additional term. The missing contributions are generally combined with the exchange contributions to the exchange-correlation term $E_{\mathrm{xc}}[\rho]$. The resulting DFT energy is consequently

$$E_{\mathrm{KS}}[\rho] = T_{\mathrm{S}}[\rho] + E_{\mathrm{ne}}[\rho] + J[\rho] + E_{\mathrm{xc}}[\rho]. \tag{2.15}$$

It remains to find an exchange-correlation functional that describes the kinetic and potential correlation energy as well as the exchange energy. Nevertheless, this formulation of DFT is considerably more robust with regard to the choice of the functional because the exchange-correlation energy is about an order of magnitude smaller than the kinetic energy.

The systems investigated throughout this work are closed shell systems without unpaired electrons. Only the exchange-correlation functional depends on the spin. This is usually formulated as a function of the total density and the spin polarisation ζ

$$\zeta = \frac{\rho_\alpha - \rho_\beta}{\rho_\alpha + \rho_\beta} \tag{2.16}$$

which is the normalised difference density that is again equal to zero for a closed shell system. Furthermore, it is common to define the effective volume with radius r_s that contains a single electron

$$\frac{4}{3}\pi r_s^3 = \rho^{-1} \tag{2.17}$$

and to formulate the exchange and correlation energy as a function of the energy per particle

$$E_{\mathrm{xc}} = \int \rho\epsilon_{\mathrm{x}}[\rho(\boldsymbol{r})]\mathrm{d}\boldsymbol{r} + \int \rho\epsilon_{\mathrm{c}}[\rho(\boldsymbol{r})]\mathrm{d}\boldsymbol{r}. \tag{2.18}$$

The Local Spin Density Approximation (LSDA) is based — similar to the Thomas-Fermi-Dirac model — on the uniform electron gas and assumes therefore that the density is slowly varying. The exchange contribution is therefore equivalent to Equation 2.12 and formulated in terms of the spin polarisation

$$\epsilon_{\mathrm{x}}^{\mathrm{LSDA}} = -C_x f_1(\zeta)\rho^{\frac{1}{3}} \tag{2.19}$$

with the spin polarisation function

$$f_1(\zeta) = \frac{1}{2}\left((1+\zeta)^{\frac{4}{3}} + (1-\zeta)^{\frac{4}{3}}\right). \tag{2.20}$$

Analytical expressions for the correlation energy are only known for the high and low density limits. Fitting functions have been devised which reproduce the known limits and high-level Quantum Monte Carlo results for the intermediate range. One example is the formula from Vosko, Wilk and Nusair (VWN) [17]

$$\epsilon_{\mathrm{c}}^{\mathrm{VWN}}(r_s,\zeta) = \epsilon_c(r_s,0) + \epsilon_a(r_s)\left[\frac{f_2(\zeta)}{f_2''(0)}\right](1-\zeta^4) + \left[\epsilon_c(r_s,1) - \epsilon_c(r_s,0)\right] f_2(\zeta)\zeta^4 \tag{2.21}$$

with

$$\begin{aligned}
f_2(\zeta) &= \frac{f_1(\zeta) - 2}{2^{\frac{1}{3}} - 1} && (2.22)\\
\epsilon_{\mathrm{c}}(x) &= A\left[\ln\frac{x^2}{X(x)} + \frac{2b}{Q}\tan^{-1}\frac{Q}{2x+b}\right. \\
&\quad \left. - \frac{bx_0}{X(x_0)}\left(\ln\frac{(x-x_0)^2}{X(x)} + \frac{2(b+2x_0)}{Q}\tan^{-1}\frac{Q}{2x+b}\right)\right] && (2.23)\\
x &= \sqrt{r_s} && (2.24)\\
X(x) &= x^2 + bx + c && (2.25)\\
Q &= \sqrt{4c - b^2} && (2.26)
\end{aligned}$$

and A, x_0, b and c being the fitting parameters. The main error of LSDA lies for molecular systems in the approximation of the exchange energy which is roughly underestimated by 10%. This error is larger than the total correlation energy which is usually overestimated, for example the strength of bonds is consequently also overestimated.

The Generalised Gradient Approximation (GGA) is the inclusion of the first derivative of the density as a variable. This goes towards a better description of a varying, non-uniform density. This approach is generally formulated as a correction to LSDA, as in this example of the B88 exchange functional [18]:

$$\begin{aligned}
\epsilon_{\mathrm{x}}^{\mathrm{B88}} &= \epsilon_{\mathrm{x}}^{\mathrm{LDA}} + \Delta\epsilon_{\mathrm{x}}^{\mathrm{B88}} && (2.27)\\
\Delta\epsilon_{\mathrm{x}}^{\mathrm{B88}} &= -\beta\rho^{\frac{1}{3}}\frac{x^2}{1 + 6\beta x\sinh^{-1}x} && (2.28)\\
x &= \frac{|\nabla\rho|}{\rho^{\frac{4}{3}}} && (2.29)
\end{aligned}$$

with the fitting parameter β that is determined using experimental data on rare gas atoms. This approach can reduce the error in the exchange energy by two orders of magnitude. Equivalent improvements can be devised for the correlation functional and one example which is often used in combination with the B88 exchange [19] has been formulated by Lee, Yang and Parr (LYP) [20]:

$$\begin{aligned}
\epsilon_c^{\mathrm{LYP}} = -4a&\frac{\rho_\alpha\rho_\beta}{\rho^2\left(1+d\rho^{-\frac{1}{3}}\right)} - \\
&ab\omega\left\{\frac{\rho_\alpha\rho_\beta}{18}\left[144\cdot 2^{\frac{1}{3}}C_F\left(\rho_\alpha^{\frac{8}{3}}+\rho_\beta^{\frac{8}{3}}\right)+(47-7\delta)|\nabla\rho|^2-\right.\right. \\
&(45-\delta)\left(|\nabla\rho_\alpha|^2+|\nabla\rho_\beta|^2\right)+2\rho^{-1}(11-\delta)\left.\left(\rho_\alpha|\nabla\rho_\alpha|^2+\rho_\beta|\nabla\rho_\beta|^2\right)\right] \\
&\left.+\frac{2}{3}\rho^2\left(|\nabla\rho_\alpha|^2+|\nabla\rho_\beta|^2-|\nabla\rho|^2\right)-\left(\rho_\alpha^2|\nabla\rho_\alpha|^2+\rho_\beta^2|\nabla\rho_\beta|^2\right)\right\} \qquad (2.30)
\end{aligned}$$

$$\omega = \frac{e^{-c\rho^{-\frac{1}{3}}}}{\rho^{\frac{14}{3}}\left(1+d\rho^{-\frac{1}{3}}\right)} \qquad (2.31)$$

$$\delta = c\rho^{-\frac{1}{3}} + \frac{d\rho^{-\frac{1}{3}}}{1+\rho^{-\frac{1}{3}}} \qquad (2.32)$$

with the parameters a, b, c and d that are fitted against exact numerical data from the helium atom.

Further improvements can be achieved by including higher-order derivatives into the functional leading to so-called meta-GGA functionals. However, the more common next step is the inclusion of HF exchange into the functional. The motivation is that if the Kohn-Sham orbitals would be identical to the HF orbitals the exact exchange for the non-interacting system would be the exchange computed by the HF method. One improvement lies in the reduction of the self-interaction error which is equal to zero in HF. However, the precise amount of exact exchange that should be included is not known, mostly system dependent and generally fitted against experimental data. One prominent example is the B3LYP functional [17, 20–22] which is defined by three parameters a, b, c combining the previously shown B88 exchange and the LYP correlation functional with a fraction of exact exchange:

$$\begin{aligned}
E_{\mathrm{xc}}^{\mathrm{B3LYP}} &= E_{\mathrm{x}}^{\mathrm{LSDA}} + a(E_{\mathrm{x}}^{\mathrm{exact}} - E_{\mathrm{x}}^{\mathrm{LSDA}}) + b(E_{\mathrm{x}}^{\mathrm{B88}} - E_{\mathrm{x}}^{\mathrm{LSDA}}) \\
&+ E_{\mathrm{c}}^{\mathrm{VWN}} + c(E_{\mathrm{c}}^{\mathrm{LYP}} - E_{\mathrm{c}}^{\mathrm{VWN}}). \qquad (2.33)
\end{aligned}$$

The computation of the DFT energy requires a procedure very similar to the HF method. Canonical Kohn-Sham orbitals are generated which are expanded in terms of atomic orbitals and an analogue of the Fock matrix is built allowing the computation of the one-electron and the Coulomb terms. However, the exchange-correlation contribution is defined in terms of the density and derivatives thereof and depends implicitly on the integration variable itself. Accordingly, these integrals cannot be solved analytically but have to be computed by means of a numeric integration. Generally, this requires the integration of a function $F(\boldsymbol{r})$ over all space

$$I = \int F(\boldsymbol{r})\mathrm{d}^3\boldsymbol{r} \approx \sum_i A_i F(\boldsymbol{r}_i) \qquad (2.34)$$

which is exact for an infinite number of grid points. However, in practice one tries to find the smallest possible number of grid points for a given accuracy. Generally, more points are required in regions of a strongly varying density for example near the nuclei and fewer in regions of a more uniform density. Some of the ideas to construct these grids for the numerical integration are illustrated in the following Section.

2.1.3 Numerical Integration in DFT

The basic idea of the following approach is that any molecular function $F(\boldsymbol{r})$ can be partitioned into a sum of contributions due to the nuclei n [23]

$$F(\boldsymbol{r}) = \sum_n F_n(\boldsymbol{r}). \tag{2.35}$$

A weighting function is assigned to each nucleus which is equal to unity close to itself and vanishes near all the other nuclei with the property

$$\sum_n \omega_n(\boldsymbol{r}) = 1. \tag{2.36}$$

Thus, the space is not partitioned into strictly separated, distinct cells but into overlapping, fuzzy and continuous cells. This allows to simplify the multi-centre integration to a simpler single-centre integration I_n

$$\begin{aligned} F_n(\boldsymbol{r}) &= \omega_n(\boldsymbol{r}) F(\boldsymbol{r}) && (2.37) \\ I &= \sum_n I_n && (2.38) \\ I_n &= \int F_n(\boldsymbol{r}) \mathrm{d}^3\boldsymbol{r}. && (2.39) \end{aligned}$$

An appropriately chosen weighting function allows using conventional single-center integration approaches in polar coordinates.

The space is separated in Voronoi polyhedra, [24] that means each nucleus i is surrounded by a polyhedron which contains all points in space that are closer to i than to any other nucleus. For periodic atomic lattices these are known as Wigner-Seitz cells. [25] Confocal elliptical coordinates which depend explicitly on the distance between two nuclei r_{ij} allow a simple definition of these polyhedra

$$\begin{aligned} \mu_{ij} =& \frac{r_i - r_j}{r_{ij}}, && (2.40) \\ \lambda_{ij} =& \frac{r_i + r_j}{r_{ij}} && (2.41) \end{aligned}$$

with r_i and r_j being the distances to the nuclei. The third coordinate is the angle about the internuclear axis ϕ_{ij}. For example $\mu_{ij} = 0$ corresponds to the plane in the middle between the two nuclei with the vector connecting the two nuclei being the normal vector of this plane. The ranges for these coordinates are defined as

$$0 \leq \phi_{ij} \leq 2\pi, \tag{2.42}$$

$$1 \leq \lambda_{ij} \leq \infty, \tag{2.43}$$

$$-1 \leq \mu_{ij} \leq 1. \tag{2.44}$$

The definition of a step function $s(\mu_{ij})$

$$s(\mu_{ij}) = \begin{cases} 1, & -1 \leq \mu_{ij} \leq 0 \\ 0, & 0 < \mu_{ij} \leq 1 \end{cases} \tag{2.45}$$

allows to define the polyhedron on a given nucleus i as

$$P_i(\boldsymbol{r}) = \prod_{j \neq i} s(\mu_{ij}) \tag{2.46}$$

which is equal to unity for any given point inside of the polyhedron and equal to zero outside.

In order to generalise the strictly separated polyhedra to fuzzy overlapping ones, the step function is replaced by a continuous analogue s_c which fulfils the following conditions

$$s(-1) = 1, \tag{2.47}$$

$$s(+1) = 0, \tag{2.48}$$

$$\left(\frac{\mathrm{d}s}{\mathrm{d}\mu}\right)_{\mu=\pm 1} = 0. \tag{2.49}$$

As stated before they are unity close to one nucleus and vanish at all other nuclei. Furthermore they are continuous and do not have any cusp at the position of the nuclei. However, this does not define a unique function. A simple polynomial $\rho(\mu_{ij})$ of two terms has been found to work well and fulfils the conditions:

$$\rho(\mu_{ij}) = \frac{3}{2}\mu_{ij} - \frac{1}{2}\mu_{ij}^3. \tag{2.50}$$

Repeatedly applying this function allows to go from very fuzzy overlapping cells towards the discrete cells until the original step function is recovered. It has been found that three times works reasonably well for general applications which leads to the final expression:

$$s_c(\mu) = \frac{1}{2}\left[1 - \rho\left(\rho\left(\rho\left(\mu\right)\right)\right)\right] \tag{2.51}$$

and allows the definition of the weights as

$$\omega_n(\boldsymbol{r}) = \frac{\prod\limits_{i \neq n} s_c(\mu_{in})}{\sum\limits_m \prod\limits_{i \neq m} s_c(\mu_{im})}. \tag{2.52}$$

For most molecules different elements are present and the polyhedra should reflect the corresponding element to some degree. Bragg-Slater radii can be used to adjust the volume to the corresponding element and allow a polyhedron to occupy correspondingly a smaller or larger space. At this point the multi-centre integration has been simplified to the single-centre integration which will be shown in the following text.

The single-centre integrals are partitioned by means of a product ansatz into the integration over the distance and the angles as defined in a polar coordinate system

$$I_n = \iiint F_n(r,\theta,\phi) r^2 \sin\theta \mathrm{d}r \mathrm{d}\theta \mathrm{d}\phi. \tag{2.53}$$

Grids and weights for the Gauss-Markov quadrature over the surface of a unit sphere have been given by Lebedev. [26] The angles can be integrated separately but it turned out to be more efficient to integrate for both angles combined over the surface. Lebedev devised a quadrature formula invariant under the octahedron group and under inversion G_8^* which defines a number of points on the surface that are equivalent:

	Points	Coordinates
a_i^1	6	$(0,0,\pm 1), (0,\pm 1,0), (\pm 1,0,0)$
a_i^2	12	$\frac{1}{\sqrt{2}}(\pm 1,\pm 1,0), \frac{1}{\sqrt{2}}(\pm 1,0,\pm 1), \frac{1}{\sqrt{2}}(0,\pm 1,\pm 1)$
a_i^3	8	$\frac{1}{\sqrt{3}}(\pm 1,\pm 1,\pm 1)$
b_i^k	24	$(\pm l_k,\pm l_k,\pm m_k), (\pm l_k,\pm m_k,\pm l_k), (\pm m_k,\pm l_k,\pm l_k)$ with $2l_k^2 + m_k^2 = 1$
c_i^k	24	$(\pm p_k,\pm q_k,0), (\pm p_k,0,\pm q_k), (0,\pm p_k,\pm q_k)$ with $p_k^2 + q_k^2 = 1.$

This leads to the quadrature formula with the grid points a_i^j, b_i^k, c_i^1 and the corresponding weights A_i, B_k, C_1

$$\begin{aligned} I_n(f) \approx A_1 \sum_{i=1}^{6} f(a_i^1) + A_2 \sum_{i=1}^{12} f(a_i^2) + A_3 \sum_{i=1}^{8} f(a_i^3) \\ + \sum_{k=1}^{N_1} B_k \sum_{i=1}^{24} f(b_i^k) + C_1 \sum_{i=1}^{24} f(c_i^1) \end{aligned} \tag{2.54}$$

with a total of $N = 26 + 24N_1$ grid points with $N_1 \leq 3$. The weights are determined by guaranteeing the exact integration on the surface of all polynomials up to order n. It can be shown that only the polynomials invariant under group G_8^* need to be considered and the resulting linear equations have been solved and tabulated up to $n = 131$. For order 11, 17, and 23 the resulting number of grid points with non-zero weights are 50, 110 and 194 respectively. Relatively large angular grids are required to integrate accurately the step-like feature in the internuclear regions and about 100–200 grid points achieve an accuracy of about 5–6 significant digits in

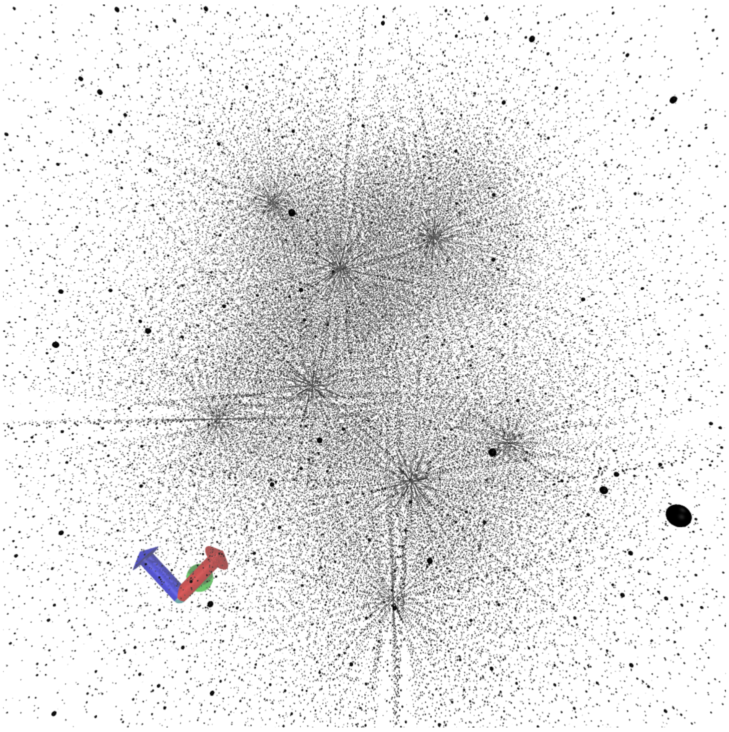


FIGURE 2.1 Grid for an ethanol structure according to Lebedev for the angular integration and Mura and Knowles for the radial integration.

the integration.

The radial integration is carried out by Chebyshev-Gauss quadrature of second order:

$$\int_{-1}^{+1} \sqrt{1-x^2} g(x)\, dx \approx \sum_{i=1}^{n} \frac{\pi}{n+1} \sin^2\left(\frac{i}{n+1}\pi\right) g\left(\cos\left(\frac{i}{n+1}\pi\right)\right). \tag{2.55}$$

This has the advantage over the widely used Gauss-Legendre quadrature that analytical formulas are known for the Chebyshev weights and points. The integration over the range -1 to 1 is transformed onto the range 0 to ∞. Different mappings have been proposed including the formula by Becke:

$$r = r_m \frac{1+x}{1-x} \tag{2.56}$$

with r_m corresponding to the midpoint of the integration interval $x = 0$. This parameter is chosen as half the Bragg radius in order to represent a meaningful physical scale for the distribution of the radial grid. Hydrogen is an exception where the Bragg radius itself has been used without applying the factor $\frac{1}{2}$. A different mapping proposed by Mura and Knowles [27] is

$$r = -\alpha \log_e (1 - x^m) \tag{2.57}$$

where α and m can be choosen freely. This mapping exhibits a balanced representation of the nuclei, bonded and long-range regions. The performance of these grids is revisited in the context of the perturbative Monte Carlo method also with regard to the number of grid points in Section 4.1. An example of a typical grid for ethanol is shown in Figure 2.1. The different number of points in nuclear, inter-nuclear and long-range regions can be easily distinguished.

2.2 Molecular Mechanics

The solution of the Schrödinger equation (Equation 2.1) is computationally very expensive which arises mainly from the large number of degrees of freedom due to the electrons. One approach to reduce the degrees of freedom has been illustrated with DFT (Section 2.1.2). However, the electrons react instantaneously to the change of the positions of the nuclei as stated by the Born-Oppenheimer approximation. This means they can be treated separately. The potential energy can be described as a parametric function of the coordinates of the nuclei. This function is called a force field which attempts to describe the QM potential energy surface.

Force fields are parametrised for atoms or groups of atoms with similar properties which is based on the chemically intuitive concept of functional groups. Such a group denotes a small number of atoms that exhibits similar functionality in different molecules, e.g. a carboxyl or phenol group. This suggests that a transferable function can be derived to describe a functional group in the context of various molecules. One example are carbonyl groups where the bond length of C-O double bonds are always around 1.2 Å, the frequencies are usually about 1700 cm^{-1} and the carbon is always found in a planar geometry. Additionally, the heat of formations of linear alkanes can be estimated based only on the chain length suggesting that all CH_2 groups contribute a constant value to the energy. [28]

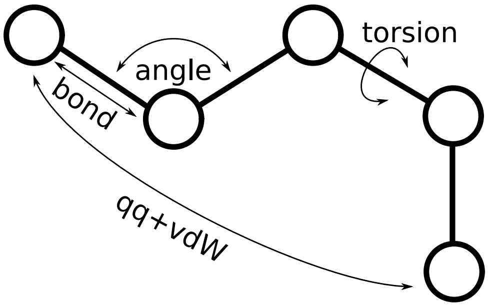


FIGURE 2.2 Schematic representation of the different terms in MM methods.

Force fields describe molecules with a ”ball and spring” model. Most commonly, the potential energy is separated in bonded terms which include contributions from bonds, angles and torsions and non-bonded terms for the electrostatic and van der Waals interactions:

$$E_{\mathrm{MM}} = E_{\mathrm{bond}} + E_{\mathrm{angle}} + E_{\mathrm{torsion}} + E_{\mathrm{qq}} + E_{\mathrm{vdW}} \tag{2.58}$$

which are illustrated in Figure 2.2. Information about the connectivities have to be specified as they do not emerge naturally as in the case of the Schrödinger equation where electrons are described explicitly. Furthermore, the type of atoms has to be given in order to distinguish e.g. a carbon in an alkane group or in a carbonyl group. Many different functions have been proposed for these terms and parameters can be derived from high-level QM methods or fitted in order to reproduce experimental results. Compared to QM methods the computational costs are reduced by several orders of magnitude. However, deriving parameters for a force field is far from trivial. If not noted differently explicit formulas are given in the following for the

Optimized Potentials for Liquid Simulations - All Atoms (OPLS-AA) force field. [29–31] The latter has been used throughout this work.

2.2.1 Bonded Terms

The bonded and angle terms are described both by a simple harmonic potential, which can be understood in terms of a Taylor expansion that is truncated after the first term:

$$E_{\text{bond}} = \sum_{\text{bonds}} K_r (r - r_{\text{ref}})^2 \tag{2.59}$$

$$E_{\text{angle}} = \sum_{\text{angles}} K_\Theta (\Theta - \Theta_{\text{ref}})^2 \tag{2.60}$$

with the force constants K which specify the stiffness of the bond and the angle. The natural bond length r_{ref} and angle Θ_{ref} are in the simplest case equivalent to the equilibrium value of the bond or angle. However, more generally speaking it is the value which reproduces the experimental equilibrium bond length or angle for the minimum energy geometry. In larger molecules the equilibrium values deviate slightly from the natural values due to further terms and e.g. bonds are generally longer. This simple description as it is used in the OPLS-AA force field has the obvious shortcoming that reactions including changes of the bonding situation cannot be described. The bonds are specified explicitly and the harmonic potential has the wrong limiting behaviour for the dissociation. Including further terms of the Taylor expansion improves the accuracy around the equilibrium position but does not describe the dissociation correctly. One way forward is to choose a different function form like the Morse potential

$$E_{\text{Morse}} = D\left(1 - \exp(-\alpha(r - r_{\text{ref}}))\right)^2 \tag{2.61}$$

which reproduces the correct dissociation energy and accurate energies around the equilibrium position. However, for simulations at standard conditions mostly the range up to 40 kJ/mol is accessible which are well described even by a harmonic potential with lower computational costs.

The torsional potential which describes the change of the energy associated with the rotation around a bond is very different from the bond or angle term. The potential can have multiple minima which are separated by relatively small barriers. Consequently, a large range of angles is accessible without a clear equilibrium value. Moreover, the potential is periodic and should give the same energy after one period. These criteria are fulfilled by a Fourier series:

$$E_{\text{torsion}} = \sum_{\text{torsion}} \sum_{i=1}^{3} \frac{V_i}{2} \left(1 + (-1)^{i+1} cos\left(i\phi + \phi_i\right)\right). \tag{2.62}$$

with the different terms for i representing 360°, 180°,… periodicity, V_i specifies the barriers for the conversion between the minima and ϕ_i shifts the position of these. Some of the V_i can be equal to zero e.g. for the rotation around the C-C bond in ethane only the V_3 term is required, which gives rise to three equivalent minima and maxima. How the combination of all three

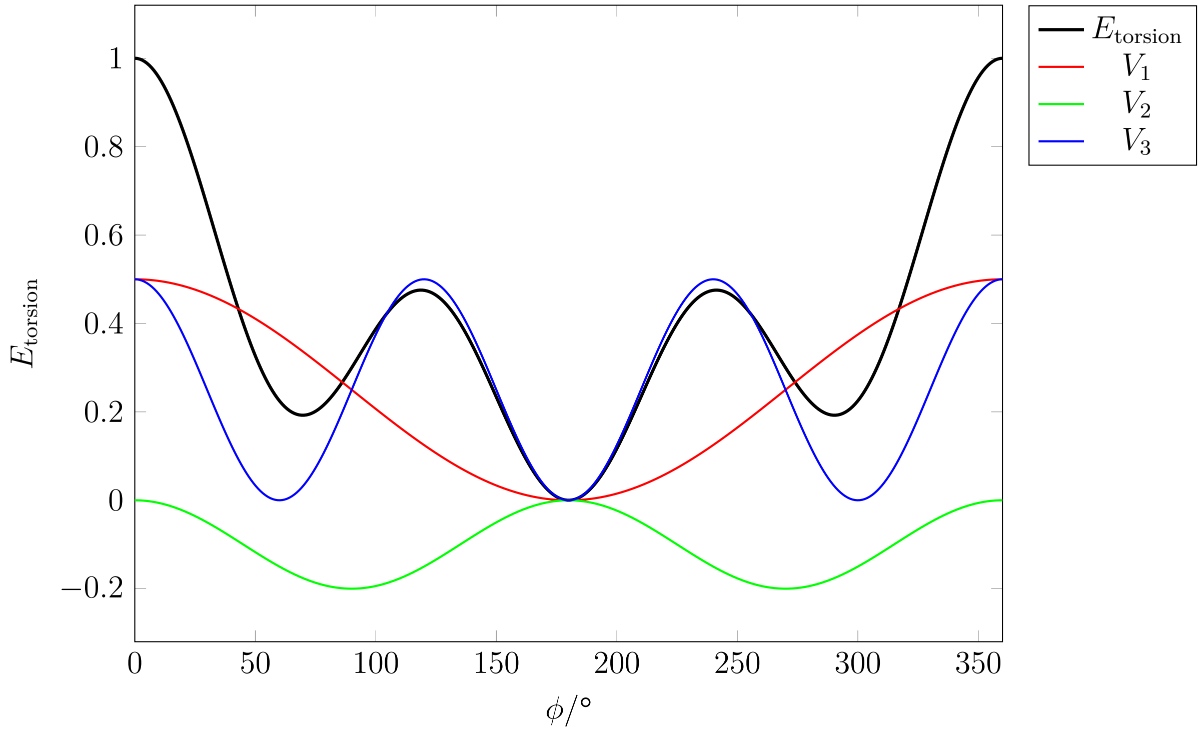


FIGURE 2.3 Rotation potential corresponding to Equation 2.62 with $V_1 = 0.5$, $V_2 = -0.2$ and $V_3 = -0.5$ and all $\phi_i = 0$.

terms gives rise to a more complex torsional potential is shown in Figure 2.3.

2.2.2 Non-Bonded Terms

The van der Waals term — one of the non-bonded expressions — describes the short range repulsion as well as the intermediate range attraction between atoms which is neither connected with bonded nor electrostatic interactions. For example the dispersion interaction gives rise to the attractive feature of the potential together with higher order electron correlation effects. However, the induced dipole-dipole interactions are the dominating contribution which therefore determine the R^{-6} decay of the van der Waals interaction for large distances. A typical example is the interaction between non-polar molecules such as alkanes or rare gas atoms. The repulsive part for short distances is due to the overlap of the electron densities which are repelling each other. Therefore, an approximately exponential decay of this repulsion should be expected. However, mostly due to practical considerations a common function is the Lennard-Jones potential

$$E_{vdw} = 4 \sum_{a>b} f_{ab} \epsilon_{ab} \left[\left(\frac{\sigma_{ab}}{r_{ab}} \right)^{12} - \left(\frac{\sigma_{ab}}{r_{ab}} \right)^{6} \right] \tag{2.63}$$

with σ_{ab} describing the beginning of the attractive part and ϵ_{ab} the depth of the well. Both parameters are computed as the geometric mean from atomic parameters. This reduces the number of required parameters with only two for each atom type. The sum includes all atoms that are separated by more than three bonds and f_{ab} is always equal to one with the exception of the 1,4-interactions which are scaled by $f_{ab} = 0.5$ in OPLS. The 1,4-interacting atoms are also interacting through the bonded torsion potential and therefore the van der Waals interaction is scaled down. The computational advantage of the Lennard-Jones potential is that it is very cheap because only the square of the distances is needed — not the distance itself — and the r^{-12} is just the square of the r^{-6} term and can be computed very efficiently. It should be noted that the van der Waals terms are parametrised against experimental data which gives rise to an effective two-body potential which includes many body effects in an average way. This model assumes an isotropic density of the atoms. Two cases where this is only approximately correct are hydrogen atoms, the single electron involved in the only bond is always displaced towards the neighbouring atom, and atoms with lone pairs, e.g. nitrogen or oxygen. However, many models do not take into account this effect and compensate for it through the treatment of the electrostatic interaction which, however, has the wrong distance dependence.

The missing half of the non-bonded interactions are the electrostatic interactions. The distribution of the electrons inside of a molecule leads to positively and negatively charged regions. These are especially important for polar compounds. The simplest approach to model the electrostatic interaction is to assign partial charges to the atoms which interact through the Coulomb potential

$$E_{qq} = \sum_{a>b} f_{ab} \frac{q_a q_b e^2}{r_{ab}} \tag{2.64}$$

again with the scaling factor $f_{ab} = 0.5$ in OPLS-AA for 1,4-interactions. Some force fields use bond dipoles instead which give in most cases equivalent results. The partial charges can be determined from QM calculations. Mostly, they are fit in order to reproduce the exact electrostatic potential near the molecule, e.g. on the van der Waals surface. Furthermore, the partial charges can be simultaneously fitted against experimental data like free solvation energies in order to construct an effective two-body potential that incorporates again many-body effects in an average way. A prominent example is water with a dipole moment of 2.5 Debye in condensed phase compared to 1.8 Debye in gas phase. Fitting against experimental data improves the accuracy considerably in this case. Alternatively, QM methods that are known to overestimate the polarisation like Hartree-Fock can be used.

The description of the electrostatic interaction solely by partial charges can lead to considerable errors because the electrostatic potential is not accurately represented. The inclusion of higher-order electric moments or of additional charges not associated with an atom can improve the accuracy and allows an anisotropic description of atoms. Nonetheless, the inclusion up to quadrupoles increases the computational costs by nearly one order of magnitude. However, if many-body effects are important atomic polarisabilities have to be taken into account explicitly. The first contribution comes from the dipole $\boldsymbol{\mu}_{\text{ind}}$ induced by the electric field $\boldsymbol{F}$ which is created by electric moments on other sites

$$\boldsymbol{\mu}_{\text{ind}} = \boldsymbol{\alpha} \boldsymbol{F} \tag{2.65}$$

with the polarisibility of the site itself. This gives rise to the polarisation contribution to the electrostatic interaction

$$E_{\text{pol}} = \frac{1}{2} \mu_{\text{ind}} \boldsymbol{F}. \tag{2.66}$$

Every induced dipole in turn contributes to the electric field and influences the induced dipoles on other sites. Hence, an iterative procedure is required until the induced dipoles are solved in a self-consistent way. This increases the computational costs at least by a factor of about two depending on the number of iterations needed until convergence.

Finally, these terms (Equation 2.58) may be coupled by so called cross terms. For instance the bond lengths which are included in the definition of an angle give rise to a stretch-bend term. In order to avoid a huge number of parameters these terms are either independent of the involved atoms or take into account a single central atom type. A closely related correction is for example the dependence of the natural bond length on the electronegativity of the involved atoms which is comparable to a stretch-electrostatic cross term.

2.2.3 Long-Range Electrostatic Interactions

The description of liquids requires an approach very similar to the one of crystals. Many molecules are needed — on the order of hundreds to thousands — to describe accurately solvent effects. However, to avoid surface effects and molecules evaporating into the surrounding vacuum Periodic Boundary Conditions (PBC) have to be employed. The condensed system is constructed inside of one of the five space filling polyhedra, most commonly a cube, and then duplicated in all directions. Therefore, the model is quasi-periodic and if a molecule leaves the central box through one wall its image enters through the opposite wall. A schematic two-dimensional representation is shown in Figure 2.4.

In the terminology of the minimum image convention the central box is called the original simulation cell and all its copies are images. Only particles and properties of the original have to be recorded if each particle always interacts with its closest image of the other particles of the original box. The distance in the minimum image convention $\tilde{r}$ can be obtained from the distance r of any two particles for a cubic system with edge length L according to

$$\tilde{r} = r - \mathrm{int}\left(\frac{r}{L}\right) L. \tag{2.67}$$

The same formula can be applied to obtain the Cartesian coordinates of the particle in the original box as long as all coordinates are positive which results in the original box ranging from 0 to L along all three axes.

The evaluation of non-bonded terms for an infinite system can be rather challenging. The van der Waals interactions with their distance dependence of r^{-6} are rather short-ranged and become negligible after a distance of about 10 Å. This allows to introduce a cut-off so that only interactions between particles below a given distance are taken into account. This can be equivalently seen as the interaction of a given particle with all other particles inside a sphere with the radius of the cut-off r_c as it is visualised in Figure 2.4. A cut-off radius larger than half of the box length should be avoided in the minimum image convention. However, the Coulomb interactions between partial charges are very long-ranged and a simple cut-off leads to sizeable discontinuities. Two different approaches are well-established to treat the electrostatic interaction correctly under consideration of PBC which will be discussed in the following.

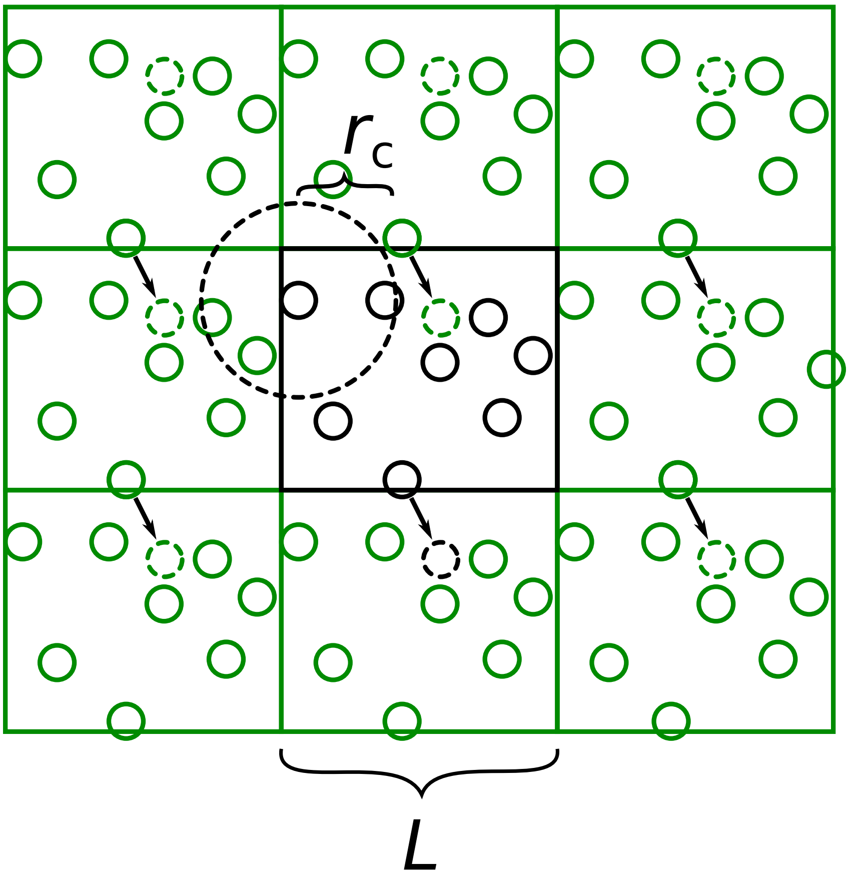


FIGURE 2.4 Schematic representation of PBC in two dimensions for a square. The original cell and particles in black and the images in green. For one of the particles a spherical cut-off is shown.

The first one is the Ewald summation [32] which divides the long-range interaction in two contributions. The short-range contribution is computed in real space while the long-range one is evaluated using a Fourier transform in reciprocal space. Both summations converge quickly in their respective spaces and can be truncated while retaining high accuracy. A number of difficulties arise with this approach. The Ewald summation formally requires a neutral system and introduces for charged systems a uniform neutralising background charge which, however, deviates from the actual charge distribution. [33] Furthermore, it has been shown that the free energy landscapes of proteins is altered because of the artificially introduced periodicity. A single protein is restricted to interact with its images in exactly the same orientation and secondary structure. This can even reverse the population of minima on the free-energy landscape as shown for a dialanine model system. [34] Another study showed that the α-helical configuration is artificially stabilised and the unfolding processes are hindered due to the Ewald summation. [35] Three factors have been identified to increase the artefacts due to the periodicity which is a charged solute, a solvent with low dielectric constant and a solute with a size that is non-negligible compare to the cell size. [36]

Pairwise alternatives [37–39] have been proposed which avoid the explicit periodicity altogether and just use a summation in real space. While this approach seems to be simple minded at first sight, it is based first on the realisation that the effective scaling of the electrostatic interaction with the distance is considerable lower than r^{-1} in condensed phase. Second, that the Ewald summation has been used historically for very small systems with correspondingly short cut-off for the real space term in order to make calculations computationally feasible. With nowadays computer power large systems of 40 Å edge length can be routinely simulated which allows using cut-offs up to 20 Å.

Truncating the summation with a spherical cut-off leads generally to a volume with a total net charge different from zero. It has been realised that this is the main cause of the poor convergence and unsystematic behaviour of the direct summation. A careful grouping of ions or atoms in neutral groups which are either completely included or excluded by the cut-off leading to a net charge equal to zero cures this erratic behaviour and gives results equivalent to the Ewald summation and converging quickly with the distance. For example the Madelung constant of an ion in a perfect crystal has been shown to depend with r^{-5} on the distance of other ions. [37] However, sorting of ions into neutral groups is computationally expensive and impractical in highly disordered systems like melts or liquids.

An alternative approach is to neutralise the spherical volume by projecting for every ion a charge with the opposite sign onto the surface of the sphere. This guarantees that the central ion interacts only with neutral pairs — the actual and the projected charge. Careful derivation lead to the realisation that the physical concept of charge neutralisation at the surface is indeed equivalent to the concept of using a shifted operator for the Coulomb interaction. The simplest formulation is the shifted potential V_{SP} which is 0 at the cut-off

$$V_{SP}(r) = \begin{cases} v(r) - v_c & r \leq r_c \\ 0 & r > r_c \end{cases} \tag{2.68}$$

with the force

$$F_{SP}(r) = \begin{cases} \frac{dv(r)}{dr} & r \leq r_c \\ 0 & r > r_c \end{cases} \tag{2.69}$$

Notwithstanding a smooth potential, the force is not continuous at the cut-off. This leads to problems especially in Molecular Dynamics (MD) simulations which rely on the gradients and leads to an energy drift during the simulation. This can be addressed by deriving a shifted force potential V_{SF} which fulfils the requirement

$$\left.\frac{dV_{SF}}{dr}\right|_{r=r_c} = 0. \tag{2.70}$$

Furthermore, the electrostatic interaction oscillates with increasing cut-off and therefore converges only slowly against the correct value. The introduction of a damping parameter leads to accelerated convergence and allows using a smaller cut-off. For the simple Coulomb potential

$$v(r) = \frac{q_i q_j}{r} \tag{2.71}$$

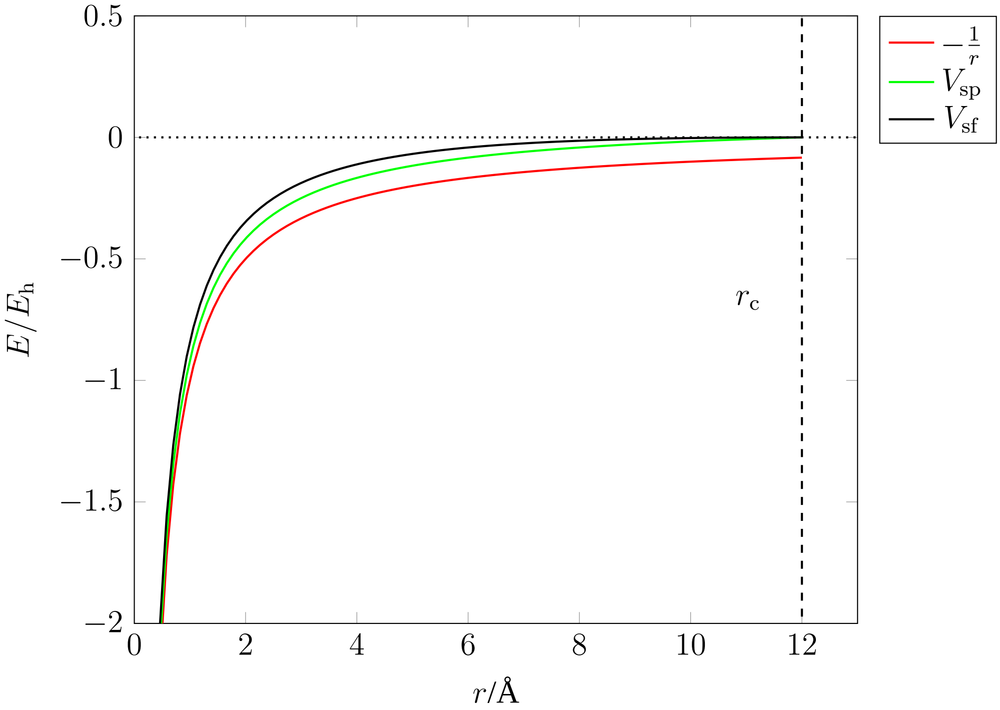


FIGURE 2.5 Coulomb interaction of two unit charges with opposing signs for the Coulomb, a shifted and a shifted force potential, both with $\alpha = 0$.

this leads to the final expressions for $r \leq r_c$ of the damped shifted force potential

$$V_{DSF}(r) = q_i q_j \left[\frac{\mathrm{erfc}(\alpha r)}{r} - \frac{\mathrm{erfc}(\alpha r_c)}{r_c} + \left(\frac{\mathrm{erfc}(\alpha r_c)}{r_c^2} + \frac{2\alpha}{\pi^{1/2}} \frac{\exp(-\alpha^2 r_c^2)}{r_c} (r - r_c) \right) \right] \tag{2.72}$$

and accordingly the forces

$$F_{DSF}(r) = q_i q_j \left[\left(\frac{\mathrm{erfc}(\alpha r)}{r^2} + \frac{2\alpha}{\pi^{1/2}} \frac{\mathrm{erfc}(-\alpha^2 r^2)}{r} \right) - \left(\frac{\mathrm{erfc}(\alpha r_c)}{r_c^2} + \frac{2\alpha}{\pi^{1/2}} \frac{\exp(-\alpha^2 r_c^2)}{r_c} \right) \right] \tag{2.73}$$

with the damping constant α and the complementary error function

$$\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty \exp(-t^2) \mathrm{d}t. \tag{2.74}$$

The accuracy of this approach has been investigated by comparing energies, forces, torques, velocity autocorrelation functions and the derived power spectra with reference simulations using the Ewald summation. Equivalent results have been obtained in all cases for $r_c \geq 12$ Å and with damping. For Metropolis Monte Carlo (MC) simulations which are only based on the energies and do not need forces it is valid to use $\alpha = 0$. This saves computational time because the expensive evaluation of the error function is not required. These approaches have been extended by enforcing that higher derivatives are as well equal to 0 at the cut-off. [39] This might be necessary for MD simulations in extremely ionic systems as shown in a study of ionic liquids. In that study 59 ionic liquid combinations of six cations and seven anions have been simulated. A correlation coefficient (R^2 value) in comparison with results form the Ewald summation of 0.99 has been found for the most accurate shifted operator. [40] It is also possible to additionally integrate group based cut-offs into these methods which increases the accuracy even further.

I conclude that pairwise direct summation methods are a viable alternative to the Ewald summation with high accuracy, linear scaling with the systems size, low computational costs and trivial parallelisability. Furthermore, it avoids non-physical artefacts introduced by the enforced periodicity of the Ewald summation and makes it suitable for systems like solid-liquid interfaces or membranes.

2.3 Solvation Models

Many QM studies are carried out in vacuum. However, most biochemical processes take place in aqueous solution. Any chemical reaction inside of cells i.e. the metabolism is embedded in water which may even participate in these reactions. Another example are enzymes which are often embedded in membranes and closely interact with them. The folding of enzymes and thus also their activity depends as well on the environment. Therefore, there is a large interest in computing the effect of the solvent in the context of QM calculations. Different effects occur upon inserting a solute into the solvent. The solvent is polarised by the solute which in turn changes the charge distribution of the solute. The electric moments, most dominantly the dipole moments, of the solvent orient with regard to the solutes electric moments. These two effects which are long-range in nature give rise to the screening of electrostatic interactions which is quantified by the dielectric constant ϵ. For example, the interaction of two charges in water is reduced by about 80, nearly two orders of magnitude. Further solvent effects are short-ranged and arise from the explicit molecular structure of the solvent molecules. This gives rise to a arrangement of solvent molecules in a shell-like structure around the solute. Specific interactions e.g. hydrogen bonds are very important in aqueous solution. Furthermore, solute and solvent do not only interact via electrostatics but also van der Waals interactions are important. Finally, charge transfer effects go beyond polarisation effects. Solvent models have been developed which focus on an accurate description of some of these effects while approximating others which gives rise to a number of different approaches. These will be discussed in the following Sections.

2.3.1 Continuum Solvation Methods

Continuum solvation models stem from the macroscopic description of the solvent as a continuum characterised by the dielectric constant ϵ. A cavity is formed inside this continuum — which requires the energy G_{cav} to form — and the solute is placed inside it (Figure 2.6). Van der Waals interactions with the solvent are generally stabilising and among these the dispersion interactions G_{disp} are dominating while for short distances the repulsion G_{rep} becomes more important due to the overlap of the wave functions. Finally, the polarisation of the solvent and in turn the polarisation of the solute results in an electrostatic stabilisation G_{el}. Consequently,

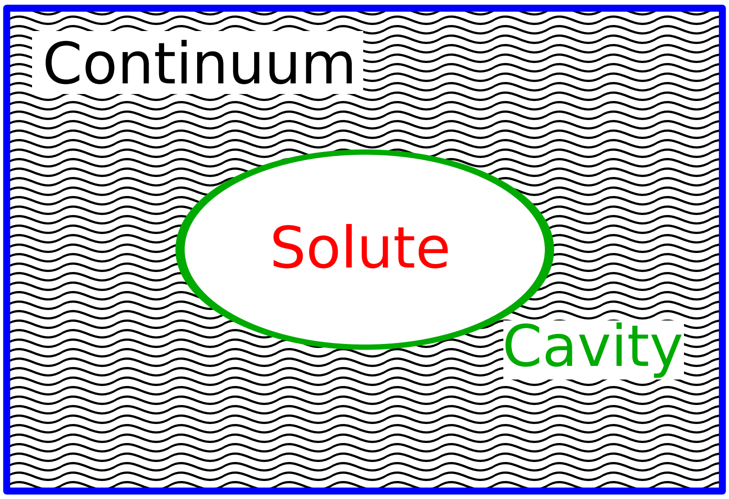


FIGURE 2.6 Schematic representation of the solute (red) inside of a cavity (green) in the continuous solvent (black waves).

the free energy of the solute can be formulated as

$$G = G_{\mathrm{cav}} + G_{\mathrm{disp}} + G_{\mathrm{rep}} + G_{\mathrm{el}} + G_{\mathrm{tm}} \tag{2.75}$$

with an additional term for the thermal motion G_{tm} in order to obtain absolute free energies.

The first step is the definition of the cavity since there is no unique correct way to define a cavity. A simple spherical or ellipsoidal one greatly simplifies the computation of further effects but it turns out that a molecular shaped cavity is required for accurate results. A computationally efficient approach is to use a scaled van der Waals surface. However, cavities embedded inside of the solute can be filled in a non-physical way with small parts of the continuum. A better approach is therefore to use the solvent accessible surface (SAS) which can be constructed by running a probe sphere on the surface of the solute. Nevertheless, the results are sensitive to the radius of this probe sphere or the factor used for scaling the van der Waals radii and they are commonly fitted in order to reproduce experimental results. Furthermore, even the SAS can lead to discontinuities in the case of geometry optimisations if suddenly a cavity becomes accessible. A third possibility is to construct a cavity based on an isodensity surface, a typical value is 0.001 of the electron density.

The energy for the creation of the cavity G_{cav} as well as interactions which are specific for the first solvation shell cannot be modelled within the framework of continuum models. Either they are neglected altogether (Conductor-like Screening Model (COSMO) [41]) or they are parametrised as a simple function of the surface of the cavity (Polarizable Continuum Model (PCM) [42] or Solvent Model Density (SMD) [43]). Dispersion contributions are generally treated together with the cavity term. More sophisticated models use a force field like approach with atom specific parameters.

Clearly, the explicit structure of the first solvation shell is inadequately represented. Therefore, specific interactions e.g. hydrogen bonds or terminally carbonyl groups in solvents like acetone cannot be described. These interactions are by definition not in accordance with a continuum description. However, it has been argued that the combination of electrostatic, dispersion and repulsion terms used in models like PCM can account even for specific interactions. [44, 45] In a hybrid fashion the explicit representation of a few solvent molecules can be combined with the continuum description. However, in the case of a parametrised dispersion and cavity term this has to be done with care because these contributions are generally fitted without explicit solvent. Furthermore, the sampling of this microsolvated system is usually neglected so that the higher accuracy of the specific interactions is traded for the loss in phase space sampling which is in an average way contained in continuum models. For example systems which require a single hydrogen bonded solvent molecule in a well defined position can benefit from this approach.

The accuracy of continuum models is most sensitive to the description of the electrostatic term G_{el} (which depends as well strongly on the definition of the cavity). The mutual polarisation resulting in an electrostatic stabilisation can be described on different levels of sophistication. The governing equation is the Poisson equation

$$\nabla(\epsilon(\boldsymbol{r})\nabla\phi(\boldsymbol{r})) = -4\pi\rho(\boldsymbol{r}) \tag{2.76}$$

with the electrostatic potential ϕ, the charge distribution ρ and the dielectric constant ϵ. The dielectric constant is usually simplified to be independent of the position so that this equation simplifies to

$$\nabla^2\phi(\boldsymbol{r}) = -\frac{4\pi}{\epsilon}\rho(\boldsymbol{r}). \tag{2.77}$$

The reaction field defined as the difference of the potential in vacuum and solution $\phi_{\text{reac}} = \phi_{\text{solv}} - \phi_{\text{vac}}$ can then be used to compute the corresponding energy

$$E_R(\rho,\rho') = \int_{R^3} \rho'(\boldsymbol{r}')\phi_{\text{reac}}(\boldsymbol{r})\mathrm{d}\boldsymbol{r}. \tag{2.78}$$

Approximate representations of the solutes charge distribution combined with simplified cavities lead to different models. The Generalised Born/Surface Area model [46] describes the solute by partial atomic charges. The Onsager model [47] uses a dipole moment for the solute and assumes a spherical cavity. This approach has been generalised to higher electric moments and ellipsoidal cavities.

A different representation is required in order to solve these equations for arbitrarily shaped cavities. The reaction field energy is reformulated into an integral over the surface of the cavity Γ

$$E_R(\rho,\rho') = \int_{\Gamma} \sigma(\boldsymbol{s}')V_M'(\boldsymbol{s}')\mathrm{d}\boldsymbol{s}' \tag{2.79}$$

with the potential V_M' due to the charge distribution of the solute ρ' given by

$$V_M'(\boldsymbol{r}) = \int_{R^3} \frac{\rho'(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}\mathrm{d}\boldsymbol{r}. \tag{2.80}$$

The surface charge density σ can be determined from the integral equation

$$\int_{\Gamma} k_A(\boldsymbol{s},\boldsymbol{s}')\sigma(\boldsymbol{s}')\mathrm{d}\boldsymbol{s}' = b_\rho(\boldsymbol{s}), \tag{2.81}$$

with the Green kernel of some operator A and b_ρ depending linearly on ρ'. In practice the apparent surface charge density is discretised into point charges associated with small segments m of the surface. The choice of A and b_ρ leads to different models for the electrostatic stabilisation.

In a matrix representation the different models can be formulated conveniently. The relevant quantities are the vector of the apparent surface charges $\boldsymbol{\sigma}$, the diagonal matrix of the segments $\boldsymbol{S}$, the vector of the electrostatic potential due to the solute on the segments $\boldsymbol{\phi}^X$ and the vector of the associated normal components of the electric field $\boldsymbol{e}_n^X$. Using the exact dielectric boundary conditions leads to the DPCM expression [48]

$$4\pi\boldsymbol{\sigma} = \frac{\epsilon - 1}{\epsilon + 1}\left(\boldsymbol{e}_n^X + \boldsymbol{D}\boldsymbol{S}\boldsymbol{\sigma}\right) \tag{2.82}$$

with $\boldsymbol{D}$ which generates the normal vectors of the electric field due to the apparent surface charges $\boldsymbol{\sigma}$. The COSMO approach [41] which is formally exact only for the limit of $\epsilon = \infty$, since it uses the boundary conditions of an ideal conductor, is given by

$$0 = \frac{\epsilon - 1}{\epsilon + 1}\boldsymbol{\phi}^X + \boldsymbol{A}\boldsymbol{S}\boldsymbol{\sigma} \tag{2.83}$$

with the Coulomb interaction matrix $\boldsymbol{A}$ for the charges $\boldsymbol{\sigma}$. For neutral solutes it is recommended to use for the empirical parameter $x = 0.5$ and for ions $x = 0$. The advantage of COSMO is that it depends only on the apparent surface charges and not on the normal components of the electric field which are computationally more expensive to obtain and more sensitive to numerical noise. Finally, the integral equation formulation (IEF) of PCM can be obtained by replacing this dependence on the normal components by an expression based on COSMO leading to

$$\left(4\pi\boldsymbol{I} - \frac{\epsilon - 1}{\epsilon + 1}\boldsymbol{D}\boldsymbol{S}\right)\boldsymbol{\sigma} = \frac{\epsilon - 1}{\epsilon + 1}\left(\boldsymbol{D} - 4\pi\boldsymbol{S}^{-1}\right)\boldsymbol{A}^{-1}\boldsymbol{\phi}^X \tag{2.84}$$

with the unit matrix $\boldsymbol{I}$. IEFPCM is an improvement over DPCM. However, it is computationally more expensive than COSMO and more sensitive to numerical problems due to the dependence on the asymmetric matrix $\boldsymbol{D}$. One of the assumptions for these models is that the charge distribution of the solute is completely contained inside of the cavity. This is generally not correct and causes considerable errors in DPCM while COSMO and consequently IEFPCM are less sensitive to this problem.

The COSMO-RS approach extends the COSMO model to realistic solvents (RS). It has been realised that a solvent that can provide the ideal opposite surface charge density for a solute is as efficient in screening the solute as an ideal conductor. Any solvent that cannot provide this ideal surface charge density leads to a mismatch of not exactly opposite surface charge densities. This explains the successful application of COSMO to water. Water exhibits a broad range of different surface charges so that it can easily match in a close to ideal way with a solute and itself. Consequently, one could argue the description of water as an ideal conductor is qualitatively correct.

In practice the surface charge density is determined from a COSMO calculation and the σ-profile, the probability function to find a certain surface charge density, is constructed for the solute and any solvent. This allows a simple treatment of the intermolecular interactions through local pairwise potentials. The misfit energy relative to the ideally screened case is computed for the optimal pairing of these fragments and a constant scaling factor of 0.64 is applied to account for the polarisibilty of involved molecules. [49, 50] The fragments are treated as independent which means that sterical constraints due to the actual geometry of the solvent molecules are not taken into account. Finally an extra term has been introduced for the treatment of hydrogen bonds in order to capture the effect of the mutual charge penetration.

2.3.2 Reference Interaction Site Model

The Reference Interaction Site Model (RISM) is based on principals from statistical mechanics and starts from a Ornstein-Zernike type integral equation. [51] The RISM approach works with spatial distributions instead of directly exploring the phase space. Ornstein and Zernike proposed to split the total correlation function into two terms: the first term including only the direct contribution and the second term all indirect contributions, e.g. the first molecule affects a third molecule which in turn influences the second molecule. The total correlation function measures the effect of a molecule on a second one for a given distance and is related to the radial distribution function $g(r_{12})$ by

$$h(r_{12}) = g(r_{12}) - 1. \tag{2.85}$$

The starting point for the three dimensional RISM approach [52–54] (3D-RISM) is given by

$$h(r_{12}) = \sum_{\alpha} \int c_{\alpha}(\boldsymbol{r} - \boldsymbol{r}')\chi_{\alpha\gamma}(\boldsymbol{r}')\mathrm{d}\boldsymbol{r} \tag{2.86}$$

with the direct correlation function $c_{\alpha}(\boldsymbol{r})$ and the site-site susceptibility $\chi_{\alpha\gamma}(\boldsymbol{r})$ which includes inter- and intramolecular terms. The indices run over all interaction sites of all included solvent species. The asymptotic behaviour for distances beyond the first solvation shell can be derived as

$$c_{\gamma}(\boldsymbol{r}) \propto -\frac{u_{\gamma}(\boldsymbol{r})}{k_{\mathrm{B}}T} \tag{2.87}$$

with the interaction potential $u_{\gamma}(\boldsymbol{r})$ between the solute and the interaction site γ of the solvent.

In order to solve Equation 2.86 an additional closure relation is needed which relates the total and direct correlation function and includes as well the interaction site potential. Two examples which both reproduce by construction the correct asymptotic behaviour are the Mean Spherical Approximation (MSA) [55] and the Hypernetted Chain Closure (HNC). [56] The first one leads to non-physical negative values in the vicinity of associative peaks for the distribution function and the latter can lead to numerical problems for charged sites in polar solvents so that no solution can be found for the given set of equations. The closure proposed by Kovalenko and Hirata (KH) [57] combines the advantages of both closures to

$$g_\gamma(\boldsymbol{r}) = \begin{cases} \exp\left[-\frac{u_\gamma(\boldsymbol{r})}{k_\mathrm{B}T} + h_\gamma(\boldsymbol{r}) - c_\gamma(\boldsymbol{r})\right] & \text{for } g_\gamma(\boldsymbol{r}) \leq 1 \\ 1 - \frac{u_\gamma(\boldsymbol{r})}{k_\mathrm{B}T} + h_\gamma(\boldsymbol{r}) - c_\gamma(\boldsymbol{r}) & \text{for } g_\gamma(\boldsymbol{r}) > 1 \end{cases} \tag{2.88}$$

with a linearised version of MSA applying to regions of solvent enrichment and HNC to regions of solvent depletion. A similar approach is the partial series expansion up to order n (PSE-n) [58] which combines the KH closure and HNC by interpolating between them.

The 3D-KH closure slightly underestimates the height of associative peaks but at the same time slightly widens the peaks. Consequently, the integral of a peak (e.g. coordination numbers), is reproduced rather accurately because these effects cancel out each other. A representative example with regard to the accuracy is water bound to the MgO surface where the coordination numbers are reproduced with 90% accuracy and the positions of the peaks with about 0.5 Å deviations compared to MD. [59] However, the main advantage of the 3D-RISM approach lies in the computational savings because direct simulations are not required. A large explicit solvent simulation with about one million solvent molecules can instead be solved on a standard workstation in a rather short calculation with 98% correlation for the density maxima compared to results from MD.

In practice, the first step is a dielectrically consistent RISM (DRISM) [60] calculation which affords the solvent susceptibility χ that is required as input for the 3D-RISM calculations. Secondly, the embedded cluster RISM method is used to determine the solvent effect on the solute electronic structure in a self-consistent way similar to the self-consistent reaction field approach (Section 2.3.1). The 3D-RISM equations are discretised on a uniform three dimensional grid which includes the first 2–3 solvent shells. Non-periodic contributions are separated out, then the convulation is solved very efficiently by a fast Fourier transform in reciprocal space and finally the non-periodic contributions are added back. [52, 61]

Analytical formulas can be derived for a number of thermodynamic properties, e.g. the excess chemical potential for the KH closure

$$\mu_\mathrm{KH}^\mathrm{ex} = \frac{1}{k_\mathrm{B}T}\sum_\gamma \rho_\gamma \int \left[\frac{1}{2}h_\gamma^2(\boldsymbol{r})\Theta(-h_\gamma(\boldsymbol{r})) - c_\gamma(\boldsymbol{r}) - \frac{1}{2}h_\gamma(\boldsymbol{r})c_\gamma(\boldsymbol{r})\right] \mathrm{d}\boldsymbol{r} \tag{2.89}$$

with the solvent density ρ_γ and the Heaviside step function Θ. The excess free energy together with free solvation energy E_solv can be related to the free energy of reaction in solution

$$\Delta G \approx \Delta E_\mathrm{solv} + \Delta\mu^\mathrm{ex}. \tag{2.90}$$

This merely requires the application of 3D-RISM to the products and educts which is computationally very efficient. [58]

2.3.3 Explicite Solvation with Hybrid QM/MM Methods

The full QM treatment of complete proteins is nowadays possible with algorithmic advances as well as an increase in the available computing power. However, this approach limits severely the possibility to explore the phase space. On the other hand, extensive sampling can be achieved with molecular mechanics which allows to study even slow folding processes of enzymes. Nevertheless, the classical potential cannot describe quantum effects which are important e.g. for reactivity where bonds are broken and formed frequently. This is the context in which hybrid Quantum Mechanics/Molecular Mechanics (QM/MM) methods have been developed. The high-accuracy of QM methods is used for a small part of special interest — the high level (HL) — like the active site of an enzyme or transition metal catalyst in solution. The efficiency of Molecular Mechanics (MM) methods is used to describe the large number of particles of the embedding environment — the low level (LL) — like the protein or the solvent. Different schemes have been developed in order to couple these methods which will be discussed briefly in the remainder of this Section. [62]

In this scheme the Hamiltonian is replaced by an effective Hamiltonian. Depending on the approach the energy can be partitioned in two different ways. In the subtractive scheme the energy for the total system (S) is given by

$$E^{\text{sub}} = E_{\text{QM}}^{\text{HL}} + E_{\text{MM}}^{\text{S}} - E_{\text{MM}}^{\text{HL}}. \tag{2.91}$$

The advantage of the subtractive scheme is its simplicity because no direct coupling between QM and MM has to be computed. However, empirical parameters have to be available even for the HL which can be especially problematic for reactions. Furthermore, the interaction between HL and LL is computed only at the MM level. In contrast, the additive scheme includes an explicit term for the coupling between HL and LL

$$E^{\text{add}} = E_{\text{QM}}^{\text{HL}} + E_{\text{MM}}^{\text{LL}} + E_{\text{QM/MM}}^{\text{HL,LL}}. \tag{2.92}$$

The MM computation is only carried out for the LL. This scheme is most commonly used for QM/MM calculations and the definition of the coupling term gives rise to different flavours of QM/MM methods. Three variants are presented in the following going from low accuracy and low computational costs to high accuracy and high computational costs. [63]

In both the substractive and additive schemes, the mechanical embedding treats the coupling completely at the MM level which is computationally very efficient. The disadvantages are that empirical parameters are required also for the HL. These do not necessarily represent the correct charge distribution of the HL. If the character of an atom changes during the course of an reaction its parameters should be adjusted. There is no clear way to do this. Finally, the electronic density does not interact in any way with the charges of the LL and is not polarised by the environment.

The electrostatic embedding scheme improves upon this by computing the electrostatic interaction at the QM level. The point charges of the LL are directly included in the Hamiltonian.

Thus, the electronic density of the HL is polarised by its environment. However, MM point charges in the direct vicinity of the HL can cause a spurious polarisation of the QM density and charge leakage from QM atoms to MM sites may occur. Nevertheless, it has been found that this is a minor problem for limited basis sets. [64] This embedding scheme is widely used and the results are reasonable although it is not clear if the interaction is described correctly. The MM partial charges are parametrised to give a balanced description of the force field and not to represent the true charge distribution. Replacing part of a system with the correct charge distribution could deteriorate the results. [65]

Finally, the most sophisticated scheme takes the polarisation of both HL and LL into account. The electrostatic embedding is combined with a polarisable force field. While the solute polarises the solvent the same holds true vice versa. Therefore, this mutual polarisation has to be solved iteratively in a self-consistent way and is associated with considerably higher computational costs. Approximate schemes can be introduced which truncate this iterative procedure before convergence is reached.

The comparison with MM simulations and the QM cluster approach, which has in principle a more accurate formulation of the potential energy, reveals the strength and weaknesses of the QM/MM approach. The optimization of geometries is an important criteria for the accuracy of any method. QM/MM approaches have a clear advantage in comparison with the QM cluster approach. The geometry of a cluster in vacuum cannot be simply optimised because severe surface effects will influence the results. For example the density of a solvent cluster is lower at the surface and solvent molecules rearrange in order to maximise the interaction with the cluster. Therefore, it may be necessary to artificially keep the position of one or several atoms of the outer molecules fixed. This constraints might introduce a bias into the results. QM/MM approaches include the complete environment explicitly and it has been shown that reliable geometries can be obtained even with relatively small QM parts. [66] Furthermore, it is not clear which parts of the environment should be included with the QM cluster approach. Obviously this choice might be biased and lead to the right results for the wrong reasons. A similar decision has to be made in QM/MM methods about the size of the QM part. However, the difference between a group at the MM level or the QM level is obviously much smaller than a group at the QM level or the continuum solvation. [67] Finally, large clusters which are expected to increase the accuracy turn out to be problematic in some cases especially in the description of reaction pathways. The study of the potential energy surface is complicated by a manifold of minima which are associated with small unrelated changes in the environment. This can be partly circumvented by using snapshots from classical simulations which can provide an approximate sampling of the configuration space. A different approach is to freeze a part of the environment after the initial geometry optimisation. This restricts the environment to a single local minimum and avoids thereby the effect of unrelated solvent rearrangements on the energy.

QM/MM methods with their smaller QM part and the associated lower computational costs can be used to sample degrees of freedom on a short time scale. However, compared to MM or semiempirical methods the computational costs are much higher and it is not possible to simulate longer time scales or to extensively sample the phase space. Therefore, it turns out to be difficult to obtain dynamic information or free energies because entropic effects cannot be easily estimated.

The energy is more sensitive to the size of the QM system than the geometry during optimisations. Therefore, it can turn out to be advantageous to first optimize the geometry with a QM/MM approach with a rather small QM system. Then one can carry out a single point calculation with a larger QM part to obtain accurate energies. [66] This also removes the dependence on the empirical force field. It has been observed that different force fields can have a large influence on the energetics. [68] However, this approach does not correct for the inherent dependence of the sampling of the configuration space on the force field.

The definition of HL and LL is critical for any QM/MM study. For the treatment of a single solute with its surrounding solvent the choice of the HL is usually only the solute itself. The situation can be more complicated for the description of enzymes. The active site which should be included in the HL usually contains a metal ion, a potential substrate and amino acid side chains in direct vicinity. However, these side chains are covalently bound to the embedding enzyme. The QM/MM boundary has to cut at some point through one of these covalent bonds. These junctions have to be dealt with at the QM as well as at the MM level of theory.

The link-atom method is technically simple and therefore widely used. The dangling bond of the QM part is saturated with an additional atom which is not present in the real system. This is usually a hydrogen atom but can be in principle any atom with a single valence. The position of the link-atom is determined by the cut bond in order to avoid additional degrees of freedom. The cut bond is only treated at the MM level. [63]

The pseudobond method avoids the introduction of additional atoms. Instead it assigns the MM atoms which participate in a cut bond a specifically designed basis set and an effective core potential to saturate the free valency of the neighbouring QM atom. These boundary atoms interact therefore with the MM part as well as the QM part. The pseudopotentials can be designed to reproduce the properties of certain groups, e.g. a methyl group. [63, 68]

Finally a frozen localized orbital approach can be used. A special localized orbital is placed at the QM atom of the cut bond and oriented correctly along the bond. This orbital is kept frozen during the self-consistent field cycles and does not mix with other orbitals. In this way it is used to satisfy the free valency of the atom. It can be imagined as a frozen lone pair which replaces the original bond. [68]

Nevertheless, the overpolarisation of the electronic density at the boundary can pose serious problems. Different schemes have been developed to remove MM charges in close vicinity to the QM part for example by reassigning them to neighbouring atoms. The charges can be also scaled to decrease the polarisation effect or described by a more accurate Gaussian distribution.

It has been observed that the results depend strongly on this boundary treatment. This influence can be decreased considerably by increasing the size of the QM part and consequently moving the junction away from the active centre. It has been recommended that the distance to the active centre should be at least two residues. [67]

Besides the treatment of the border, the computation of long-range electrostatics is technically difficult and there is no unique approach which turns out to be ideal. For pure MM computations the Ewald summation method is routinely applied. [32] First implementations of the Ewald summation for QM/MM calculations neglected the QM-QM interaction. However, implementations including all interactions have been developed in the context of plane-wave calculations [69] and more recently also with a real space multigrid approach by Parrinello and coworkers. [70, 71] A simple approach is to use a cutoff for the MM charges which are included in the Hamiltonian. This works rather well because in most cases the long-range polarisation of the QM density can be neglected. Alternatively, point charges or multipoles which represent the true electrostatic potential of the QM part can be computed. These electric moments can be used subsequently for the Ewald summation. [68]

By and large, QM/MM methods have been established as a state-of-the-art approach and can be routinely used to study localized electronic events embedded in a large environment. However, the setup remains complicated and the methods are far from a black box approach. The results can in principle be improved by increasing the size of the QM region. However, one encounters the same problems as with QM cluster calculations: the sampling of the QM system is problematic and the choice on which parts to include in the QM system might be biased.

2.4 Molecular Simulations

Traditionally a vast majority of chemical experiments is carried out in solution at a finite temperature, most commonly room temperature. Also biochemically relevant processes occur mostly in aqueous solution with a number of further molecules like ions, acids and bases setting a certain pH or even whole proteins or aggregates like membranes making up very complex systems. The thermodynamic state in liquid phase is controlled by a few macroscopic variables with the most natural being the number of particles N, the pressure P and the temperature T. Any thermodynamic property depends on the state and not only on the instantaneous coordinates and momenta of the particles. Therefore, approaches treating the particles as isolated according to an ideal gas which is the basis of the rigid rotor/harmonic oscillator approach in QM cannot be used for this type of systems.

Any measurement of an observable is in fact the time average of the instantaneous realisations of this property which can be formulated as

$$\begin{aligned} A_{\mathrm{obs}} &= \left\langle A\left(\Gamma(t)\right)\right\rangle_{\mathrm{time}} && (2.93) \\ &= \frac{1}{t_{\mathrm{obs}}} \int_0^{t_{\mathrm{obs}}} A\left(\Gamma(t)\right) \mathrm{d}t. && (2.94) \end{aligned}$$

The instantaneous coordinates and momenta define a point of the phase space Γ which evolves along a trajectory with time. This evolution is defined by Newton's equations of motion for a classical system. The MD method described in the following Section 2.4.1 uses this approach to compute thermodynamic properties.

The basis for statistical mechanics, however, is based on a slightly different approach. Gibbs suggested to replace the average over time by an average over an ensemble

$$A_{\mathrm{obs}} = \langle A | \rho_{\mathrm{ens}} \rangle_{\mathrm{ens}} \tag{2.95}$$

$$= \sum_{\Gamma} A(\Gamma) \rho_{\mathrm{ens}}(\Gamma). \tag{2.96}$$

An ensemble is here a collection of points in phase space that are distributed according to the probability ρ_{ens}. Instead of following a system moving through the phase space with time $\Gamma(t)$ one can equivalently observe the probability distribution for a fixed phase point with time $\rho(\Gamma, t)$. For a system in equilibrium the distribution does not change with time $\frac{\partial \rho_{\mathrm{ens}}}{\partial t} = 0$ and it is consequently denoted as equilibrium distribution. The MC method is based on the description of ensembles and will be discussed in Section 2.4.2.

Both of these methods rely on the important property of ergodicity. If we observe the trajectory of a system starting from an arbitrary point in phase space and it passes through all points of the phase space returning to the initial position the system is considered ergodic. If on the other hand regions of the phase space are not accessible and contain cyclic trajectories then the system is not ergodic. In that case the region of phase space that is sampled is dependent on the initial conditions. Deep wells or barriers can lead to inaccessible regions in phase space. However, in practice it is impossible to prove the ergodicity of a realistic system and only independent simulations with different initial conditions gives some evidence towards this important property.

The averages presented so far are computed either for an infinite amount of time or for an infinitely large ensemble which affords identical results for any simulation under the condition of ergodicity and a constant equilibrium distribution. However, the difficulty lies in sampling sufficiently long the important part of phase space. These rather vague formulations emphasise the difficulties in defining what is sufficiently long and which parts of phase space are actually important. Many configurations can be discarded due to the overlap of atoms. Nevertheless, the remaining space cannot be easily defined.

2.4.1 Molecular Dynamics

Molecular dynamics is a method which computes the integral over time as formulated in Equation 2.93 in order to evaluate thermodynamic properties as function of the state. The equations governing the evolution in time for a potential which is independent of the time t and the momenta $\boldsymbol{p}_i$ can be formulated as $6N$ first order differential equations

$$\dot{\boldsymbol{r}}_i = \frac{\boldsymbol{p}_i}{m_i} \tag{2.97}$$

$$\dot{\boldsymbol{p}}_i = -\nabla_{\boldsymbol{r}_i} V \tag{2.98}$$

or equivalently $3N$ second order differential equations

$$\ddot{\boldsymbol{r}}_i = \frac{\dot{\boldsymbol{p}}_i}{m_i} \tag{2.99}$$

with the number of particles N and the coordinates $\boldsymbol{r}_i$. These equations fulfil a number of properties. For a potential that depends only on the distances between particles and without external field the total linear momentum is conserved for an isolated system as well as under periodic boundary conditions. Since the potential is not time dependent, the derivative of the Hamiltonian is constant as well $\frac{\partial H}{\partial t} = 0$. Hence, the energy of the system is conserved. Finally, the equations are reversible in time. Upon change of the sign of all momenta a system follows its trajectory in reverse direction. These properties should be maintained by any algorithm devised to solve the differential equations at hand.

For a continuous potential V as they have been introduced in Section 2.2 a finite difference approach can be used. The idea is to evolve the system stepwise in time according to the equations of motion and thereby to numerically evaluate the desired integral over time. The predictor-corrector algorithm is a widely used finite difference method and variations thereof give rise to a number of integrators. The predictor step is based on a Taylor expansion at the time t which allows to predict the system at time $t + \Delta t$

$$\boldsymbol{r}^{\mathrm{P}}(t+\Delta t) = \boldsymbol{r}(t) + \Delta t \boldsymbol{v}(t) + \frac{1}{2}\Delta t^2 \boldsymbol{a}(t) + \ldots \tag{2.100}$$

$$\boldsymbol{v}^{\mathrm{P}}(t+\Delta t) = \boldsymbol{v}(t) + \Delta t \boldsymbol{a}(t) + \frac{1}{2}\Delta t^2 \boldsymbol{b}(t) + \ldots \tag{2.101}$$

$$\boldsymbol{a}^{\mathrm{P}}(t+\Delta t) = \boldsymbol{a}(t) + \Delta t \boldsymbol{b}(t) + \ldots \tag{2.102}$$

with the coordinates $\boldsymbol{r}$ and their first, second, etc. derivatives $\boldsymbol{v}$, $\boldsymbol{a}$, $\boldsymbol{b}$. Any of these derivatives can be formulated as well as a numerical derivative according to a backward two-point scheme based on previous steps e.g. $\boldsymbol{r}(t-\Delta t)$ and $\boldsymbol{r}(t-2\Delta t)$. Different schemes include the expansion up to different order which allows balancing the accuracy with the computational costs.

A corrector step is needed to incorporate the equations of motion into the algorithm and generate a correct trajectory. Therefore, the exact accelerations are computed at the predicted positions which allows to estimate the error of the prediction $\Delta \boldsymbol{a}(t+\Delta t) = \boldsymbol{a}^{\mathrm{exact}}(t+\Delta t) - \boldsymbol{a}^{\mathrm{P}}(t+\Delta t)$. Next, this estimate can be used to correct the positions

$$\boldsymbol{r}^{\mathrm{c}}(t+\Delta t) = \boldsymbol{r}^{\mathrm{P}}(t+\Delta t) + c_0 \Delta \boldsymbol{a}(t+\Delta t) \tag{2.103}$$

with some constant c_0. Generally, this correction step is carried out only once because the associated costs with the evaluation of the forces is very high. However, in principle the correction may be carried out iteratively converging to arbitrary accuracy.

Different realisations of the predictor-corrector algorithm vary in their accuracy and the computational costs, to which degree the conservation laws are obeyed and if the algorithm is time reversible. The conservation of the energy is important because it ensures that the trajectory stays on the correct hypersurface in phase space and only then the correct ensemble is generated.

The natural ensemble that is generated by MD — considering liquid phase simulations under PBC — is the microcanonical ensemble (NVE) with a constant number of particles N, a constant volume V and a constant energy E. Different ensembles can be generated by introducing a thermostat to obtain the canonical ensemble (NVT) or by adding furthermore a barostat to obtain the grand canonical ensemble (NPT). The simplest approaches achieve this by scaling the velocities or the volume respectively. However, this does not generate the canonical distribution because the temperature and pressure do not have the correct fluctuations. Based on the work of Nosé a thermostat has been developed by Hoover. [72] This approach introduces an additional degree of freedom which represents a heat bath. It can be shown that this approach produces the canonical ensemble. Furthermore, the algorithm can be used equivalently to obtain the grand canonical ensemble with another additional degree of freedom. These are examples for global thermostats because the temperature is defined as a function of all particles. On the other hand, a local thermostat dissipates energy locally and heats up parts of the system while others are cold down. An example is the Langevin dynamics [73] which has been introduced for the simulation of polymers. The effect of a viscous solvent is mimicked which allows controlling temperature as well as pressure and thereby approximating the canonical or grand canonical ensemble.

2.4.2 Metropolis Monte Carlo Simulations

The ensemble average of any property depends on the partition function Z which is defined for a continuous canonical system as the integral over the phase space Γ

$$Z = \int \exp(-\beta V(\Gamma))\mathrm{d}\Gamma \tag{2.104}$$

with the total energy $V(\Gamma)$ and $\beta = \frac{1}{k_\mathrm{B}T}$. Based on the partition function any observable A can be computed as an ensemble average by

$$\langle A \rangle_\mathrm{NVT} = \frac{\int A \exp(-\beta V(\Gamma))\mathrm{d}\Gamma}{Z}. \tag{2.105}$$

These integrals could in principle be computed by Monte Carlo simulations. However, while the numerator depends only on likely configurations where A is significant the partition function is rather sensitive to the huge amount of unlikely configurations which in their entirety contribute still significantly. Additionally, the large number of degrees of freedom resulting in a huge phase space cannot be sampled easily and most time is spent on configurations that are energetically so unfavourable due to the overlap of the repulsive cores of atoms that they are physically irrelevant.

This lead to the development of the importance sampling which focuses the sampling on the important parts of the phase space. Instead of sampling an even distribution and weighting every sample with their corresponding weight $\exp(-V(\Gamma)/(k_\mathrm{B}T))$ as suggested by Equation 2.105, the samples are generated with the probability $\exp(-V(\Gamma)/(k_\mathrm{B}T))$ and do not have to be weighted at all. Therefore, an ensemble average is simply given by the sum over the realisations of A

$$\langle A \rangle_\mathrm{NVT} \approx \frac{1}{N} \sum_x A(x) \tag{2.106}$$

with the number of configurations N. In other words, the sampling is carried out according to the desired distribution which is in known beforehand and is in this case the Boltzmann distribution. As a result, the partition function does not have to be computed explicitly at any given time with this approach. However, the difficulty is shifted to finding an efficient algorithm to generate points in phase space according to the Boltzmann distribution.

For this purpose, a Markov chain is constructed which satisfies the conditions that each sample belongs to the phase space and that every step depends only exactly on the current configuration but is independent of all previous configurations. The equilibrium distribution ρ can then be reached by successive application of the transition matrix $\mathbf{\Pi}$ which is built from the transition probability to move from a state n to any state m including to remain in the current state given by the diagonal elements. The equilibrium distribution satisfies the eigenvalue equation

$$\rho\mathbf{\Pi} = \rho \tag{2.107}$$

with the eigenvalue unity. The transition matrix is a stochastic matrix because the sum of all elements of a row is unity $\sum_m \Pi_{mn} = 1$. It has exactly one eigenvalue that is unity with the corresponding eigenvector giving the equilibrium distribution while all other eigenvalues are positive and smaller than unity. They govern how fast a given distribution converges towards the equilibrium distribution upon application of the transition matrix. It is guaranteed for any transition matrix that the equilibrium distribution is reached and this is in fact independent of the start conditions but in the interest of computational costs it is important that this is reached as fast as possible.

Microscopic reversibility can be imposed in the definition of the transition matrix

$$\rho_m \Pi_{mn} = \rho_n \Pi_{nm} \tag{2.108}$$

but is not strictly necessary. The matrix $\mathbf{\Pi}$ only has to obey the weaker condition given by Equation 2.107. The transition step is separated into two parts. First a new state is proposed according to the proposal matrix $\boldsymbol{P}$ and second the step is accepted according to the acceptance matrix $\boldsymbol{A}$. The proposal matrix is chosen to be symmetric which reflects the detailed balance condition (Equation 2.108) and the elements of the acceptance matrix are given by

$$A_{\mathrm{mn}} = \min\left(1, \frac{\rho_n}{\rho_m}\right) \tag{2.109}$$

as it has been suggested by Metropolis and coworkers. [74] Important to note is that the acceptance depends only on the ratio $\frac{\rho_n}{\rho_m}$ and therefore does not depend on the partition function.

In order to show that indeed the equilibrium distribution is obtained by this procedure we write down the probability to move from state s to r as a function of the energy of these states

$$p_{s\to r} = \begin{cases} 1 & E_r < E_s \\ e^{-\frac{E_r - E_s}{kT}} & E_r > E_s. \end{cases} \tag{2.110}$$

If we consider now an ensemble of states and carry out a Metropolis Monte Carlo step in all the systems, the net number of systems going from state s to r is

$$\overline{p}_{s\to r} = \frac{\nu_s}{\nu_r} \frac{\exp\left(\frac{-E_r}{kT}\right)}{\exp\left(\frac{-E_s}{kT}\right)} - 1 \tag{2.111}$$

with $E_r > E_s$. In the case of an equilibrium distribution the ratio of the probabilities of the two states is

$$\frac{\nu_s}{\nu_r} = \frac{\exp\left(\frac{-E_s}{kT}\right)}{\exp\left(\frac{-E_r}{kT}\right)} \tag{2.112}$$

which we can insert into Equation 2.111 resulting in $\overline{p}_{s\to r} = 0$. This means if we have already the equilibrium distribution this procedure will not change the distribution and therefore obeys Equation 2.107. In the case that more systems are in state s than in state r compared to the equilibrium distribution

$$\frac{\nu_s}{\nu_r} > \frac{\exp\left(\frac{-E_s}{kT}\right)}{\exp\left(\frac{-E_r}{kT}\right)} \tag{2.113}$$

and consequently we obtain $\overline{p}_{s\to r} > 0$. Hence the distribution changes until equilibrium is reached.

The proposed steps are carried out by choosing randomly a single particle and translating it to a randomly selected position inside of a cube. The size of the cube is critical for the number of accepted steps and consequently for the convergence. If the steps are very short many steps are accepted but the generated states are highly correlated and the system progresses only slowly. If on the other hand too large steps are choosen most steps are rejected and the sampling is also inefficient. The step size is highly system dependent and a target acceptance ration of about 50% is widely used although it can be shown that acceptance ratios of 23.4% are more efficient. [75] Furthermore, it has been shown that the asymptotically ideal acceptance ratio depends on the target distribution. [76]

For anisotropic particles (e.g. molecules) additional rotational steps have to be carried out. There is no unique solution for the step size of translation and rotation for a given target acceptance ratio hence the rotation step size is usually fixed and only the translation step is varied. Furthermore, this samples only the intermolecular degrees of freedom with rigid molecules. The sampling of internal degrees of freedom requires special attention and is an ongoing field of research. The different conformers due to the secondary structure of a protein can be sampled by carrying out steps in the dihedral angles of the backbone. However, the overlap between atoms becomes very likely for larger molecules and no unique approach has been established so far for the sampling of internal degrees of freedom. Based on the work of Maginn and coworkers [77] a method has been implemented in the MC program package DICE [78] which breaks the molecule of interest into fragments, separates the hard and soft degrees of freedom and finally reconnects these fragments in order to generate very efficiently new configurations based on a configurational bias MC approach.

The acceptance criteria depends on the potential of the involved states but only the change of the energy has to be computed. That means for a two-body potential only terms involving the changed positions M have to be recomputed instead of the total energy

$$\Delta E = \sum_{i \in M} \sum_{j \neq i} f(r_{ij}) \tag{2.114}$$

which means that the computational costs for the change of the energy scale only linear with the system size. The acceptance criteria proposed by Metropolis and coworkers [74] in fact maximizes the acceptance of steps and therefore the phase space exploration. It is important to note that in case a state is not accepted the previous state has to be taken into account another time in order to obtain the correct averages.

2.5 Free Energies in Solution

The free energy cannot be computed as an ensemble average with the Metropolis Monte Carlo method because it depends directly on the partition function. However, the change of the free energy can be computed by employing a procedure that is similar to calorimetry. The change of the free energy along some path from a reference state to the state of interest can be computed with the Free Energy Perturbation (FEP) method. Since the free energy is a state function any path can be chosen and opposite to calorimetry this path does not have to be physically accessible and can even include changes of the Hamiltonian.

The free Helmholtz energy A is defined as a function of the partition function Q by

$$A = -k_{\mathrm{B}} T \ln Q. \tag{2.115}$$

Since the free energy cannot be computed directly we write an expression for the change of the free energy from reference r to target t

$$\begin{aligned} A_t - A_r &= -k_{\mathrm{B}} T (\ln Q_t - \ln Q_r) && (2.116) \\ &= -k_{\mathrm{B}} T \ln \frac{Q_t}{Q_r}. && (2.117) \end{aligned}$$

Next, the partition function of the target can be separated by adding and subtracting the potential of the reference

$$\begin{aligned} A_t - A_r &= -k_{\mathrm{B}} T \frac{\int \exp\left(-\frac{1}{k_{\mathrm{B}} T}(V_t - V_r + V_r)\right) \mathrm{d}\Gamma}{Q_r} && (2.118) \\ &= -k_{\mathrm{B}} T \frac{\int \exp\left(-\frac{1}{k_{\mathrm{B}} T}(V_t - V_r)\right) \exp\left(-\frac{1}{k_{\mathrm{B}} T} V_r\right) \mathrm{d}\Gamma}{Q_r}. && (2.119) \end{aligned}$$

Now it is obvious that the latter expression is indeed an ensemble average of the change of the potential energy and does not depend on the partition function directly in the framework of Metropolis MC simulations. This relationship has been first derived by Zwanzig [79] and the final formulation is

$$\Delta A_{\mathrm{r}\rightarrow\mathrm{t}} = -k_{\mathrm{B}}T \ln \left\langle \exp\left(-\frac{1}{k_{\mathrm{B}}T}(V_t - V_r)\right)\right\rangle_{\mathrm{r}} \tag{2.120}$$

evaluated at the reference state r.

The expression for the reverse direction can be obtained by simply exchanging the indices of target and reference and ideally the same change of the free energy should be obtained just with the opposite sign. However, the formal requirement of FEP is that all configurations of the target configuration space are included in the reference space. With a decreasing overlap between reference and target the energetic difference between the forward and reverse simulation becomes larger. It is possible to estimate the free energy change as the average of forward and reverse directions. However, this is strongly discouraged because the underlying assumption is that the phase space of the reference which is not included in the target is as large as the phase space of the target which is not included in the reference. Nevertheless, this cannot be known in advance and many times one direction is actually clearly preferred and only associated with a small error. Using a simple average in such a case will consequently result in an even worse estimate of the free energy change.

The best possible estimator for the free energy is the Bennett Acceptance Ratio (BAR) method which uses results of the forward as well as the reverse simulation simultaneously to achieve the best possible accuracy. [80] Let us consider a simulation with additional moves which keep the configuration fixed but switch from the reference to the target potential or vice versa. For the acceptance ratios of those moves it can be formulated

$$M(U_t - U_r)\exp(-U_r) = M(U_r - U_t)\exp(-U_t) \tag{2.121}$$

with the Metropolis function $M(x) = \min(1, \exp(-x))$. Integrating this expression over the configuration space and multiplying it with $\frac{Q_r}{Q_r}$ and $\frac{Q_t}{Q_t}$ respectively leads to

$$Q_r\frac{\int M(U_t - U_r)\exp(-U_r)\mathrm{d}\Gamma}{Q_r} = Q_t\frac{\int M(U_r - U_t)\exp(-U_t)\mathrm{d}\Gamma}{Q_t}. \tag{2.122}$$

Both fractions are actually ensemble averages of either the target and reference and can be rewritten as

$$\frac{Q_r}{Q_t} = \frac{\langle M(U_r - U_t)\rangle_t}{\langle M(U_t - U_r)\rangle_r}. \tag{2.123}$$

Moreover, it can be shown that this expression can be further improved by substituting the Metropolis function by the Fermi function $f(x) = \frac{1}{1+\exp(x)}$ and shifting the potentials by a constant C

$$\frac{Q_r}{Q_t} = \frac{\langle f(U_r - U_t + C)\rangle_t}{\langle f(U_t - U_r - C)\rangle_r}\exp(C) \tag{2.124}$$

with C depending on the unknown ratio of the partition functions

$$C = \ln\frac{Q_r n_t}{Q_t n_r} \tag{2.125}$$

and the steps sampled in the reference n_r and the target state n_t. This leads to the final equations to estimate the free energy change

$$\Delta A_{\text{est}} = \ln \frac{\sum_t f(U_r - U_t + C)}{\sum_r f(U_t - U_r - C)} + C - \ln \frac{n_t}{n_r} \tag{2.126}$$

$$\Delta A_{\text{est}} = C - \ln \frac{n_t}{n_r} \tag{2.127}$$

which have to be solved self-consistently. The results are rather insensitive to the actual number of steps sampled in the target or reference state and in practice just the results of two separate simulations are analysed by this procedure instead of carrying out a single simulation with switching moves.

If the overlap is still insufficient then the free energy cannot be estimated at all and further intermediate states have to be constructed. As pointed out before these do not have to correspond to actual physical systems. Changes might include geometrical degrees of freedom or changes in the Hamiltonian. A common option is to build a linear combination of the Hamiltonian of target and reference state $H = \lambda H_t + (1 - \lambda)H_r$ and to stepwise change λ from zero to unity.

2.6 Computation of Electronic Spectra

The following formalism to compute electronic spectra will be applied in Chapter 6. Upon the interaction of a molecule with an external field the former may undergo a transition from the ground state to an excited state. Computationally this physical phenomenon may be described by linear response theory and in the framework of DFT the response of the density to a time-dependent external field is derived which allows determining the response to any external field.

The proof of Hohenberg and Kohn [13] has been extended by Runge and Gross [81] to time-dependent systems and shows that at any given time the density is uniquely determined by the external potential. In order to obtain the linear response function a time-dependent external potential is added to the Hamiltonian

$$H = H_0 + V_{\text{ext}}(t). \tag{2.128}$$

Starting from the time-dependent Schrödinger equation in the interaction picture, which means that both state vectors as well as operators are time-dependent, the time evolution is given by

$$i\frac{\mathrm{d}}{\mathrm{d}t} |\Psi(t)_I\rangle = V_I |\Psi(t)_I\rangle \tag{2.129}$$

with $|\Psi(t)_I\rangle = \exp(iH_0t)|\Psi(t)_I\rangle$ and $V_I = \exp(iH_0t)\, V_{ext}(t) \exp(-iH_0t)$. Expanding this expression as a Dyson series and truncating after the first term, which corresponds to the linear response, leads to

$$|\Psi(t)_I\rangle = |\Psi(0)\rangle - i\int_{-\infty}^{t} V_I(t')\,|\Psi(0)\rangle\,\mathrm{d}t'. \tag{2.130}$$

Because of the truncation after the first term — linear response — this expression is only exact for small perturbations. A second approximation is visible as the lower limit of $-\infty$. This adiabatic approximation means that the perturbation is turned on slow enough that the unperturbed wave function can follow the perturbation. This holds only true for small perturbations. The response function for arbitrary operators can be derived but in Time-Dependent Density Functional Theory (TD-DFT) mostly the response of the density is of interest. Upon performing a Fourier transform the frequency-dependent response is obtained

$$\delta \left\langle \rho(\boldsymbol{r},\omega) \right\rangle = \int \chi(\boldsymbol{r},\boldsymbol{r}',\omega) V_{\mathrm{ext}}(\boldsymbol{r}',\omega) \mathrm{d}\boldsymbol{r}' \tag{2.131}$$

with the external potential in the frequency domain

$$V_{\mathrm{ext}}(\boldsymbol{r}',\omega) = \int_{-\infty}^{\infty} V_{\mathrm{ext}}(\boldsymbol{r}',t) \exp(i\omega t) \mathrm{d}t. \tag{2.132}$$

Using the ground state as initial state and inserting the set of unperturbed states the spectral representation of the response function can be obtained

$$\chi(\boldsymbol{r},\boldsymbol{r}',\omega) = \lim_{\eta\to 0} \sum_{n\neq 0} \left[\frac{\langle \Psi_0 | \rho(\boldsymbol{r}) | \Psi_n \rangle \langle \Psi_n | \rho(\boldsymbol{r}') | \Psi_0 \rangle}{\omega - (E_n - E_0) + i\eta} - \frac{\langle \Psi_0 | \rho(\boldsymbol{r}') | \Psi_n \rangle \langle \Psi_n | \rho(\boldsymbol{r}) | \Psi_0 \rangle}{\omega + (E_n - E_0) + i\eta} \right]. \tag{2.133}$$

The excitation energies can be determined from the poles of this function. [12]

TD-DFT can be also used in the context of QM/MM calculations for the QM part allowing to study the influence of the environment on the electronic excitations. If the excitations are computed for fixed solute and solvent configurations then vertical excitation energies are obtained. This is a good approximation because the solvent rearranges only slowly upon excitation. Furthermore, using a force field with fixed partial charges there is only an indirect influence on the excitation energies. Due to the inclusion of the charges in the Hamiltonian the unperturbed density is influenced which in turn affects the excitations. However, the electronic response of the solvent may need to be considered for certain systems which can be done in combination with a polarisable force field.

On the other hand, in the formalism of continuum models the response of the continuum can be separated into a fast and a slow response. In the context of electronic excitations the slow term corresponds to the response of the nuclear coordinates while the fast term describes the response of the electronic density of the solvent molecules. This separation leads to the non-equilibrium regime. Consequently, the unperturbed state is influenced by the embedding which influences indirectly the excitations but an additional term represents the explicit dependence of the excitations on the fast response of the medium. [82]

CHAPTER 3

The Perturbative Metropolis Monte Carlo Method

The computational costs in QM/MM based simulations is solely determined by the QM calculation. The latter requires many costly iterations in the SCF cycles. It has been realised by a number of groups at about the same time that the influence on the wavefunction of a single MC step in a QM/MM based simulation is very small. [84–86] Compared to MD generally only a single molecule is moved in MC simulations. The change of the total energy is consequently much smaller. It is also important to note that MD simulations require the fully converged wave function because the energy as well as the forces are needed for the algorithm. The forces which are the first derivates of the energy are much more sensitive to numerical noise and a stricter convergence is necessary. MC simulations, however, are based only on the energies.

Therefore, an approximation can be used in order to compute the change of the energy of a single MC move without doing a full SCF cycle. Truong and Stefanovich proposed to use first order perturbation theory and thereby were the first to establish the terminology for this approach. [86] In other words, the Perturbative QM/MM Metropolis Monte Carlo (PMC) approach is a method to perform simulations at the QM/MM level with huge computational savings by introducing a single approximation for the calculation of the change of the electrostatic interaction term between QM and MM part. The working equations as derived by Truong and Stefanovich will be described in the following Sections. Subsequently, these will be extended to systems under periodic boundary conditions where special care has to be taken of the long-range electrostatic interactions. An efficient implementation in the context of hybrid architectures has been developed [83] which combines the strength of a Graphics Processing Unit (GPU) from graphic cards and the conventional Central Processing Unit (CPU) into a single algorithm. This allows carrying out simulations which are several orders of magnitudes faster than full QM/MM MC or MD simulations. Furthermore, these simulations can be run on commodity hardware and at the same time allow energy savings of up to 64%. Finally, further implementation details, used libraries and additional modules for the analysis of trajectories will be presented.

The material in this Chapter was presented in part in Reference [83].

3.1 Basic Formulation

The effective Hamiltonian that is used for PMC simulations is partitioned as in the additive QM/MM approach discussed in Section 2.3.3. The terms correspond to the QM part which describes the solute in vacuum, the purely classical part which describes the internal energy of the solvent and the interaction between those two subsystems:

$$H_{\text{eff}} = H_{\text{QM}} + H_{\text{QM/MM}} + H_{\text{MM}} \tag{3.1}$$

The total energy is given as

$$E_{\text{tot}} = \left\langle \Psi \middle| H_{\text{QM}} - \sum_i \sum_\alpha \frac{q_\alpha}{r_{i\alpha}} \middle| \Psi \right\rangle + \sum_A \sum_\alpha \frac{q_\alpha Z_A}{r_{A\alpha}} + E^{\text{vdW}}_{\text{s-S}} + E^{\text{MM}} \tag{3.2}$$

with the index i running over the electrons and A over the nuclei of the solute, while α runs over the solvent interaction sites. The partial charges are given by q, nuclear charges by Z and the distances between sites by r. The effective Hamiltonian as defined in Equation 3.2 changes upon the move of a single molecule by the term

$$\Delta H = \sum_i \sum_{\alpha \in m} -q_\alpha \left(\frac{1}{r'_{i\alpha}} - \frac{1}{r_{i\alpha}} \right) \tag{3.3}$$

with the new distances given by r'. It is important to note that this term depends only on the interaction sites m of the single molecule that has been moved while the first term in Equation 3.2 depends on the coordinates of all solvent molecules. Given that the perturbation is small the change of the energy can be computed by first order perturbation theory

$$\Delta E_{\text{tot}} = \left\langle \Psi \middle| \sum_{\alpha \in m} -q_\alpha \left(\frac{1}{r'_{i\alpha}} - \frac{1}{r_{i\alpha}} \right) \middle| \Psi \right\rangle + \sum_A \sum_{\alpha \in m} q_\alpha Z_A \left(\frac{1}{r'_{A\alpha}} - \frac{1}{r_{A\alpha}} \right) + \Delta E^{\text{vdW}}_{\text{s-S}} + \Delta E^{\text{MM}}. \tag{3.4}$$

The change of the energy depends only on the wave function of the previous step. Consequently, no SCF cycle has to be performed and only a few one-electron integrals have to be evaluated which considerably reduces the computational costs. The first term describing the electrostatic interaction between the QM and MM subsystem can be expressed in the dependence of the density matrix by

$$\Delta E^{elec.}_{\text{QM/MM}} = \sum_{\mu\nu} P_{\mu\nu} \left\langle \mu \middle| \sum_{\alpha \in m} -q_\alpha \left(\frac{1}{r'_{i\alpha}} - \frac{1}{r_{i\alpha}} \right) \middle| \nu \right\rangle \tag{3.5}$$

with the density matrix elements $P_{\mu\nu}$ and the basis functions $|\,\mu\rangle$ and $|\,\nu\rangle$. The number of one-electron integrals therefore scales with $m \times k^2$ for k basis functions.

In the conventional QM/MM formulation the Hamiltonian $H_{\text{QM/MM}}$ describing the interaction between QM and MM systems is tightly coupled to the Hamiltonian of the QM part H_{QM}. Any change in a degree of freedom requires recomputing both terms which includes costly SCF cycles until a fully converged wave function is obtained. The approximation introduced only for the $H_{\text{QM/MM}}$ operator allows to decouple these terms. It simplifies the description of the interaction between the QM and MM system to a two-body potential. The change of the energy is therefore independent of the size of the total MM system and depends only on a single molecule. It also allows to independently sample the configuration space of the QM and MM systems. This is especially important because the number of solvent degrees of freedom is considerably larger than the solute one's. Extensive sampling is required to obtain well converged and statistically relevant results.

A slightly different route has been developed by Gao and coworkers [87] which is based on the gas phase wave function. In effect, the wave function does not have to be optimised during the simulations and with embedding potential at any given moment. As a consequence, the perturbation approach cannot be restricted to first order but the expansion has been derived up to second order. In every step the molecular electrostatic potential $V(\boldsymbol{R})$ from the QM molecule at the position $\boldsymbol{R}_m$ of the MM charges q_m has to be computed. This allows computing the vertical interaction energy which is the first order perturbation energy

$$E^{(1)} = \sum_{m=1}^{M} q_m V(\boldsymbol{R}_m) \tag{3.6}$$

for M charge sites. $V(\boldsymbol{R}_m)$ is explicitly given as a function of the gas phase density matrix $P^0_{\mu\nu}$ by

$$V(\boldsymbol{R}_m) = \sum_{a=1}^{A} \frac{Z_a}{R_{ma}} - \sum_{\mu,\nu} P^0_{\mu\nu} \left\langle \mu \left| \frac{1}{|\boldsymbol{R}_m - \boldsymbol{r}_1|} \right| \nu \right\rangle . \tag{3.7}$$

The electronic polarisation energy approximated by the second term of the perturbation can be computed as

$$E^{(2)} = \sum_{i}^{\text{virt}} \sum_{j}^{\text{occ}} \frac{1}{\epsilon_i - \epsilon_j} \left(\sum_{\mu,\nu} c_{\mu i} c_{\nu j} \sum_{m=1}^{M} \left\langle \mu \left| \frac{1}{|\boldsymbol{R}_m - \boldsymbol{r}_1|} \right| \nu \right\rangle \right)^2 \tag{3.8}$$

with the indices i and j running over occupied and virtual space and the orbital coefficients $c_{\mu i}$. The advantage of this particular formulation of the perturbation terms is that only the gas phase wave function is needed which can be computed once at the start of the simulation. This approach has been termed the Generalized Molecular Interaction Potential with Polarization Correction (GMIPp). Studies of hydrogen bonded interactions showed that the polarisation energy deviates by about 3% from the exact result. In order to reduce the computational costs further the orbitals considered for the polarisation energy can be truncated with an energy cut-off criterion (e.g. 3 E_h).

A closely related approach by Pulay and coworkers [88] computes the first order term in the same manner but switches for more distant solvent sites to a multipole representation for the solute in order to increase the computational efficiency. In a preparation phase the generalised multipoles and polarisibilities are determined which are used subsequently in the simulation. For the evaluation of the electrostatic polarisation term the electrostatic potential from the solvent in the solutes volume has to be evaluated on a grid which is used together with the predetermined polarisibilities to approximate the second order term. An expansion of the electrostatic potential in a Fourier series leads to numerical stability and allows the efficient evaluation of the second order term. First benchmarks lead to speed-ups of more than four orders of magnitude with an average error of 1–2 kcal/mol but an extension to PBC has not been brought forward until now.

Both of these approaches require an unnecessarily high order of the perturbation because they suffer from the choice of the gas phase wave function as their unperturbed reference which is a completely unpolarised wave function. However, the computational costs of converging the wave function embedded in the partial charges is comparable to the gas phase calculation even for thousands of solvent molecules. Ultimately, our choice of reference is closer to the target function requiring much smaller corrections. This means that the perturbation due to a solvent move is considerably smaller and can therefore be accurately described by first order perturbation theory. Furthermore, the limit of the full QM/MM simulations can be easily recovered in the PMC approach by simply increasing the number of updates. However, in the approaches based on a gas phase wave function the only way to improve the accuracy is by increasing the order of the perturbation expansion which is not guaranteed to converge.

3.1.1 Limits of the Perturbation Approach

The same wave function could indeed be used for a complete MC simulation which would correspond to computing the density for the initial configuration and freezing it subsequently. However, the error introduced by using first order perturbation theory accumulates throughout the simulation. In order to limit this error it is advised to update the wave function in regular time intervals. Initial studies used a very conservative criterion with updates being carried out for every step inside a cut-off radius of about 5 Å around the QM part and at least every 20 steps. [84, 85]

It has been found later that even for very small ion-water clusters with only one or two water molecules the enthalpies of binding are very stable for up to 2000 perturbative steps and the deviation is on the order of magnitude of the statistical error of those simulations. [89] For every update the energy can be computed with the old as well as the new wave function, which allows estimating the error of the perturbation theory

$$\sigma_{\text{PMC}} = E(P_{\mu\nu}^{\text{previous}}) - E(P_{\mu\nu}^{\text{current}}). \tag{3.9}$$

This error increases consistently with more perturbative steps between successive updates. It has also been found that the errors for anions are larger than for cations which shows that more updates are required for solutes with a larger polarisability. [86] In liquid systems with a much larger number of solvent molecules the number of steps between two updates can be increased considerably. This will be further investigated in Section 4.2.

While it has been obvious from the beginning that the error of the perturbation approach depends on the distance to the solute molecule, the approximate dependence can be derived from classical electrostatics. The electric field $\boldsymbol{F}$ that a solvent molecule experiences due to the presence of the solute is given by

$$\boldsymbol{F} = \sum_i \frac{q_i}{r_i^3}\boldsymbol{v}_i \tag{3.10}$$

with the distance r between solute and solvent and the unit vector $\boldsymbol{v}$ along this distance. The induced dipole moment $\boldsymbol{\mu}$ of the solvent molecule is given as the product of the electric field and the polarisability $\tilde{\alpha}$

$$\boldsymbol{\mu} = \tilde{\alpha} \cdot \boldsymbol{F} \tag{3.11}$$

and the resulting energy due to the polarisation is

$$E_{\text{pol}} = -\frac{1}{2}\boldsymbol{F} \cdot \tilde{\alpha} \cdot \boldsymbol{F}. \tag{3.12}$$

The error (Equation 3.9) can be seen as the change of the energy with respect to the position of a solvent molecule

$$\sigma_{\text{PMC}} = \frac{\Delta E}{\Delta r}. \tag{3.13}$$

Inserting Equation 3.12 and considering the limit of $r \to 0$ the distance dependence is approximately given by

$$\sigma_{\text{PMC}} \propto \frac{1}{r^7}. \tag{3.14}$$

This reveals a very short-range nature of the error which is consequently rather insensitive to the solute itself. This formal distance dependence has been used to define a weighting function to estimate the error of the perturbative approach and reduce the number of updates. It has been found that the number of perturbative steps can be increased up to about 2000 steps to guarantee an error in the total electronic energy as small as 0.01 kcal/mol for $\text{Na}^+(\text{H}_2\text{O})_{125}$. This error is well below any statistical accuracy commonly reached by simulations and the number of water molecules — while larger than previous examples — is still an order of magnitude smaller than typical simulations in solution. [90]

3.1.2 Computation of the Electrostatic Perturbation Term

The computation of the electrostatic perturbation term (Equation 3.5) requires the evaluation of a sum of one-electron integrals. These are also part of every conventional QM calculation and are generally known as three-center nuclear attraction integrals:

$$\int \phi_i\left(\boldsymbol{A}, \alpha_A, n, l, m\right) \frac{1}{r_q} \phi_j\left(\boldsymbol{B}, \alpha_B, n, l, m\right) \mathrm{d}V. \tag{3.15}$$

with Gaussian functions at the centres **A** and **B** and a point charge at the centre **Q**. The Gaussian functions are defined as

$$\phi\left(\boldsymbol{A}, \alpha_A, n, l, m\right) = n_A(n, l, m, \alpha_A)(x - x_A)^n (y - y_A)^l (z - z_A)^m \exp\left(-\alpha_A r_A^2\right) \tag{3.16}$$

with the normalising constant

$$\begin{aligned} n_A(n,l,m,\alpha_A) &= n_A(n,\alpha_A)n_A(l,\alpha_A)n_A(m,\alpha_A) \\ &= \prod_{k\in n,l,m} \left(\frac{2\alpha_A}{\pi}\right)^{\frac{1}{4}} (4\alpha_A)^{\frac{k}{2}} \left[(2k-1)!!\right]^{-\frac{1}{2}} . \end{aligned} \tag{3.17}$$

These functions are categorised according to the sum $L = n + l + m$: $L = 0 \rightarrow$ s, $L = 1 \rightarrow$ p, $L = 2 \rightarrow$ d, and so forth. I will show as an example the evaluation for two s functions.

The inverse distance operator $\frac{1}{r_q}$ commutes with the Gaussian functions. This allows building the product of the two Gaussians which, according to the Gaussian product theorem, is again a Gaussian function

$$n_A(n,l,m,\alpha_A)\exp\left(-\alpha_A r_A^2\right) n_B(n,l,m,\alpha_B)\exp\left(-\alpha_B r_B^2\right) = n_P \exp\left(-\alpha_P r_P^2\right) \tag{3.18}$$

with $\alpha_P = \alpha_A + \alpha_B$ and the combined normalising constant

$$n_P = \exp\left(\frac{\alpha_A \alpha_B}{\alpha_P} \left|\boldsymbol{A} - \boldsymbol{B}\right|^2\right) n_A n_B. \tag{3.19}$$

The centre **P** of this Gaussian is on a line between the two original centres

$$\boldsymbol{P} = \frac{\alpha \boldsymbol{A} + \beta \boldsymbol{B}}{\alpha_P}. \tag{3.20}$$

Boys [91] derived that an integral of this type can be evaluated as

$$\int \frac{1}{r_q} \exp\left(-\alpha_P r_P^2\right) \mathrm{d}V = \frac{2\pi}{\alpha_P} \int\limits_0^1 \exp\left(-T u^2\right) \mathrm{d}u \tag{3.21}$$

with the argument T

$$T = \alpha_P \left|\boldsymbol{Q} - \boldsymbol{P}\right|^2 . \tag{3.22}$$

In practice the integral is evaluated with different approaches according to the value of T, which are summarised in Table 3.1. Recursive formulas can be derived for Gaussians with $L > 0$ which allows the evaluation based on the here presented formulas. A more recent investigation established an optimized algorithm for the evaluation of the Boys functions which achieved an up to 19% reduced number of floating point operations. [92]

TABLE 3.1 Computation of the integral (eq. 3.21) for different values of T in atomic units. The values $F_k(T^*)$ are precomputed and tabulated.

T	$F_0(T)$
$T \leq 10^{-16}\ a_0$	$F_0(T) \approx 1$
$10^{-16}\ a_0 < T \leq 10\ a_0$	$F_0(T) \approx \sum_{k=0}^{6} \frac{(T^*-T)^k}{k!} F_k(T^*)$
$10\ a_0 < T \leq 34\ a_0$	$F_0(T) \approx \frac{1}{2}\sqrt{\frac{\pi}{T}} - \frac{1}{2T}\exp(-T)$
$T > 34\ a_0$	$F_0(T) \approx \frac{1}{2}\sqrt{\frac{\pi}{T}}$

Difficulties arise in evaluating these integrals when going from cluster models to periodic systems. Initial studies assumed that the electrostatic interaction decays quickly enough with the distance. Under the condition that the integrals evaluate approximately to zero for large r_Q a rigorous cut-off can be introduced. Although this may be true for the short-range van der Waals (vdW) interactions, the electrostatic interactions decays formally only with $\frac{1}{r}$. Consequently, the potential approaches only very slowly zero. Therefore, using a cut-off leads to interface effects. With this in mind a shifted force operator will be used as it has been introduced in Section 2.2.3. The downside is that such an operator cannot be integrated analytically. The resulting integrals are not over all space but only over the volume of the cut-off centred on the partial charges and not on the atoms. Therefore, the resulting space is not symmetric with respect to the basis functions. This leads also to the situation where parts of the electron density are influenced by a point charge while others do not feel its influence at all.

Consequently, a numerical integration scheme has been used. Based on established approaches of DFT as presented in Section 2.1.3 the density is first evaluated on a numerical grid and partial charges are generated by multiplying with the weights of the grid points. Finally, these partial charges are used to compute the interaction with the molecule that has been moved during a MC step. The accuracy of this numerical integration will be investigated further in Section 4.1. However, the shifted operator is now only applied to compute the interaction between a grid point and a partial charge of a solvent molecule. This means that the interaction computed during the perturbation steps is not consistent with the density and interaction computed in the QM calculations. In other words, when the electronic density is generated the solute is embedded in unscaled charges and the shifted operator is used to compute the energies after the SCF cycles have been already converged. It is expected that this has only a minor effect on the results. In order to obtain consistent results the shifted operator would have to be applied during the SCF cycles of the QM program. This can be straightforwardly implemented by building the Coulomb operator in the same grid used for the QM/MM run.

3.2 Implementation in Hybrid Architectures

The PMC method is a hybrid QM/MM method. The classical and QM calculations have very different computational requirements, different algorithms and use different data structures. With this in mind it is clear that no single computer architecture can perfectly match all these requirements and that on the level of the hardware a hybrid approach has to be implemented as well. After the initial implementation of a reference algorithm working on conventional CPUs only a solid framework and the basic PMC implementation has been developed for a hybrid GPU-CPU algorithm in close collaboration with Sebastião Miranda from the group of Pedro Tomás [83] on which all further extensions are based. The program is developed in C++11 and interfaces a development version of the Molpro program suite [93] for QM calculations. The GPU functionality has been implemented in OpenCL and is therefore neither restricted to a certain hardware vendor nor to graphic cards as the only possible accelerator.

The structure of the algorithm is shown schematically in Figure 3.1. After the initial charge

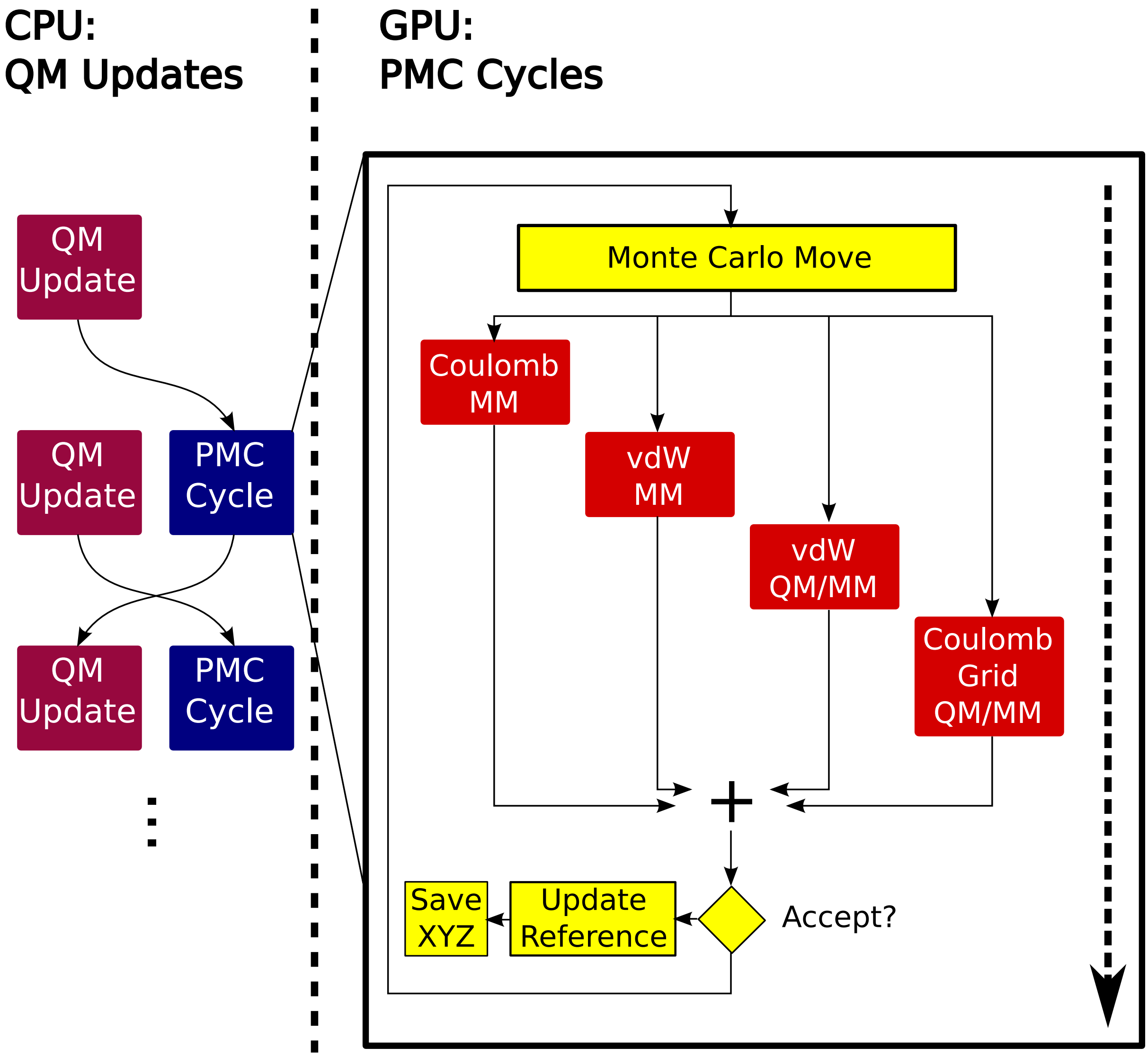


FIGURE 3.1 Schematic representation of the PMC algorithm with the functions computing the Coulomb and vdW interaction of the MM system and the Coulomb and vdW interaction between QM and MM system.

density is constructed, the first cycle of perturbative MC steps can be executed. At first a MC step is proposed and subsequently all changes of the energy terms are computed which includes the change of the Coulomb and vdW energy of the MM part as well as the coupling term. The QM/MM Coulomb term includes the interaction of the charge density as well as the charge of the nuclei with the partial charges describing the solvent molecules. These contributions are accumulated and the move is either rejected or accepted. Averages are computed on the fly and in regular intervals the energies as well as the configuration can be saved to a file for later analysis. Finally, another charge density is generated by the QM program suite and this procedure repeats until the desired number of steps have been carried out.

Based on the CPU implementation, timings have been obtained for a capped arginine cation in the QM system embedded in 1301 classically treated water molecules (Table 3.2). They show clearly that a large majority of the time is spent on computing the electrostatic coupling term between QM and MM part. The large number of data points of the numerical grid can be perfectly exploited by a fine grained data level parallelism on the GPU which is reflected by the

TABLE 3.2 Kernel execution times of a CPU (i7-4770K) reference implementation and a GPU (GTX 780Ti) implementation.

Kernel	CPU		GPU		Speed-up
Step Generation	32 μs	0.04 %	17 μs	3.0 %	1.88
$\Delta E^{\mathrm{elec}}_{\mathrm{S-s}}$	76077 μs	98.80 %	473 μs	83.3 %	160.84
$\Delta E^{\mathrm{MM}}_{\mathrm{s}}$	791 μs	1.03 %	40 μs	7.0 %	19.77
$\Delta E^{\mathrm{vdW}}_{\mathrm{S-s}}$	4 μs	0.01 %	18 μs	3.2 %	0.22
Decision	94 μs	0.12 %	20 μs	3.5 %	4.70
Total	76998 μs	100.00 %	568 μs	100.0 %	135.55

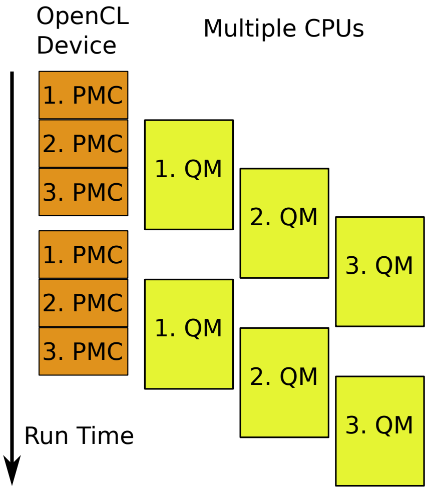


FIGURE 3.2 Schmematic representation of the PMC algorithm with multiple Markov chains being generated in parallel.

speed-up of about 160 compared to the CPU implementation. The subsequent developments showed, however, that it is essential to reduce the overhead resulting from the communication between the CPU and GPU. Therefore, the complete cycle of perturbative steps between two subsequent QM calculations has been implemented with OpenCL. While some parts of the Metropolis MC algorithm are inherently serial and not well suited for GPUs as it can be seen from the five times slow-down of the vdW QM/MM term the overall speed-up obtained is still significant. Furthermore, the different energy terms can be computed independently and exploit a task grained parallelism. This leads to an overall speed-up of about 135 for a whole cycle of perturbative steps.

While a new charge density is generated by the QM module the simulation cannot continue until this task is completed. In order to avoid this bottleneck a coarse-grained parallelism at the Markov chain level is exploited. The memory-less property of a Markov chain allows to carry out several shorter simulations which are subsequently combined instead of a single long simulation. Therefore, multiple charge densities are generated on the available CPU cores while the GPU carries out the perturbative steps (Figure 3.2). This overlapping execution guarantees that all available resources are utilised as efficiently as possible.

In the event that the GPU modules become the bottleneck, additional graphics cards can be included in the calculation. Therefore, a performance aware load-balancing scheme has been implemented which divides the grid points for the electrostatic QM/MM term for the different devices and guarantees after a couple of iterations that all GPUs need about the same time to compute the energy for the assigned grid points. Therefore, the implementation scales very well with the number of CPU cores — which is often unfavourable for conventional QM methods — as well as GPU devices.

A more practical measurement shows that this allows the calculation of about 500 perturbative steps per second for a small test system (6 QM Atoms, 1494 MM atoms). The run time of a whole set of simulations consisting of 25 times 24.8 M PMC steps can be reduced from 283 days on a standard workstation to about 2 days with OpenCL acceleration. This alleviates the current restriction of conventional QM/MM methods with regard to the time scale of simulations and allows sufficient sampling of the configuration space which can otherwise only be achieved by a substantial amount of computational resources.

3.3 Additional Modules of the PMC Program Suite

Further functionalities and details of the implementation of the PMC program suite which extend its applicability, increase the usability and efficiency and allow a number of analyses are outlined in this Section. A schematic overview of all modules is given in the mindmap in Figure 3.3. One example are the classical simulations, which could be run with other program packages but many times technical details are different, files are not compatible and the set-up has to be carried out for two program packages. Here, the simulations can be carried out for initial equilibration or to diverge the chains from a common starting structure and conveniently continued with PMC.

In order to take advantage of the existing set of Tinker tools version 7.1 [94] the file format for geometries is specified in the tinker xyz file format and is also saved automatically in the same except for trajectories which are written directly in the standard xyz format reducing the amount of data. Furthermore, force field parameters can be specified in a native PMC format but also standard Tinker parameter files can be read directly. Therefore, solutes can be prepared and solvated with the program xyzedit, subsequently optimized and then directly used to start a simulation with the PMC program.

Different levels of theory and coordinates systems can be employed to run MC simulations. QM can be used in gas phase, in combination with continuum models or also for clusters and the sampling is carried out in cartesian coordinates. MM and PMC can be used with rigid molecules in combination with PBC. In that case only the intermolecular degrees of freedom are sampled. The three-centre nuclear interaction integrals for the perturbative steps without periodic boundary conditions — therefore without shifted force operator — have been implemented in a stand-alone version in the programming language C during my master thesis. [95] This module was directly included in the PMC program and allows an extremely fast evaluation without the overhead of calling a QM suite which makes it about three orders of magnitudes faster. Besides the electron density it requires additional information about the basis set and molecular orbitals from the QM program.

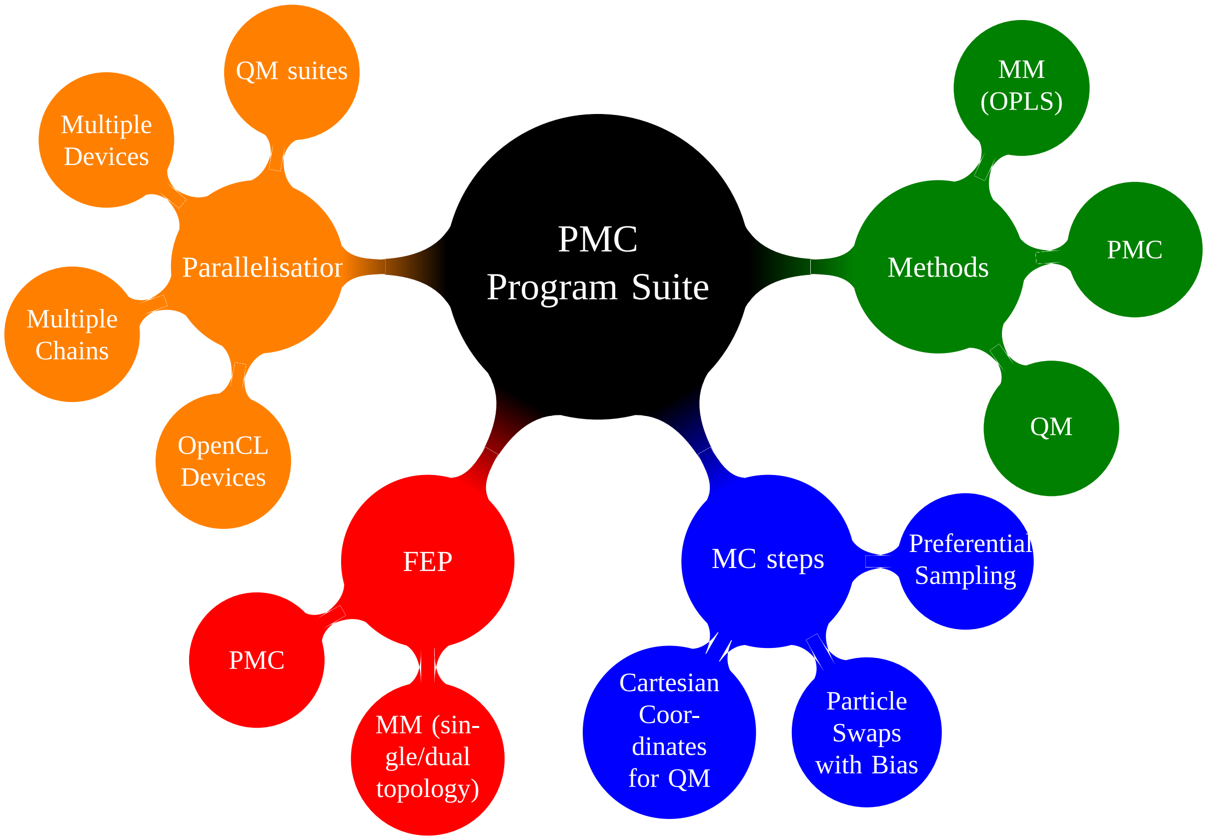


FIGURE 3.3 Mindmap of modules and functionalities included in the PMC program suite.

Free energy perturbation theory can be used in combination with the MM as well as the PMC module. For the classical simulations the single as well as the dual topology approach have been implemented, meaning whether or not the initial and final states are described in the same topology. It has been argued that the single topology is always advantageous and at worst equivalent to the dual topology approach. [96] However, there are techniques that can only be used in the latter and they can increase the efficiency considerably. [97] For the computation of free solvation energies the vdW interactions between solute and solvent are turned off stepwise. Numerical instabilities at the endpoints of this step require the use of soft-core potentials. [98] These potentials explicitly depend on the alchemical variable λ but they are identical to the vdW potential for $\lambda = 0$ and $\lambda = 1$

$$E_{\text{vdW}} = 4\epsilon\lambda^n \left[\left(\alpha\left(1-\lambda\right)^m\right) + \left(\frac{r}{\sigma}\right)^6 \right)^{-2} - \left(\alpha\left(1-\lambda\right)^m\right) + \left(\frac{r}{\sigma}\right)^6 \right)^{-1} \right] \tag{3.23}$$

with the common values for the parameters $m = n = 1$, $\alpha = 0.5$ — which can be adjusted if required — while ϵ and σ define the vdW potential.

The classical energy contributions needed for the pure MM as well as the PMC simulations are carried out in reduced units. The natural length unit under PBC is the box length L so that all distances are expressed as multiples of the box length $r_{\text{red}} = \frac{r}{L}$. Subsequently, according to the minimum image convention the distances are computed as

$$\tilde{r} = r_{\text{red}} - \text{round}(r_{\text{red}}) \tag{3.24}$$

with round returning the nearest integer. Using reduced units avoids the division by the box length every time a distance is computed. To be able to use reduced units directly for the computation of the energy with the force field the parameters of the vdW σ and Coulomb q_i term as well as any cutoffs r_c are transformed according to the following equations

$$\tilde{\sigma} = \frac{\sigma}{L}, \tag{3.25}$$

$$\tilde{r}_c = \frac{r_c}{L}, \tag{3.26}$$

$$\tilde{q}_i = \frac{q_i}{\sqrt{L}}. \tag{3.27}$$

Constants which are required for the computation of the shifted force potential e.g. the reciprocal cutoff r^{-1} and r^{-2} are precomputed. Only for the QM calculation step (which is carried out without PBC) the coordinates are transformed back and translated into the central unit cell.

The force field computations have been optimized to some degree in order to exploit memory locality and consequently to reduce L3 cache misses as measured by the tool Cachegrind which is part of the program package Valgrind. [99] Exploratory studies showed that the limiting factor due to the system size is the memory bandwidth and it turned out that for example the computation of all $\epsilon_i \epsilon_j$ or $q_i q_j$ products (for the vdW and Coulomb term respectively) at the beginning of the simulation reduced naturally the number of floating point operations but slowed down the simulations by requiring even more data to be loaded from memory.

Special attention has been paid to the computation of the distances not only because most of the CPU time is spent on it but also because different implementations can vary widely in their memory footprint. The memory layout of these is shown in Figure 3.4 while on the GPU it is always advantageous to compute the distances on-the-fly. The first naive implementation, storing the complete distance matrix, turned out to use too much memory for large systems in the context of multi-chain simulations. However, this implementation has advantages as well. Namely, trivial access to the distance r_{ij} as well as r_{ji} is guaranteed and the distances are sequentially stored which allows efficient access due to data prefetching. Using only one triangle of the matrix can speed-up the calculations because the limiting factor has been the memory bandwidth. Furthermore, the memory footprint can be reduced by mapping one triangle of the matrix onto a vector with the index v defined according to

$$v = i + (j-1)\frac{j}{2}. \tag{3.28}$$

The correct order of loops accessing the distances should be guaranteed in order to read them as much as possible sequentially.

A flexible implementation could be easily realised by using inheritance and overloading a function e.g. `getDistance(i, j)` but virtual function calls are associated with a cost and this should be avoided in this case of a low-level function that is called many times and has only a very short execution time. Furthermore, many times composition should be preferred over inheritance. [100] In order to achieve the required polymorphism the delegation pattern is used and a very fast implementation of delegates has been presented based on templates by Ryazanov

	(1,0)	(2,0)
(0,1)		(2,1)
(0,2)	(1,2)	

	(1,0)	(2,0)
		(2,1)

(1,0)
(2,0)
(2,1)

FIGURE 3.4 Schematic representation of the memory layout of different algorithms storing the distances. Left: full matrix without diagonal, middle: only the upper triangle, right: the upper triangle mapped onto a vector.

[101] which has been subsequently improved and modified for C++11 by Kryukov. [102] This allows a good trade-off between flexibility and efficiency.

The random numbers which are essential for MC simulations are generated by a Pseudo Random Number Generator (PRNG) which can be easily changed should this be required for certain applications. The default is set to the 64bit Mersenne Twister developed by Matsumoto and Nishimura. [103] In the current implementation the random numbers are generated on the CPU and then transferred for the perturbative steps to the GPU. However, a more recent implementation of the Mersenne Twister specifically for GPUs has been developed and could be used instead. [104] The initialisation of the PRNG is done with true random numbers from the operating systems which might fall back to pseudo random numbers if the entropy pool is exhausted. The state of the PRNG is saved for any completed simulation and can be loaded to conveniently continue a simulation.

Commonly required ensemble averages are computed on-the-fly with an updating algorithm avoiding numerical instabilities especially for variances as developed by LeVeque and coworkers. [105] The analysis of simulations in order to compute changes of free energies is carried out by the Python library pymbar. [106] The computation of the free energy change over several windows specified by an xyz-trajectory is automated and one or both directions for the integration can be chosen. The wrapper for pymbar includes algorithms to analyse time series [107] which allow to select a subset of uncorrelated samples — a requirement for the statistical analysis and especially the error estimation. A generic framework allowing the analysis of trajectories has been developed and can be easily extended for further properties. Implementations to generate Radial Distribution Functions (RDFs) as well as angular RDFs and to compute displacements and dipole moments are included.

The library Libconfig [108] has been used to read configuration as well as input files. It is a library to parse structured configuration files which are, however, more compact than xml files making them especially suitable for input files. Additionally, it is type aware which adds an extra layer of automatic checks for the input that does not have to be implemented and avoids tedious string parsing. All logic that constructs run-time parameters based on the input or available hardware is cleanly separated and crashes loudly upon encountering errors in order to avoid spending time on faulty simulations. This module could be separated into a stand-alone module for a queuing system in order to check for errors before even submitting any calculation. Error messages include the line number of the input where the error has been encountered and the type of error which is again facilitated by the library Libconfig.

CHAPTER 4

Benchmarks

The PMC approach depends mostly on two parameters. The first one is the size of the grid employed for the numerical integration. The second one is the frequency with which the electronic density is updated with QM calculations. In this Chapter the influence of these parameters will be assessed. Thereby, guidelines will be established which allow carrying out simulations at the level of full QM/MM simulations but at considerably reduced computational costs. The benchmark systems have been selected either due to their low computational cost which makes them accessible for carrying out many simulations with varying parameters or due to the available experimental and theoretical reference data.

These findings will be used in the second half where I will establish that this QM/MM approach is capable to reproduce – at least qualitatively – experimental findings and captures the essential physics which are needed to describe solute-solvent interactions. Free energies of solvation as well as the free energy torsional potential of hydrogen peroxide will be computed by means of PMC simulations.

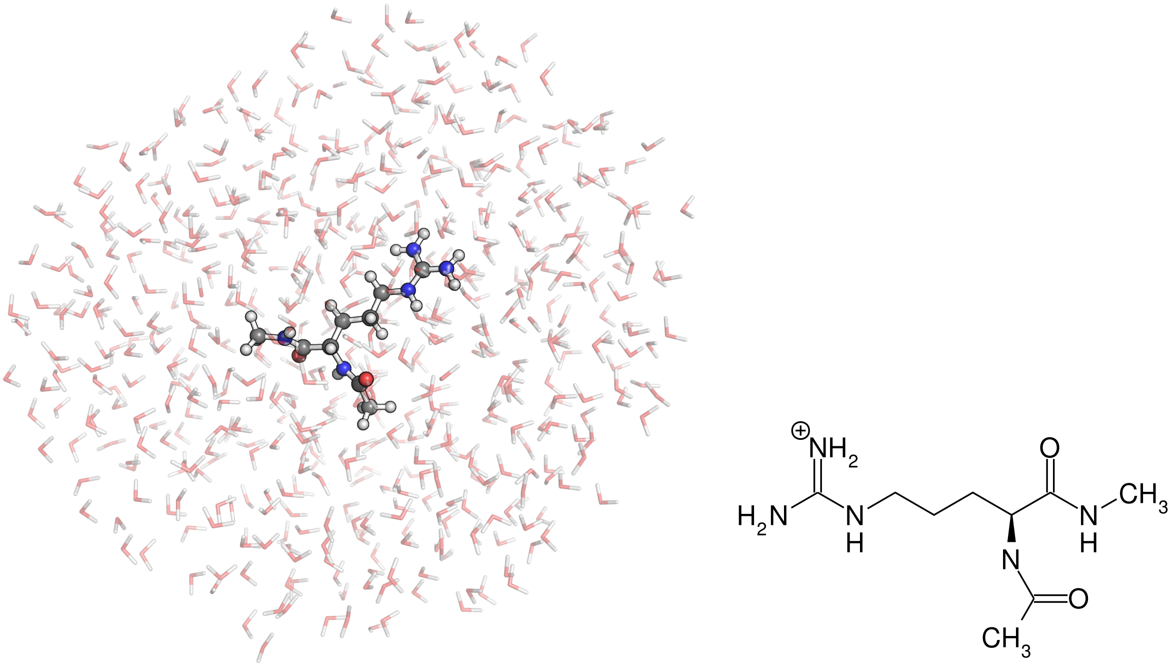
4.1 Numerical Integration

A discrete representation of the electronic density is obtained on a radial grid. This allows integrating over the density with a numerical integration scheme. Different ways of constructing the radial integration grid have been discussed in Section 2.1.3. The convergence of Becke's [23] and Mura and Knowles' [27] variants has been compared for several different grid sizes. The actual number of grid points used in the radial integration are determined by a dynamic algorithm in the Molpro program package [93] according to a specified target accuracy. This accuracy would be obtained for the integration of the Slater-Dirac functional. Eighty-two snapshots of a PMC simulation of a capped arginine cation in TIP3P water have been used for the analysis. A sphere of solvent molecules has been constructed including all molecules with at least one atom inside a radius of 12 Å. A representative snapshot as well as the Lewis structure of the solute are shown in Figure 4.1. This system without PBC has been deliberately constructed to compare with the analytical integrals. The energies have been computed at the PBE/def2-SVP [110, 111] level of theory.

The Root Mean Squared Deviation (RMSD) of the electrostatic interaction relative to the

---

The material in this Chapter was presented in part in Reference [109].



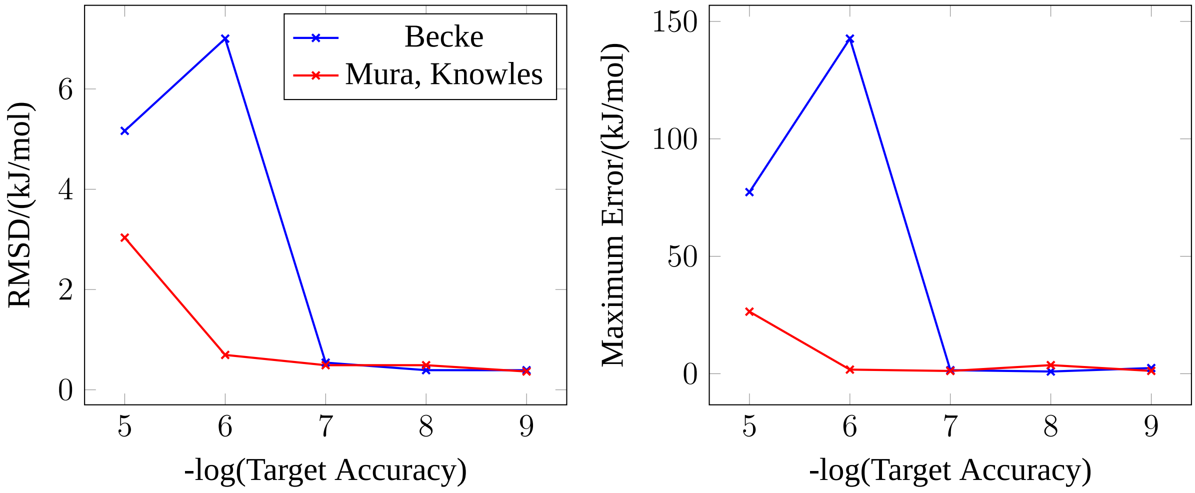
(a) Snapshot with solvent.

(b) Lewis structure of solute.

FIGURE 4.1 Arginine cation with NMe cap at the C terminus and acetate cap at the N terminus with a sphere of TIP3P water.

analytical results are shown in Figure 4.2a. Given a very large number of grid points corresponding to a very high target accuracy both grids converge to the exact result. However, the Mura and Knowles grid converges faster and leads to a RMSD value of only about 1 kJ/mol for a target accuracy of 10^{-6} $\mathrm{E_h}$. One should keep in mind that the target accuracy is for the Dirac-Slater functional and does not translate directly to the accuracy for the electrostatic interaction. This is in average about -400 kJ/mol for this system. Therefore, the deviation of 1 kJ/mol is less than 1%. The general convergence pattern also appears more consistent for the grid by Mura and Knowles (Figure 4.2b). This confirms that this grid exhibits a balanced description for the nuclei, the bonded and the long-range region which becomes in this benchmark apparent in the more robust behaviour with regard to changes in the configuration of the solvent partial charges as represented especially by the lower maximum deviation. Furthermore, this grid is the default grid for conventional DFT calculations in Molpro [93] which confirms that the accurate description of the electronic density is reflected in the accurate description of the electrostatic interaction.

A considerable advantage of the PMC approach are the immense computational savings which allow extensive sampling at the accuracy of full QM/MM simulations. Therefore, it is very important to ascertain that indeed the same configuration space is sampled with the chosen numerical grid. The numerical grid might add noise to the exact results and while the average deviation in the energy is very low the sampling of the configuration space might be affected more severely given that not a fixed set of configurations is studied. In MD two simulations with identical initial conditions are required to result in the same trajectories. However, in MC simulations this is not required due to the stochastic nature of the algorithm. In fact, the



(a) RMSD

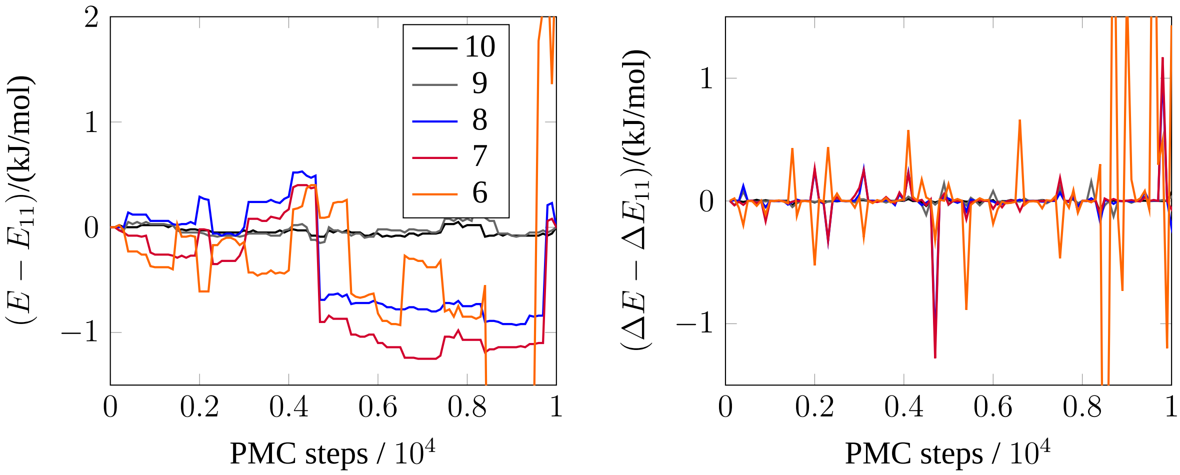
(b) Maximum error

FIGURE 4.2 Electrostatic interaction for different grids with the exact results as a reference for snapshots of a PMC simulation of a capped arginine cation in TIP3P water at the PBE/def2-SVP level of theory.

same trajectories can be only obtained because a pseudo random number generator is used. Nevertheless, we will assess our simulations by this criteria although it is not strictly required. Furthermore, the potential energy as well as the change of the potential energy can be monitored throughout a simulation. Ideally, the error in the potential energy should not accumulate during the whole perturbative cycle. However, only the change of the potential energy in every step is used for the MC acceptance criteria and it is important that the error is constant in this value.

Since the convergence for very large grids has been established a grid with 10^{-11} $\mathrm{E_h}$ target accuracy. The latter will be used as the reference which allows analysing an actual simulation under PBC. Results for a series of simulations with decreasing target accuracy up to 10^{-6} $\mathrm{E_h}$ are represented in Figure 4.3a. The QM solute ethanol has been simulated for one PMC cycle in 498 molecules of TIP3P water in a box with the edge length of 24.6 Å at the temperature of 298 K. The energies have been computes at the B3LYP-D3/def2-TZVP [17, 20–22, 111–114] level of theory and with OPLS-AA [41] vdW parameters for the solute. The deviation of the absolute potential energy from the reference has been computed for every step. For 10^{-10} and 10^{-9} $\mathrm{E_h}$ target accuracy in the grid the error of the potential energy for every structure stays well below 0.5 kJ/mol. This deviation is only caused by the differences in the electrostatic interaction so that it is about 1% of the electrostatic interaction. Only for 10^{-8} and 10^{-7} $\mathrm{E_h}$ the error increases to about 1.2 kJ/mol in the second half of the PMC cycle. It is also apparent that the error starts to accumulate throughout the simulation. For all of these simulations a single step associated with a large error leads to the larger error in the potential energy in the second half of the simulation. However, it can be also seen in the case of 10^{-8} and 10^{-7} $\mathrm{E_h}$ target accuracy in the grid that the error can be reduced again. This can be either due to the same step but in the reverse direction or due to a different step with an error of the opposite sign.

Moreover, it should be noted that the MC algorithm uses only the change of the energy for the acceptance criteria of a step. The error in the differences is shown in Figure 4.3b and it can be seen that only a few steps have an error larger than 1 kJ/mol. While it is not a strictly



(a) Energy

(b) Change of energy

FIGURE 4.3 Comparison of relative potential energies of PMC simulations with B3LYP-D3/def2-TZVP and OPLS-AA parameters of the same trajectory with different grid sizes. $-\log(\mathrm{Target\ Accuracy})$ is given in the legend with the reference being 11.

necessary criterion that different MC chains do not diverge it is a sufficient one. Only the simulation with 10^{-5} target accuracy in the grid diverges at about 8200 steps and it appears to have suddenly much larger errors. This is because it is not directly comparable to the reference which is based on a different set of configurations from that point onwards. These benchmarks show that the same configuration space is sampled even with a relatively low target accuracy in the grid of 10^{-6} $\mathrm{E_h}$ for several thousand steps. Based on these short simulations with the fine-grained analysis it appears to be only a random error, introducing no systematic bias. Next, I will investigate the influence on more complex observables e.g. the free energy that cannot be expressed as a simple ensemble average.

Simulations have been carried out and analysed with BAR in order to determine free energies of solvation. These results will be analysed in more detail in Section 4.3. Here, the free energy change due to the decoupling of the electrostatic interaction between solute and solvent will serve as the final benchmark for the accuracy of the numerical grid. Compared to the potential energy this is a very sensitive criterion because the relative free energy mainly depends on structures far from equilibrium which are rarely sampled. Any bias with regard to the exploration of the configuration space will be revealed in these simulations. The results for the previously used system — QM ethanol in TIP3P water with B3LYP-D3/def2-TZVP and OPLS-AA parameters — are shown in Figure 4.4. For up to a target accuracy of 10^{-8} $\mathrm{E_h}$ in the grid, the error in the free energies stays within 5% of the free energy of the reference simulation and does not accumulate noticeably over the four λ steps. For the target accuracy in the grid up to 10^{-6} $\mathrm{E_h}$ the error is slightly larger than 5% but only for even smaller grids a clear and systematic deviation becomes apparent.

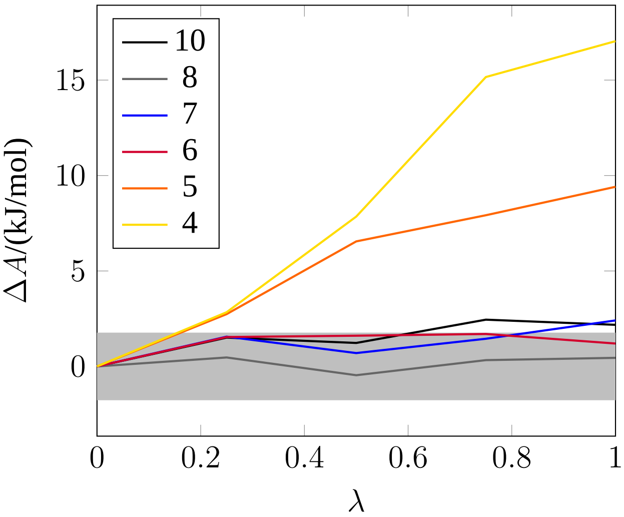


FIGURE 4.4 Decoupling of the electrostatic interaction of an free energy perturbation calculation for ethanol in water with PMC (B3LYP-D3/def2-TZP and OPLS-AA vdW parameters) with different grid sizes. $-\log(\text{Target Accuracy})$ is given in the legend and the energies are relative to 11. The shaded area denotes 5 % of the free energy (≈ 1.8 kJ/mol).

4.2 Updates of the Density

The number of QM calculations to update the electron density is the most important factor determining the computational costs. The accuracy that can be obtained for a given number of updates depends on the properties of the solute and the solvent. The limit of many updates corresponds to the full QM/MM approach with the exactly polarised density in every step which naturally comes also with the costs of a full QM/MM simulation. The limit of no updates corresponds to the frozen density approximation. A single electron density is generated for the initial configuration and used throughout the whole simulations and the polarisation effects are described only by perturbation theory. With the following benchmark set the lowest possible number of updates will be established in order to reach a given accuracy. The number of required updates naturally depends on the solute as well as the solvent which allows at this point to establish only empirical guidelines. These have nevertheless proven their value in further studies and applications that have been carried out during this work. Furthermore, for a given combination of solute, solvent, system size and QM method it might be possible to increase the number of updates and therefore the accuracy by merely using the available hardware more efficiently. This is possible due to the overlapping execution of QM and perturbative steps (see Section 3.2) as long as the QM calculations are the more time consuming portion.

Free energy calculations have been carried out and will be analysed in the following Section 4.3. Here the results for the electrostatic decoupling of a single QM water embedded in the TIP3P water model are shown in Figure 4.5. The solvent box had the edge length of 24.6 Å and contained 498 TIP3P water molecules. The simulation has been carried with a temperature of 298 K. These are the contributions to the free solvation energy that depend on the accuracy of the perturbation approach and consequently on the frequency of the density updates. Additionally, the QM system is very small, the solute as well as the solvent is polar, donating and accepting hydrogen bonds are expected between the QM and MM part which means that this system is very sensitive to any error in the electrostatic coupling term. The results show that the free solvation energy is very stable for updates carried out every 5 k–100 k steps with deviations of about 2–3 kJ/mol. This is about 10% of the free energy change for this example and of the order of magnitude of the statistical accuracy. If the number of updates is reduced even further, a systematic underestimation of the free solvation energy can be observed. The maximum error reaches about 10 kJ/mol for one update every 10 M steps which corresponds to only eight QM calculations for the whole simulation.

The systematic change of the free solvation energy can be understood in terms of the distributions of the interaction energies (Figure 4.5). The interaction energies are reduced considerably with a maximum at about -120 kJ/mol for a 5 k update frequency to about -70 kJ/mol for the 10 M update frequency. This change of 50 kJ/mol is rather surprising because at the same time the free energy of solvation changes only by about 10 kJ/mol. Further investigations are warranted into the relation between these two values. A possible explanation might be that when the density is updated it has the best possible polarisation for the interaction with the current configuration. Then the simulation proceeds, newly generated configurations are very similar and most moves occur for solvent molecules far from the solute so that the perturbation approach works very well. However, if the same density is used for many steps slow processes like exchange of solvent molecules in and rearrangements of the first solvation shell can occur and the density is not properly polarised. The perturbation approach breaks down and the interaction energy is underestimated.

Nevertheless, the free energies are remarkably insensitive even for a nearly frozen density. The reason for this is that the instantaneous difference of the potential energy which goes into the exponential averaging or BAR depends on the error made for the reference state $\sigma_{\mathrm{reference}}$ as well as the target state σ_{target}

$$\Delta E^{\mathrm{PMC}} = (E_{\text{target}} + \sigma_{\text{target}}) - (E_{\text{reference}} + \sigma_{\text{reference}}) \tag{4.1}$$

so that the error of ΔE^{PMC} is given by

$$\Delta E^{\text{exact}} - \Delta E^{\text{PMC}} = \sigma_{\text{target}} - \sigma_{\text{reference}}. \tag{4.2}$$

Given that the error of the perturbation approach is similar in the reference and target states the error in the energy difference is close to zero. If the error of both states changes similarly with the update frequency, the combined error is rather independent of the number of updates. However, the potential energy, however, which depends only on a single state cannot benefit from this fortuitous error cancellation. The analysis of the average potential energy shown in Figure 4.6 confirms these findings. The average is stable for updates carried out every 10 k–1 M steps and decreases suddenly as soon as the updates are more than 1 M steps apart by about 500 kJ/mol. This is not accompanied by a larger standard error of the determined averages. However, the standard deviation, which is the measure of the variability of the distribution, increases by about one order of magnitude. This corresponds to the different and it seems larger

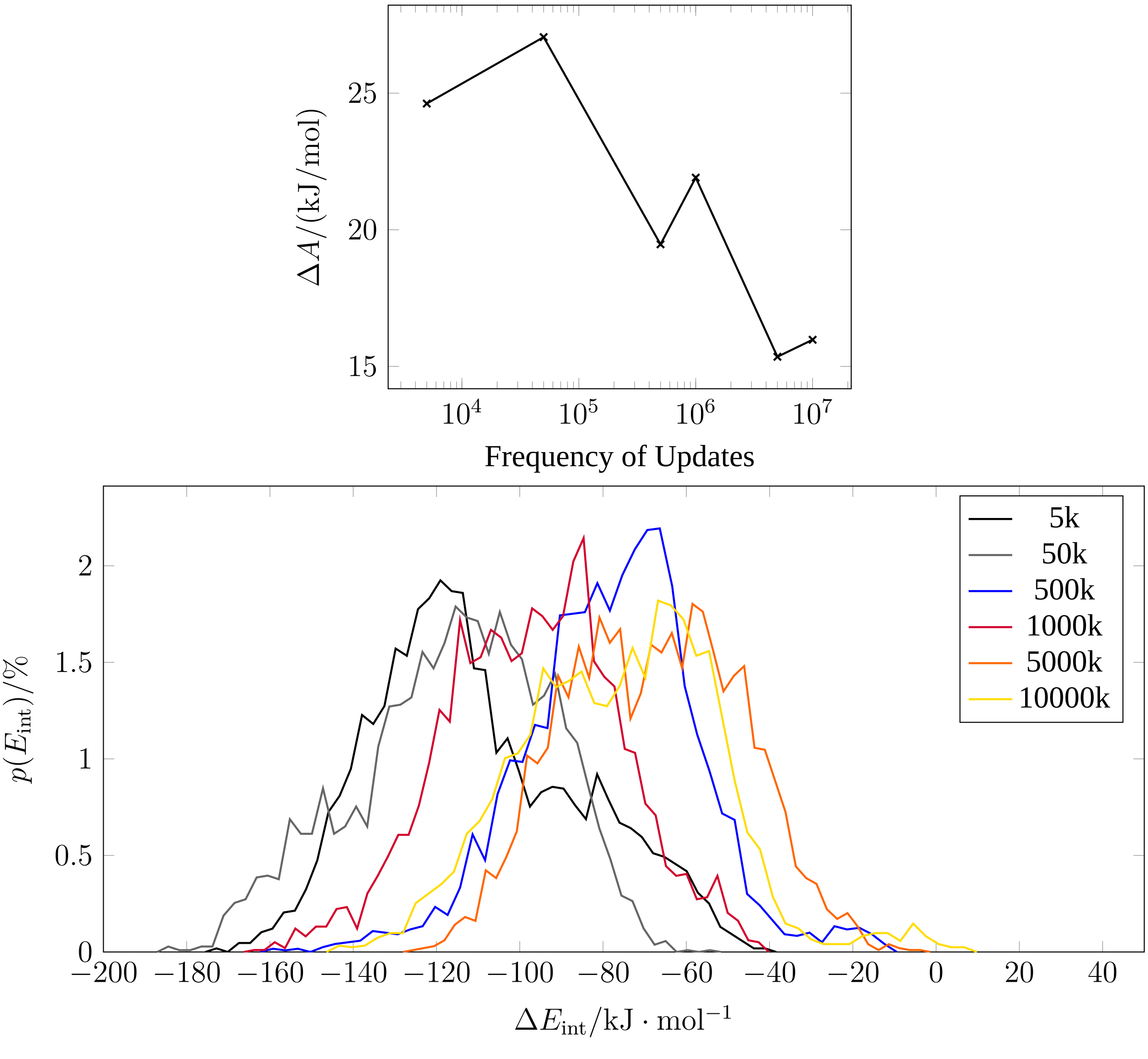


FIGURE 4.5 Free solvation energy simulations with PMC (B3LYP-D3/def2-TZVP and OPLS-AA vdW parameters) with increasing frequency of density updates for one QM water in TIP3P solvent. Top: First λ step. Bottom: Distributions of interaction energies between QM and environment.

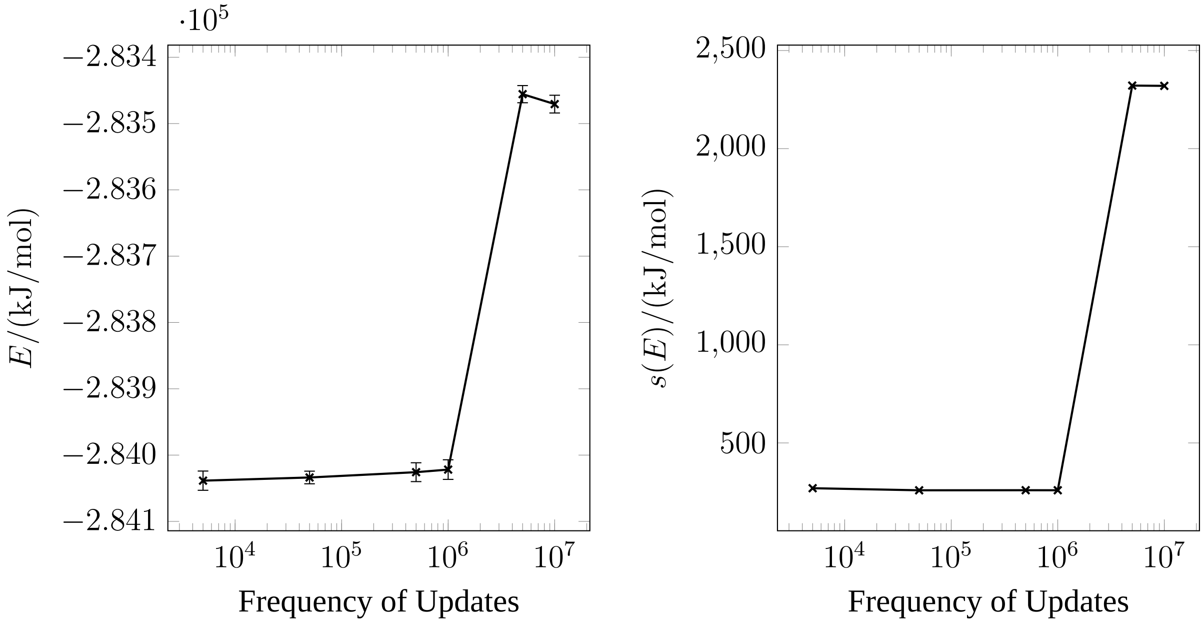


FIGURE 4.6 The potential energy of a simulation with PMC (B3LYP-D3/def2-TZVP and OPLS-AA vdW parameters) with increasing frequency of density updates for one QM water in TIP3P solvent. Left: ensemble average and standard error, right: standard deviation.

configuration space that is sampled because of the larger error in the perturbation approach. Nevertheless, the difference in the potential energy is still less well below 1%.

4.3 Free Energies of Solvation

Free energies of solvation are experimentally and theoretically well studied which makes them an important target of any new method attempting to describe solvent effects. [98] Many biological processes are governed by free solvation energies especially in the context of proteins or membranes. Any two molecules interacting in solution need to be at least partially desolvated, e.g. a ligand binding to a protein or a catalyst. The transport of drug molecules is of great importance for pharmacological applications. [115]

The theoretical basis for the computation of free solvation energies is straightforward but sufficient sampling as well as an accurate description of the interaction between solute and solvent are required. The free solvation energy is computed in the context of PMC simulations by a stepwise decoupling of the interactions between solute and solvent and a subsequent analysis with FEP or BAR. First, the electrostatic interaction is turned off linearly in four steps. In case large differences between the forward and reverse simulation have been encountered, which are indicative of a poor configuration space overlap the number of steps has been doubled. In the second phase the vdW interactions are decoupled which involves a soft-core potential in order to avoid numerical problems in the analysis of the free energies at the end points ($\lambda = 0$ and $\lambda = 1$) as discussed in Section 3.3. The λ protocol for this part has been chosen according to the study by Sherman and coworkers [115] in order to improve the accuracy and reduce the variance of the results. The solute geometries have been optimised with B3LYP-D3/def2-TZVP [17,20–22,111–114] and COSMO [41] correction ($\epsilon(\mathrm{toluene}) = 2.379$,

$\epsilon(\text{chloroform}) = 4.806$, $\epsilon(\text{acetonitrile}) = 35.88$ and $\epsilon(\text{water}) = 80.0$). Thermodynamic corrections have been included based on the rigid rotor-harmonic oscillator approximation in gas phase and COSMO. Geometry relaxation effects have been considered as well at the level of the continuum model. The simulations have been carried out at 298 K with B3LYP-D3/def2-TZVP and OPLS-AA vdW parameters for the solute. The number of molecules n and the box length l for the solvents are: acetonitrile ($n = 395$, $l = 41.0$ Å), water ($n = 2132$, $l = 40.0$ Å), chloroform ($n = 479$, $l = 39.98$ Å) and toluene ($n = 263$, $l = 36.0$ Å).

Reactions in solution and many other processes involve solute-solvent interactions in the initial as well as in the final state. For this reason, the error depends only on the differential error of the potential between the two states. In this case, the final state is naturally without any interaction between solute and solvent which makes free solvation energies rather sensitive to inaccuracies in the potential because the result cannot benefit from any error cancellation.

Classical force fields lack not only the transferability but also the accuracy in many cases. The PMC approach is expected to improve the description of the electrostatic interaction. However, a balance between electrostatic and van der Waals interactions is needed for the free solvation energies. Standard force fields such as OPLS-AA [29–31] have never been parametrised with QM/MM applications in mind. Hence, the replacement of only the electrostatic term of the force field might lead to an unbalanced description and no improvement in the results. In fact, recent studies [116] argued that a systematic improvement of free energies with QM/MM is not generally possible.

Here, in total twenty-nine energies for four different solvents and solutes with varying properties have been computed. PMC simulations of 80 M steps with updates every 20 k steps have been carried out and classical simulations with the same number of steps for the vdW part. The results are summarised in comparison to experimental results [117] in Figure 4.7 and in Table 4.1 which lists the exact combination of solute and solvent. The experimental results are Gibbs free energies while the simulations have been carried out in the canonical ensemble so that the computed results are Helmholtz free energies. It can be safely assumed that this effect is negligible for the small solutes of this benchmark set and very large simulation boxes of about 40 Å. Exploratory studies with ethanol in water, increasing the box edge length from about 24 Å, showed that the change in the free solvation energies is smaller than the statistical accuracy of the simulations.

Focusing on the results with water as solvent, first of all it can be noted that even small solutes with a correspondingly small QM region like water and ammonia in water can be described surprisingly accurate by QM/MM calculations. These systems are both strongly influenced by specific interactions, here most notably hydrogen bonds. Other hydrogen bonded systems like the series of alcohols show larger errors in comparison with the experimental results. However, the qualitative description and most importantly the energetic ordering is very well preserved. Examples are 1-propanol and 2-propanol which show the same free solvation energy or the ordering of ethane and ethene which is characterised by an energetic difference of only 2.5 kJ/mol experimentally which is still reproduced by the simulations. Last, the standard error of the free solvation energy in water is about an order of magnitude larger than with other solvents. This is due to the rigid structure of water including rather strong hydrogen bonds and consequently slow dynamics for example for solvent exchanges in the first solvation shell. This can be also seen in the considerably higher correlation between consecutive configurations and the associated high statistical inefficiency. To reach a similar statistical certainty for all systems 10–100 times more steps would be required for water but for the purpose of the benchmark the results were deemed sufficient.

An interesting outlier is the alcohol ethane-1,2-diol which is the only example with two hydroxyl groups. The experimental free solvation energy of -38.9 kJ/mol is overestimated by a factor of nearly two with -64.1 kJ/mol. A combination of different factors may lead to this large difference. The geometries of the solutes are optimised in continuum solvation models. Here, the two hydroxyl groups might lead to a number of hydrogen bonds involving one or more water molecules that are connecting these groups in a bridging manner. This very specific structure of the first solvation shell would distort the equilibrium geometry of the solute and could not be described even qualitatively by the continuum model. Additionally, this could also involve hydrogen bonds with more unusual geometric parameters. In this context dynamic effects of the solute in concert with environment as well as polarisation effects can play an important role. First, dynamic effects are only considered by the rigid rotor-harmonic oscillator model because a rigid solute geometry is used. This is a severe approximation especially in the case of rotations around the dihedral angle which orients the hydroxyl groups with respect to each other. Second, the water molecules in bridging positions might be differently polarised which cannot be reproduced by the effective potential with fixed charges. If hydrogen bonds occurring between water molecules and the solute and among water molecules of the first solvation shell are very different from average hydrogen bonds in bulk solvent then large errors can be expected.

Aromatic systems which might include specific interactions involving the π-system are described rather accurately for all solvents with errors up to 5 kJ/mol with the exception of toluene in water. Also, when the solvent molecules include an aromatic system as in the case of toluene, only described by a force field, the results agree well with the experimental findings.

The comparison between the QM/MM results and equivalent results obtained by purely classical simulations where the solute has been described as well with the OPLS-AA force field reveal to which degree the results can be improved with the hybrid approach. The results are directly be compared in Figure 4.8. The free solvation energies with water and toluene as solvent are clearly improved for 10 of 12 and for 3 of 4 combinations respectively. In these cases clearly the insufficient description of the electron distribution by the effective potential of the force field cannot accommodate for the polarisation of the solute and consequently describes the electrostatic interaction between solute and solvent incorrectly. Next, only two solutes have been simulated in toluene with rather good accuracy but it can be seen that the results are nearly identical to the classical simulations. Nevertheless, this shows the robustness of the QM/MM approach.

Lastly, we discuss the chloroform solution results. The latter show a large systematic underestimation of about 10 kJ/mol in the MM as well as with QM/MM. The hybrid approach does

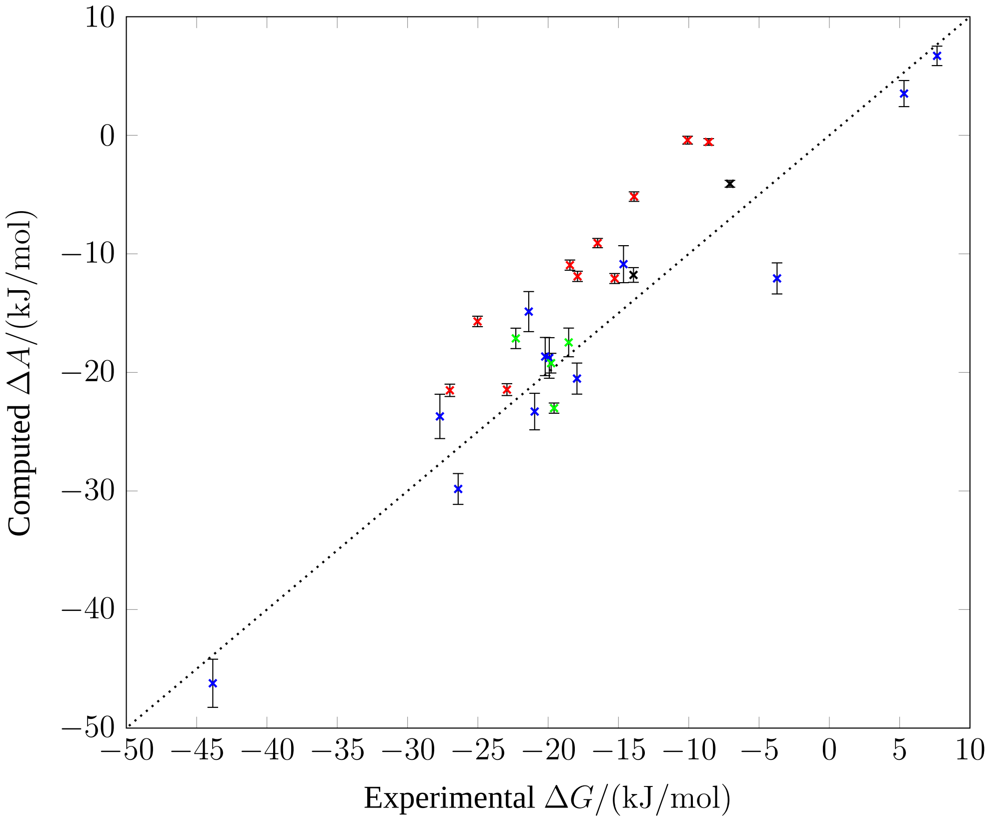


FIGURE 4.7 Solvation free energies computed with PMC (B3LYP-D3/def2-TZVP and OPLS-AA, OPLS-UA for chloroform) in comparison with experimental results. The solutes and solvents are listed in table 4.1. The colour coding is water→blue, toluene→black, acetonitrile→green and chloroform→red.

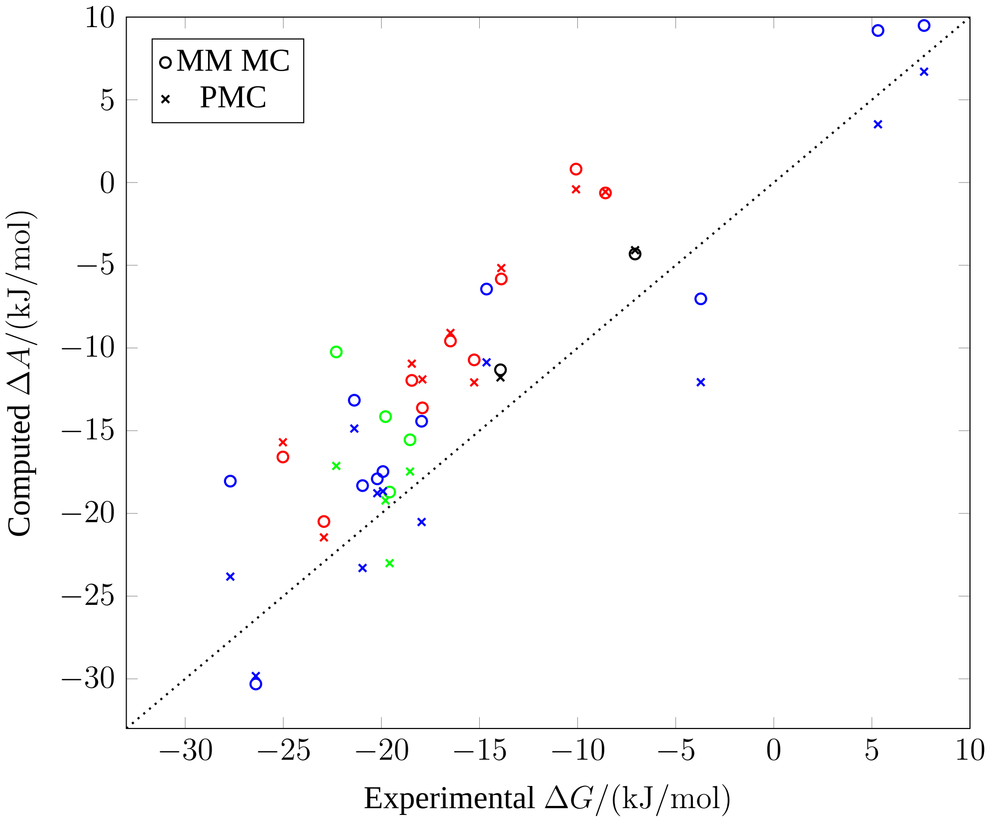


FIGURE 4.8 Solvation free energies computed with PMC (B3LYP-D3/def2-TZVP and OPLS-AA, OPLS-UA for chloroform) and MM MC (OPLS-AA, OPLS-UA for chloroform) in comparison with experimental results. The solutes and solvents are listed in table 4.1.

TABLE 4.1 All combinations of solute and solvent that have been used to compute free energies of solvation with the standard error σ and experimental results (Exp.). [117] All energies are given in kJ/mol. Computations (Comp.) have been carried out with PMC (B3LYP-D3/def2-TZVP and OPLS-AA, OPLS-UA for chloroform).

Solute	Solvent	Comp.	σ	Exp.
Water	Water	-29.8	1.3	-26.4
Ammonia	Water	-20.5	1.3	-18.0
Ethane	Water	6.7	0.8	7.7
Ethene	Water	3.5	1.1	5.3
Methanol	Water	-14.9	1.7	-21.4
Ethanol	Water	-23.3	1.5	-21.0
Ethane-1,2-diol	Water	-64.1	1.7	-38.9
1-Propanol	Water	-18.8	1.7	-19.9
2-Propanol	Water	-18.7	1.6	-20.2
Ethanal	Water	-10.9	1.6	-14.6
Toluene	Water	-12.1	1.3	-3.7
Phenol	Water	-23.8	1.9	-27.7
4-Hydroxybenzaldehyde	Water	-46.2	2.0	-43.9
Ethanol	Toluene	-11.8	0.6	-13.9
Water	Toluene	-4.1	0.3	-7.1
Ethanol	Acetonitrile	-17.5	1.2	-18.5
Butanone	Acetonitrile	-19.2	0.8	-19.8
1,4-Dioxane	Acetonitrile	-17.1	0.9	-22.3
Toluene	Acetonitrile	-23.0	0.4	-19.6
Water	Chloroform	-0.6	0.3	-8.6
Ammonia	Chloroform	-0.4	0.3	-10.1
Methanol	Chloroform	-5.2	0.4	-13.9
Ethanol	Chloroform	-9.1	0.4	-16.5
Ethane-1,2-diol	Chloroform	-15.7	0.4	-25.0
1-Propanol	Chloroform	-11.0	0.4	-18.5
2-Propanol	Chloroform	-11.9	0.4	-17.9
Ethanal	Chloroform	-12.1	0.4	-15.3
Toluene	Chloroform	-21.5	0.5	-22.9
Pyridine	Chloroform	-21.5	0.5	-27.0

not change the results considerably and does also not consistently improve upon the classical simulations. This suggests a problem in the potential of the solvent which cannot be improved with a better description of the solute. Additional simulations with a set of modified partial charges for chloroform confirm these findings with the results shown in Figure 4.9. The dipole moment has been computed with Coupled Cluster with Singles and Doubles (CCSD) and a cc-pVTZ basis set including the COSMO ($\epsilon = 4.806$) correction at the HF level (PTE scheme) which results in a dipole moment which is 1.3 times larger with the continuum model than in the gas phase. Subsequently, the modified chloroform potential has been constructed by scaling the original partial charges by the same factor of 1.3. Classical simulations of pure chloroform have been carried out in order to confirm that the solvent structure is not fundamentally disturbed by these modifications. The corresponding radial distribution functions are shown in Figure 4.10. Only the CH-CH radial distribution function shows very minor changes of the height of the first peak while the position is perfectly retained. Conversely, the free solvation energies computed with this modified potential show a systematic shift towards the experimental values which reduces the error by about 2.5 kJ/mol as obtained by a linear fit. This shows clearly that the description of the electron distribution of the solute as well as of the solvent molecules is important. The partial charges are fitted in order to obtain a balanced description and internal consistency for the force field but it can be seen that this does not necessarily make the potential suitable in the context of QM/MM simulations as opposed to the water model that works very well without further adjustments. Even though the importance of the dipole moment has been shown, further aspects might be relevant. If the dipole moment is estimated in order to reproduce the experimental results this would lead to a huge overestimation with about 5 D compared to the experimental value [118] in liquid phase of 1.25 D. The example of chloroform shows that attention has to be paid to the classical as well as the QM side in order to achieve high accuracy in the coupling term.

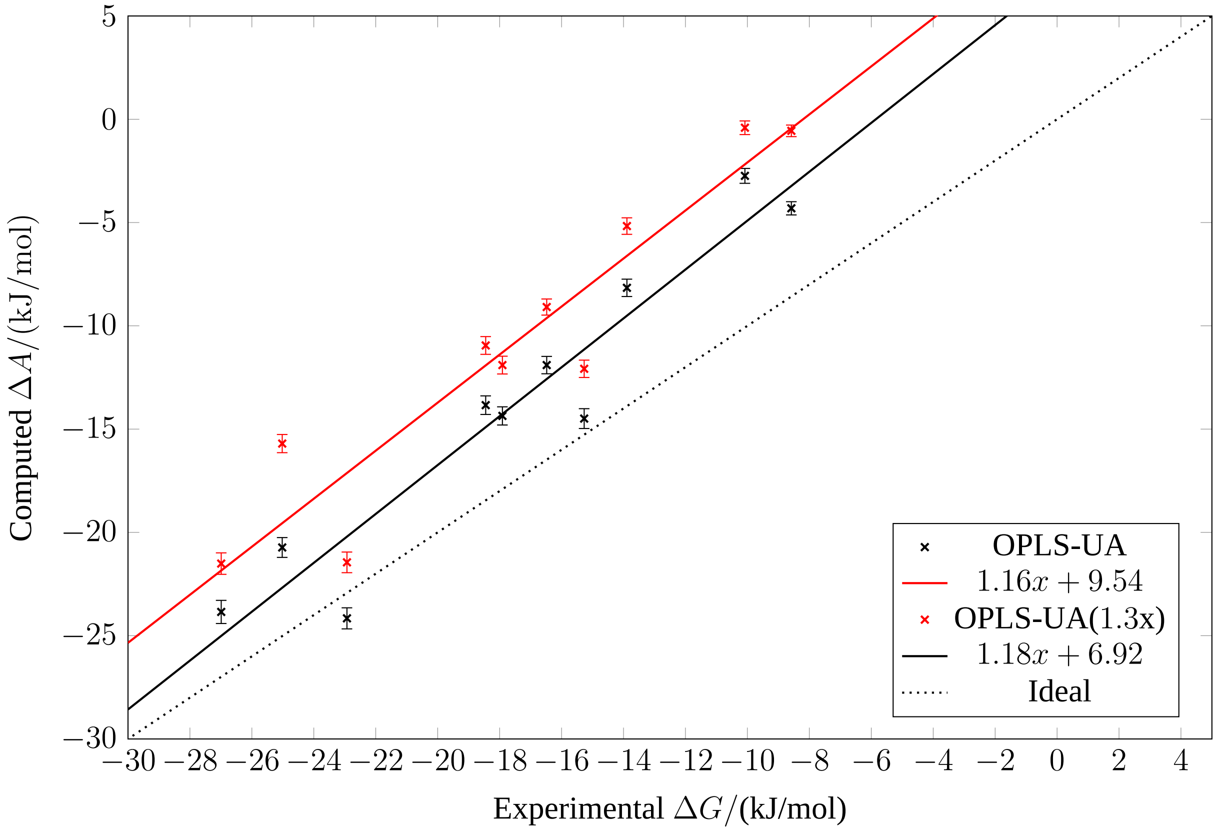


FIGURE 4.9 Solvation free energies computed with PMC (B3LYP-D3/def2-TZVP and OPLS-AA vdW parameters) in chloroform with standard parameters (OPLS-UA) and chloroform with an increased dipole moment (OPLS-UA(1.3x)).

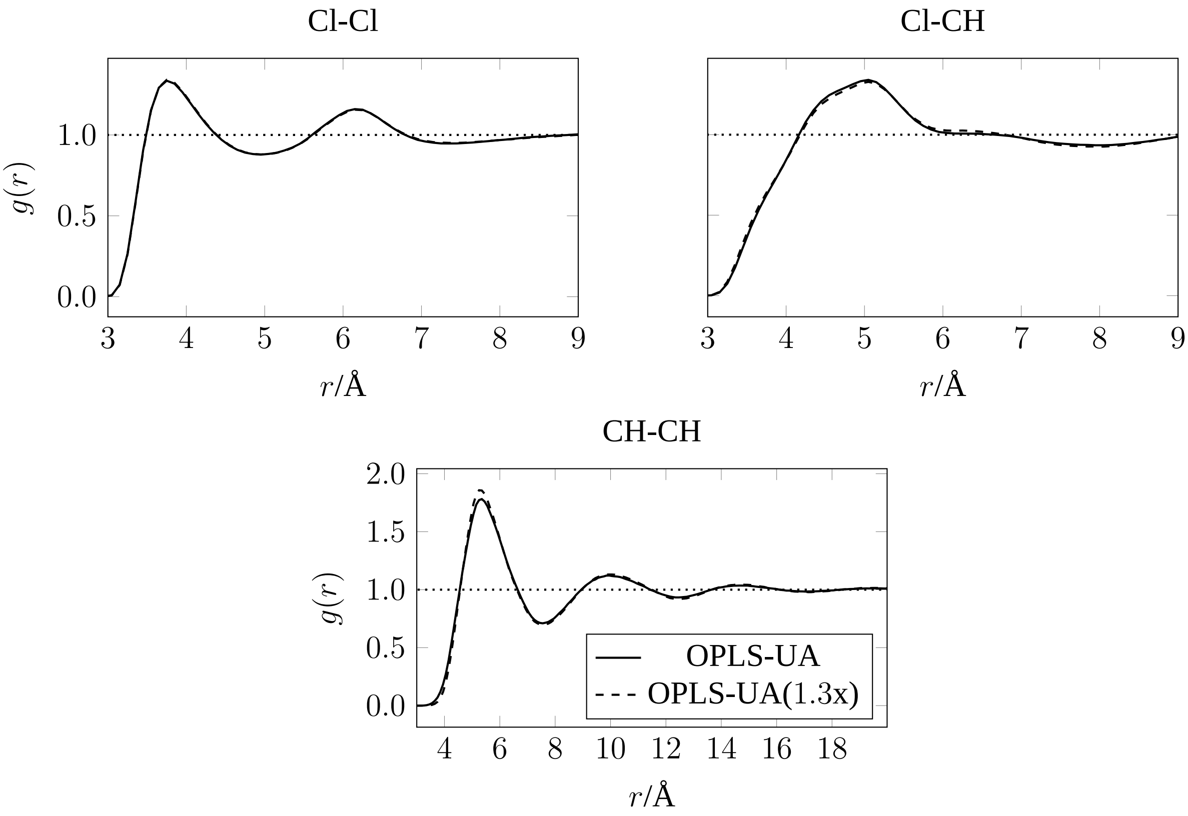


FIGURE 4.10 MC Radial distribution functions for pure chloroform with OPLS-UA and an increased dipole moment (OPLS-UA(1.3x)). Top left: Cl-Cl, top right: Cl-CH, bottom: CH-CH.

4.4 Free Energy of Torsional Potentials

Reaction pathways are fundamentally affected by the surrounding media through non-covalent interactions and polarisation effects. It might also come to be that solvent molecules participate in reactions. Thus whole new pathways might come accessible and not only the stability of stationary points relative to each other is effected. In this Section I will investigate the possibility to describe reaction pathways in the context of PMC simulations combined with FEP and the BAR analysis. As a model, hydrogen peroxide, a seemingly simple system will be investigated and the free energy for the internal rotation of the same will be determined.

Hydrogen peroxide is subject to intense study in several different fields. In atmospheric chemistry it is relevant as the dominating oxidant for SO_2 in clouds, [119] in industrial processes it is used for advanced and environmentally friendly water treatment due to its antibacterial and oxidising properties [120] and in numerous biochemical processes it plays an important role. For example, in the context of membrane transports and the modulation of transcription factors, [121–123] as a signalling agent of pathways that regulate the reactive oxygen species concentrations of the intracellular volume [124] and as a source of oxidative stress. [125] As a by-product of metabolic reactions like the conversion of hypoxanthine to xanthine it is generated directly inside of the cells. [126] Especially hydrogen bonded complexes of hydrogen peroxide are relevant in metabolic environment and for the modulation of biological processes. [126,127] Finally, it is the simplest molecule that exhibits helical chirality with the conversion from P to M being possible due to the hindered internal rotation. Therefore, the barrier for this transition has been the target of experimental as well as theoretical studies. [128–131] The solvent structure will be investigated under a different context in Section 5.1.

PMC simulations have been carried out with a temperature of 298 K varying the dihedral angle in 2.5° steps with 20 M steps for every value and 2 M intermediate steps in order to reach equilibrium after changing the geometry of the solute. Density updates have been carried out every 1000 steps throughout. The B3LYP-D3/def2-TZVP [17, 20–22, 111–114] method has been used for the QM calculations and the TIP3P water model [132] as well as vdW parameters for hydrogen peroxide from the work of Margulis and coworkers. [133] The box with the edge length of 40.0 Å contained 2132 water molecules. Additionally, the potential has been computed in gas phase and with COSMO [41] ($\epsilon = 80$) corrections at the same level of theory. These results are shown in Figure 4.11.

Independent of the environment the barrier of the cisoid transition state at 0° is considerably higher than the transoid barrier at 180°. However, the former is decreased from about 33 kJ/mol in gas phase to about 22 kJ/mol with PMC and 19 kJ/mol with COSMO. In contrast, the transoid barrier is only increased with the continuum description while PMC decreases this barrier as well. This shows clearly that an explicite solvent structure is important for the description of specific interactions like hydrogen bonds which are here accepting and donating with respect to the solute. The effect on the transoid transition state is qualitatively wrong with the continuum model. Our findings are in agreement with the QM/MM replica exchange MD simulations by Choi and coworkers. [134] However, in their study only the HF/3-21G level of theory has been

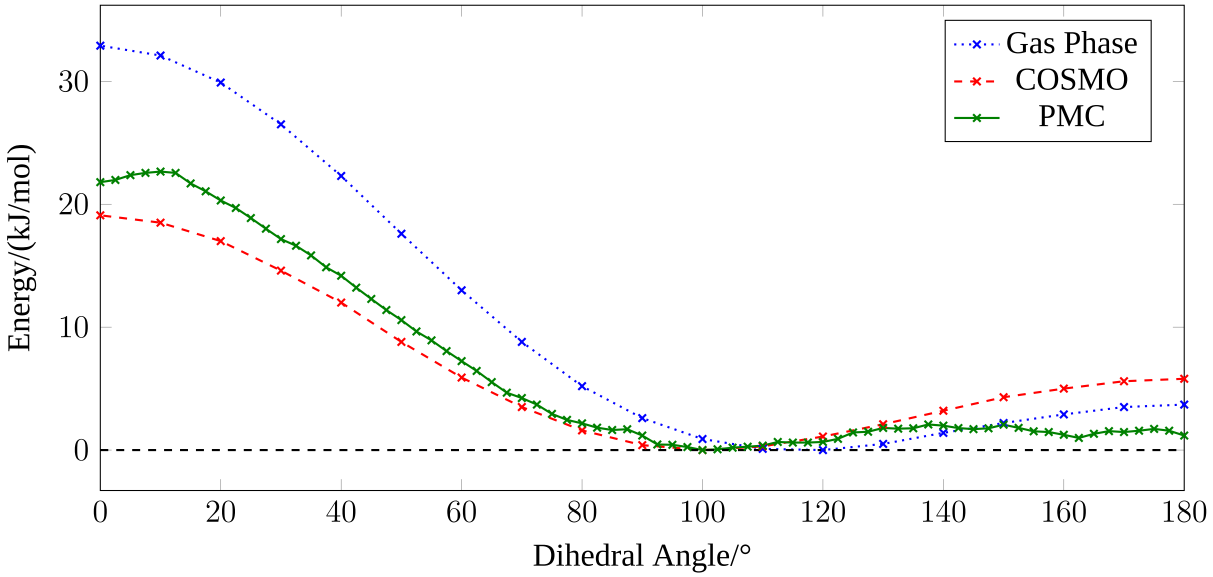


FIGURE 4.11 The torsional potential of hydrogen peroxide computed in gas phase, with COSMO ($\epsilon = 80$) and in TIP3P water with PMC (B3LYP-D3/def2-TZVP).

used for the QM system. Furthermore, they used the TIP5P water model which we found to be unsuitable for QM/MM applications (see Chapter 5). Their results have to be considered with caution and the agreement might be coincidental. They found that the cisoid barrier is lowered by about 15 kJ/mol while the transoid transition state is stabilised by about 2 kJ/mol. The position of the minimum had been located at about $90°$ which is in fair agreement with $100°$ from both COSMO and PMC while the gas phase minimum is at about $120°$. This shows another important effect of solute-solvent interactions. The latter not only influence the relative stability of the stationary points but also their geometry.

CHAPTER 5

Impact of QM/MM on Solvent Structures

The arrangement of the solvent molecules around a solute and the properties of the solute itself depend on an accurate description of the energetics of such a system. The agreement with experimental findings of the solvent structure is a good measure of any method that attempts to simulate processes in solution.

The solvent can influence the solute in different ways. The direct effect changes the energies of stationary points with respect to each other, e.g. a reaction barrier is lowered due to a favourable interaction at the transition state or an unfavourable one at the minima. An example is the Cope elimination which experiences a million fold increase in the reaction rates going from a protic to an aprotic solvent. Hydrogen bonding can stabilise the reactant and thereby increases the effective barrier and slows down the reaction. [135] This emphasizes again the importance of hydrogen bonds and directional interactions which are not well described with continuum and classical approaches.

Furthermore, the geometric effect influences the actual position or even existence of stationary points. At last, the curvature of the potential energy surface of internal degrees of freedom of the solute is affected. This is called the vibrational effect because the vibrational levels and consequently the free energy surface is changed. In particular, a correct description of the first solvent shells is fundamental and requires a balanced description of the energetics between solute and solvent.

The solvent structure is typically analysed in terms of radial distribution functions. These represent the short range order of a liquid compared to the ideal gas for a pair of elements. Values larger than 1 represent an increased number of atoms and values smaller than 1 show regions which are depleted. For large distances the distribution converges to 1 for liquids because they do not exhibit any long range order. The distribution can be analysed for pairs of elements, e.g. between oxygen atoms in water, which reveals the shell-like arrangement of water molecules around another water molecule. Typically, 2–3 solvent shells can be distinguished in such distributions.

Radial distributions functions are experimentally accessible via X-ray or neutron scattering data. However, the analysis of the data and the derivation of radial distribution functions via

---

The material in this Chapter was presented in part in Reference [109].

the structure factor is quite involved. One of the difficulties is that intramolecular contributions are strongly dominating the experimental data. Furthermore, hydrogen atoms are very difficult to detect with X-ray diffraction. Many experimental results are only available for the heavier atoms.

On the other hand, molecular computer simulations can be trivially analysed in terms of radial distribution functions. The analysis amounts to merely counting the number of pairs as a function of the distance for a trajectory. The angle can be additionally measured which results in the angular radial distribution functions. Especially for hydrogen bonds the combination of bond length and angle are important criteria to analyse the strength and existence of hydrogen bonds. A typical geometric criterion is a bond length up to 2.5 Å and an angle of up to $30°$ as it is defined for example in Figure 5.3. [136] Furthermore, the distributions can be collected along a reaction pathway and differential solvent effects on stationary points can be assessed.

5.1 Hydrogen Peroxide

Here, hydrogen peroxide is revisited under the aspect of the solvent structure and hydrogen bonding. The free energy potential of the dihedral angle has been studied in Section 4.4 where the biological relevance of this system has been described as well. PMC simulations at the minimum of the potential have been carried out for 640 M steps at a temperature of 298 K with the same computational settings as before. In short, the B3LYP-D3/def2-TZVP [17,20–22,111–114] level of theory combined with different water models [137] and the vdW parameters from the study of Margulis and coworkers. [133]

The solvent structure as well as the hydrogen bond interactions can be studied with the radial distribution functions which are shown in Figure 5.1 for the TIP3P water model. The maxima of the peaks denoting the position of the solvent shells have been compared in more detail with a wide range of theoretical studies in Table 5.1. A prominent feature of the distributions is the difference between the bands corresponding to donating and accepting hydrogen bonds. The former is at about 0.22 Å shorter distances than the latter which is in agreement with the generally known higher hydrogen bond donor capability of hydrogen peroxide than its capability to accept hydrogen bonds. A second peak can be seen in the H···OW distribution while two more peaks are present in O···HW. The oxygen-oxygen distribution (O-OW) matches closely the distinguishable accepting and donating hydrogen bonds with two close peaks. However, the distance between them is two times larger with 0.45 Å than the corresponding distance in the H···OW.

It comes as no surprise that the results match very closely the QM/MM MD study by Martins-Costa and Ruiz-López [125] which uses nearly the same combination of methods even though a fixed solute geometry has been used in the PMC simulations. Born-Oppenheimer MD simulations [138] show a slightly longer hydrogen bond length but the results agree very well. The elongated hydrogen bonds can be explained by the better QM description of the solvent molecules which allows capturing dynamic and polarisation effects of the solvent and even different protonation states as the potential is inherently reactive. However, these seem to play

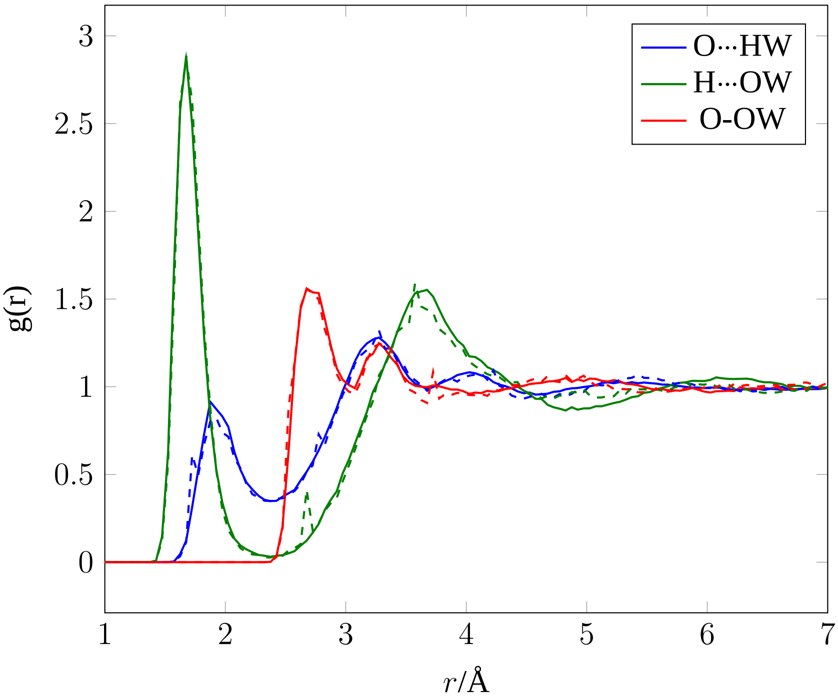


FIGURE 5.1 PMC Radial distribution function $g(r)$ of donating (O···HW) and accepting (H···OW) hydrogen bonds of hydrogen peroxide and the oxygen-oxygen function (O-OW). The full lines have been simulated with B3LYP and the dashed ones with PBE.

a minor role or in the case of the polarisation it is already captured sufficiently well by the effective potential of the water model.

However, the comparison with further simulations using only force fields show considerable deviations especially for the first solvation shell. The fixed point charge model employed by Coutinho and coworkers [127] combined with TIPS, which is the original parametrisation of TIP3P, shows a shift of all bands up to 0.2 Å and misses the third solvent peak of the O···HW distribution altogether. Since the water model is rather similar this can be attributed to the less accurate description of the electron density of the solute. The QM derived polarisable force field atom-bond electronegativity equalization method (ABEEM) used by Yang and coworkers [139] improves upon this especially for the more distant solvent shells. However, the error in the peaks corresponding to the hydrogen bonds have still an error of more than 0.1 Å. This is especially surprising because terms are included which describe the donating and accepting hydrogen bonds. It is not expected that the reference geometries and energies are the root of the problem because the MP2/AVTZ//MP2/AVDZ level of theory including corrections for the basis set superposition error is widely used in studies of water clusters and only minor differences are found with respect to the standard CCSD(T). [141–144] However, only small clusters of hydrogen peroxide and 1–6 water molecules have been used which includes in the best case only the first solvation shell and might be too restrictive in this case. Finally, the setup of the system is different with about 30 % mass fraction of hydrogen peroxide instead of a single molecule which might account for the differences in the long-range region of the O···HW distribution.

The comparison with the best available data in the literature shows that a quantitatively correct description of the solvent structure is obtained with the PMC approach. Dynamic and polarisation effects of the solvent beyond the effective potential are negligible at least for this

TABLE 5.1 Comparison of the peak positions of the RDFs shown in Figure 5.1 with results from Cabral [138]: BOMD (B3LYP-D3), Ruiz-Lopéz [125]: QM/MM MD (B3LYP/6-31G*, TIP3P), Coutinho [127]: MM MC (TIPS), Yang [139]: MM MD (ABEMM), Rode [140]: QMCF MD (HF/DZP, BJH-CF2) and Choi [134]: QM/MM-MD (MP2/6-31G*, TIP5P). Not reported distributions are left empty while peaks that are not present are denoted by ”–”.

Ref.	Method		O···HW			H···OW		O-OW	
	QM	MM	1st	2nd	3rd	1st	2nd	1st	2nd
[138]	B3LYP-D3		1.91	3.24	4.18	1.74	3.65		
[125]	B3LYP	TIP3P	1.85	3.26	4.02	1.64	3.61		
[127]		TIPS	2.08	3.42	–	1.81	3.81		
[139]		ABEMM	2.02	3.25	6.00	1.75	3.68		
[140]	HF	BJH-CF2	2.10	3.37	–	1.77	3.56	3.00	
[134]	MP2	TIP5P						2.88	
	B3LYP-D3	TIP3P	1.89	3.28	4.03	1.67	3.65	2.72	3.29
	B3LYP-D3	TIP4P	1.90	3.23	4.02	1.72	3.66	2.74	3.32
	PBE-D3	TIP3P	1.91	3.28	4.02	1.68	3.58	2.69	3.30

specific system. Furthermore, water is an ubiquitous solvent and good transferability of these findings is expected towards other solutes. This is supported by the analysis of the free solvation energies computed in Section 4.3. The interaction between solute and solvent appears to be accurately described by QM/MM approaches. The large differences between the different theoretical studies cannot be clearly attributed to inaccuracies in the solute or the solvent potential. Therefore, the influence of different different density functionals as well as solvent models will be investigated in the following.

It is expected that the PMC simulations depend only very little on the functional used for the DFT calculations of the solute. While the energies might differ significantly, the densities generated by different functionals are expected to be very similar. Therefore, simulations have been carried out with the PBE [110] functional. The comparison with results from the hybrid functional B3LYP in Figure 5.1 shows that they are nearly identical and the small deviations are in the range of the statistical error. The HF method used in the study by Rode might show larger errors but little dependence on the functional or basis set is expected.

Next, different water models have been used to investigate the influence of the MM model in hybrid QM/MM simulations. The results are shown in Figure 5.2 and the peak positions are listed in Table 5.1 except for TIP5P where no meaningful peaks can be assigned. While the results from TIP3P and TIP4P are very similar, the TIP5P distribution functions are fundamentally distorted and appear to be in parts poorly converged. In a purely classical simulation the results can be clearly improved going from TIP3P over TIP4P to the TIP5P water model, increasing the number of partial charges. However, it is known from simulations of liquid water with QM/MM methods, that the TIP5P water model can cause large errors and can disturb the solvent structure completely. [65] It appears that the additional negative charge sites in TIP5P which describe the lone pairs (LP) of the oxygen get embedded in the electron density of the QM molecule. The main reason is that no vdW parameters are associated with these sites. This leads to a strong polarization of the density and very short distances for hydrogen bonds as it can be seen in the additional RDF for H-LP which places the lone pair at about 0.65 Å. The consequences are very short distances for the peaks of the first solvation shell of all RDFs involving an oxygen of the water (H···OW, O-OW). Another problem can appear especially in MC simulations. Due to the short distance between lone pair and proton of 0.65 Å, the lone pair can reach with a single move the nucleus of the proton and becomes immediately trapped there due to the highly attractive Coulomb interaction which even further distorts the solvent structure around the solute and also leads to insufficient sampling.

Based on these results the strong influence of the employed force field — in this case the water model — becomes apparent which explains the large differences between the available literature values. The results reported by Choi [134] have to be considered with care. However, only the O-OW distribution has been reported which allows no further conclusions. On the other hand, the results for TIP3P and TIP4P are very similar. The additional site in the TIP4P is located close to the oxygen in the middle of the water molecule and cannot be embedded in the electron density. It has been shown that for TIP4P water the RDF is least changed upon switching a single MM water to a QM description and that TIP4P exhibits therefore the best compatibility within a QM/MM approach. [65]

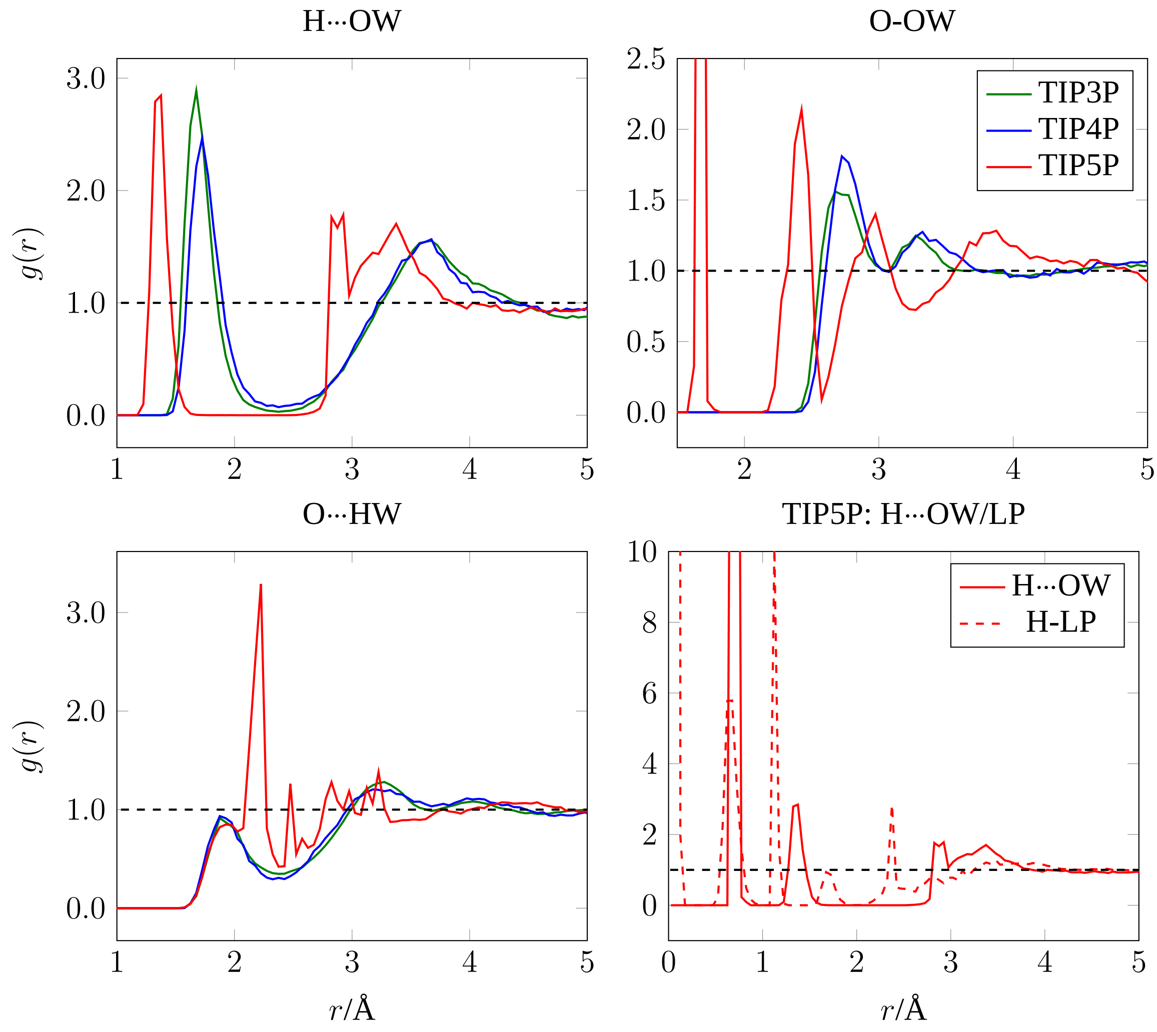


FIGURE 5.2 PMC (B3LYP-D3) Radial distribution functions for pairs X-Y with the QM atoms X and the MM site Y. HW and OW denote the oxygen and hydrogen atom of the respective water model and LP the site describing the lone pairs of a water molecule (only for TIP5P).

5.2 Methylchloride

Nucleophilic substitution reactions of methyl alkanes

$$\mathrm{X}^- + \mathrm{CR}_3\mathrm{Y} \rightarrow \mathrm{XCR}_3 + \mathrm{Y}^-$$

are well known for being strongly influenced by solvent effects. The double-well potential in gas phase changes to an unimodal reaction in aqueous solution. Experimentally measured transfer rates are reduced by 20 orders of magnitude in solution [145] which demonstrates the drastic effect the solvent has on the reaction. It is argued that the barrier is influenced due to the required desolvation of the nucleophile and a greater charge dispersal in the transition state. [146] Computational studies with a continuum model confirmed indeed that a large electron density redistribution occurs near the transition state. [145] A selection of the computational studies that have been carried out for $\mathrm{X} = \mathrm{Y} = \mathrm{Cl}$ include classical simulations, [146] continuum solvation models [145, 147] and Car-Parrinello Molecular Dynamics (CPMD) simulations. [148] Additionally, the reaction with $\mathrm{X} = \mathrm{Br}$ and $\mathrm{Y} = \mathrm{Cl}$ has been investigated with CPMD simulations. [136] Here, I will carry out PMC simulations of the reaction with $\mathrm{X} = \mathrm{F}$ and $\mathrm{Y} = \mathrm{Cl}$. The solvent structure as a fundamental property will be studied at the vdW complex, i.e. when the fluoride coordinates to chloromethane but no bond breaking or forming takes place yet. Simulations with B3LYP-D3/def2-TZVP [17, 20–22, 111–114] and the TIP3P water model have been carried out at the temperature of 298 K for 640 M steps with density updates every 20 k steps. The RDFs have been sampled every 10 k steps. The box with 2132 water molecules had the edge length of 40.0 Å.

The radial distribution functions around the leaving chlorine have been collected in this educt state and compared to results from CPMD simulations for very similar systems with chloride [148] or bromide [136] instead of fluoride which are all shown in Figure 5.4. In those studies the simulations have been carried out with about 30 water molecules in a simulation box with the edge length of 10 Å using a proton as counterion. For the Cl system 3–5 ps have been simulated in the canonical ensemble ($T = 300$ K) with the Perdew-Zunger LSDA functional combined with the BP86 GGA functional. A frozen core approximation has been used in combination with an augmented plane wave basis for the valence electrons. It should be noted that

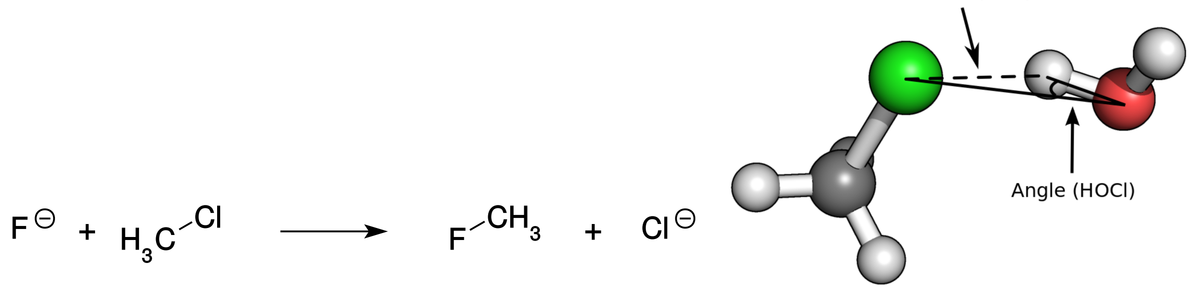


FIGURE 5.3 Left: Nucleophilic substitution involving chloride and fluoromethane. Right: Definitions of distance and angle for the analysis of the angular radial distribution functions.

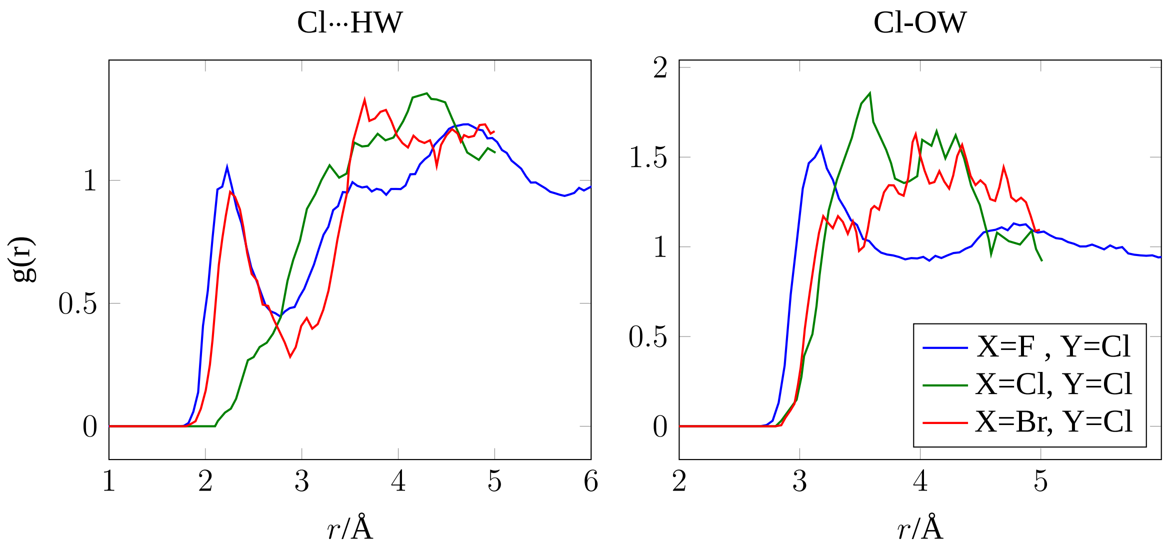


FIGURE 5.4 PMC Radial distribution functions of the leaving Cl in the educt state with B3LYP-D3/def2-TZVP and OPLS-AA parameters. Left: Cl to H of water, Right: Cl to O of water with the literature results X=Cl, Y=Cl [148] and X=Br, Y=Cl [136] (at the transition state).

these simulations have been carried out at the transition state of the S_N2 reaction. A pseudo potential approach has been used for the Br system with the HCTH GGA functional. Production runs have been carried out for 4–12 ps in the micro-canonical ensemble.

Comparing the PMC with the CPMD simulations it can be seen that the first peak in the Cl···HW distribution is located at about 2.25 Å for the F and Br system. However, for the Cl system this peak is missing altogether. Additionally, a rather broad second peak can be found roughly around 4 Å for all systems. In the Cl-OW distribution the picture is less clear. It is rather unstructured in the case of the Cl and Br system while in the F system a peak can be found at about 3.2 Å. In the other systems an accumulation of oxygen between 3 and 5 Å is noticeable but without a distinctive maximum. Most surprising is the missing hydrogen bonding region in the case of the Br system. In the same study the comparison with CH_3Cl has been drawn which defines the lower limit for the height of this peak. However, the educt state does not conform to this limit which suggests that there might be a severe approximation in the choice of the computational methods or the set-up of the simulations which strongly influences the hydrogen bonding. It can be seen that the short simulation time scales for both Cl and Br result in poorly converged distributions. Another consequence is the large asymmetry of more than 5 kcal/mol in the free energy of the S_N2 reaction which has been computed for the Cl system. However, this reaction is by definition thermoneutral. It has been suggested to include the solvent degrees of freedom into the reaction coordinate of the constrained sampling procedure in order to accelerate the slow solvent rearrangement upon change of the solutes geometry. Nevertheless, such a coordinate cannot be easily constructed.

Another limitation due to the high computational costs of the CPMD simulations is the small simulation box which leads to an artificially high concentration compared to experimental conditions. In contrast, the volume is an order of magnitude larger in the PMC simulations and the interaction between the solute and its image is by construction zero. This is also most

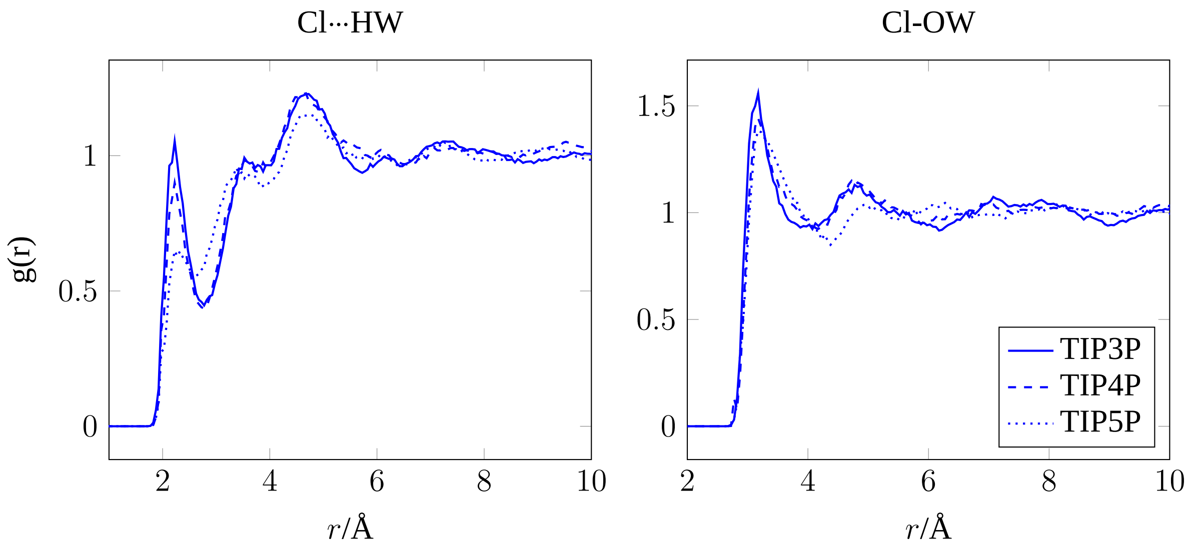


FIGURE 5.5 PMC Radial distribution functions of the leaving Cl in the educt state with the water models TIP3P, TIP4P and TIP5P and the QM approach B3LYP-D3/def2-TZVP.

likely responsible for the shift of the second peak in the Cl···HW distribution which in the PMC simulations has the maximum at about 4.7 Å and extends beyond 5 Å. However, with a system size of 10 Å in the CPMD simulations the peak cannot go beyond 5 Å and is artificially shifted to smaller distances in the range of 3.5–4.5 Å. Consequently, the slightly ordered structure in the range of 5–10 Å is not represented in the CPMD simulations as well as the unstructured part representing the bulk solvent for distances larger than 10 Å. It is expected that a considerable finite size effect is present in the CPMD simulations.

The experiments are carried out at neutral pH with potassium as counterion and in the CPMD simulations a proton has been used. The shifted force operator in PMC allows to simulate charged systems so that an investigation is possible without spurious effects of the counterion.

The influence of the employed force field has been reassessed. As opposed to hydrogen peroxide in the previous Section the solute is charged and is exclusively a hydrogen bond acceptor. The RDFs of simulations with TIP3P, TIP4P and TIP5P are shown in Figure 5.5. Again, the results of the TIP3P and TIP4P models are very similar and show only minor differences in the height of the peaks while their positions are not changed. Furthermore, the results with the TIP5P model are less disastrous because the water molecules only build donating hydrogen bonds towards the solute which means that the additional charge sites describing the lone pairs point away from the solute. However, the first feature in the Cl···HW distribution at about 2.3 Å is very unstructured and corresponds to very unstable hydrogen bonds. This does not agree with the more structured distribution as suggested by the CPMD results.

The angular radial distribution function has also been constructed and is in excellent agreement with the results from the Br system by Schettino and coworkers. [136] At this point specific attention is paid to the influence of the level of theory on the configuration space that is sampled. In Figure 5.6 these distributions are compared for PMC and classical MD simulations with the equivalent parameters. The radial distribution is qualitatively different, the PMC results show a first hydration shell at about 2.2 Å which is completely missing with the force field description.

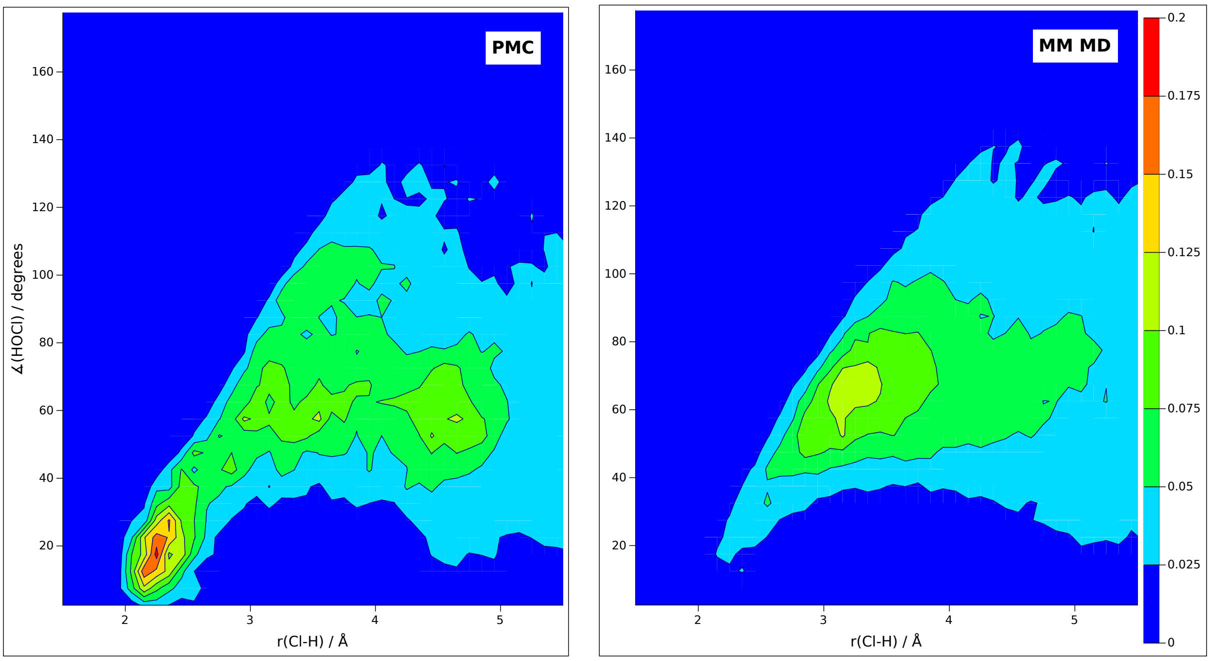


FIGURE 5.6 Angular radial distribution function for the educt of the S_N2 reaction with PMC (B3LYP-D3/def2-TZVP and OPLS-AA vdW parameters) using the TIP3P water model and MM (OPLS-AA and TIP3P).

Furthermore, the PMC results show H-O-Cl angles lower than $40°$ for short distances. These values are typical for a halide anion solvation shell determined by hydrogen bonds. While the radial distribution is normalised the angular one is commonly not normalised. Therefore, the expected peak centred around $0°$ is missing and seen at slightly larger angles.

The angular radial distribution functions with the water models TIP4P and TIP5P are shown in Figure 5.7. They confirm that TIP3P (Figure 5.6 left) and TIP4P match very closely also with respect to the angular distribution. Unexpectedly, they also reveal that the angular distribution is maintained even for the TIP5P water model. While the hydrogen bond strength is severely reduced with this model the angular distribution typical for hydrogen bonds is reproduced.

These results reveal a restriction of the widely used sequential QM/MM approach in which first classical simulations are carried out and subsequently the approximate energies are improved with a hybrid QM/MM approach. [149, 150] However, the configurations are only generated at the MM level of theory which can be problematic for example for the system that has been studied here. Even in the theoretical situation that the classical simulations generate more accurate configurations than the high level method the sequential approach can lead to a large variance and inaccurate averages. This is the case whenever the configuration spaces are different or in other words the position of the minima on the potential energy surface are not the same. Schematically, this situation is illustrated in Figure 5.8. However, these problems do not occur in the PMC simulations since QM/MM itself can be used for the sampling.

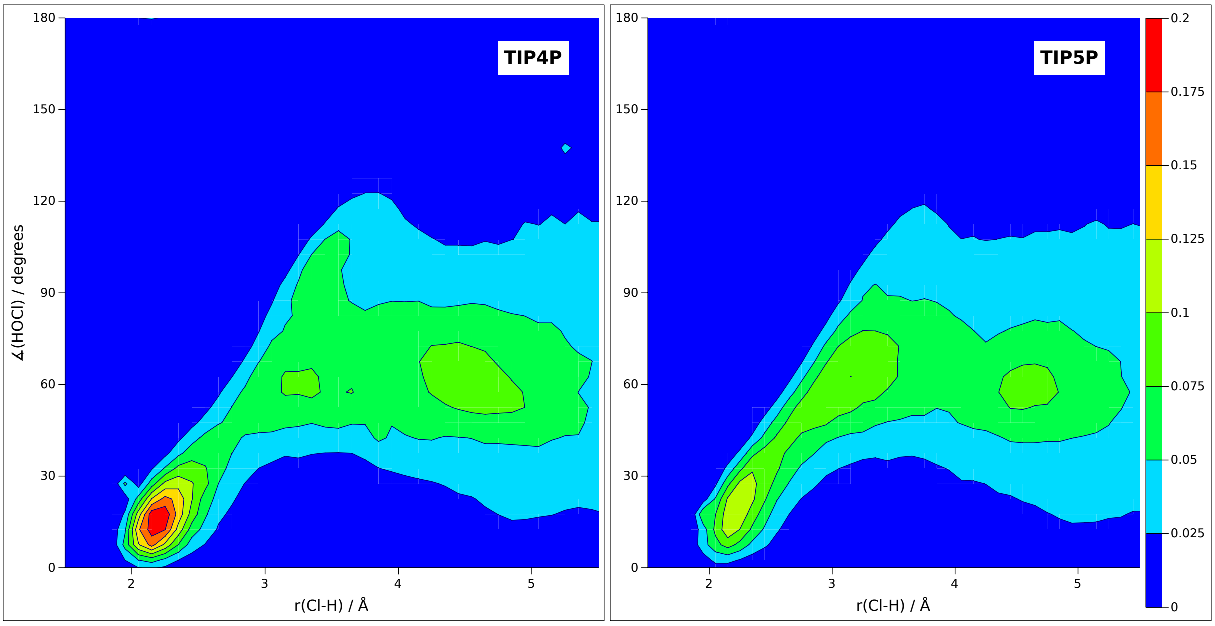


FIGURE 5.7 Angular radial distribution function for the educt of the S_N2 reaction with PMC (B3LYP-D3/def2-TZVP and OPLS-AA vdW parameters) and the water models TIP4P (left) and TIP5P (right).

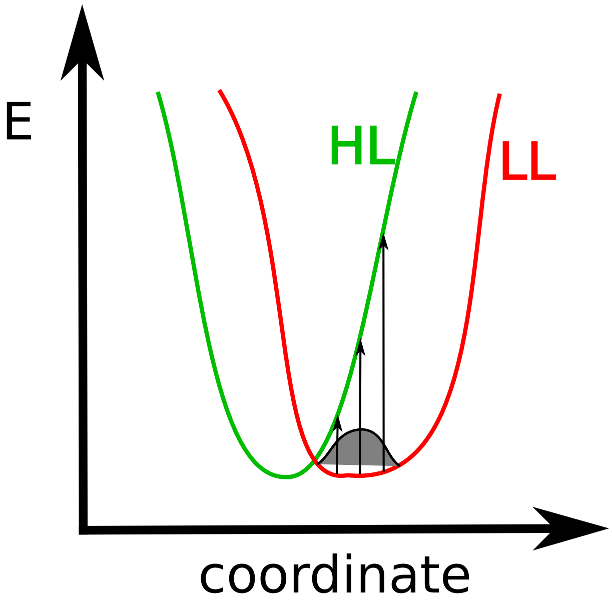


FIGURE 5.8 Schematic representation of the sequential QM/MM approach in the case of a strong dependence of the sampling on the level of theory. The low level (LL) is shown in green, the high level (HL) in red and the sampled distribution in gray. The energy corrections are represented by the arrows.

CHAPTER 6

QM/MM Calculations of Electronic Spectra in Solution

It is well known that the environment has a great influence on absorption and emission spectra. [151, 152] The associated change of the colour of a substance in liquid phase is commonly termed solvatochromism. The solvatochromic shift describes the shift of an absorption band either upon going from gas phase to liquid phase. The ground state and excited state are stabilised differently in solution which can cause a blue shift — a decrease in the wavelength — or a red shift which corresponds to an increase in the wavelength. Therefore, solvatochromism is a measure for the solute-solvent interactions [152, 153] and experimentally chromophores have been used as probes in order to study these interactions. [154–157]

A theoretical description of solvatochromism is challenging due to the combination of many different interactions and dynamic effects. [151] For the description of the environment, continuum models have been used [42, 48, 158] as well as hybrid QM/MM approaches [159–161] including sequential QM/MM studies. Continuum models have the obvious advantage of computational efficiency. Additionally, they allow modelling the important fast response of the continuum. This effect approximates the response of the electronic density of the solvent due to the changed charge distribution of the solute upon excitation with the usual consideration of the mutual polarisation in a self consistent manner. The slow term of the continuum is only determined by the ground state and therefore this solvation is also termed non-equilibrium solvation.

However, specific interactions cannot be described accurately, which is the strength of explicit solvent models like QM/MM methods. Combined with a fixed charge force field the important polarisation of the solvent is, however, not taken into account. Only a polarisable force field or a QM system including a number of solvent molecules can describe specific interactions as well as polarisation effects. In the sequential QM/MM approach the computational costs of the sampling are reduced by carrying out classical simulations to generate a set of representative configurations. However, as pointed out before and demonstrated on the example of methylchloride (Section 5.2) the configuration space can be very different and a simple correction of the energies with a high-level method cannot correct for this error. A practical limitation is that few force field parameters are available for solutes in their excited state which are required for the simulation of emission spectra. One way around this is to parametrise a force field for ground state as well as the excited state which has been done successfully (e.g. Reference [162]) but requires a tedious parametrisation procedure for every solute and excited state. Finally, QM cluster approaches may be used which, however, restrict the study of the potential energy to a local minimum or mostly a limited number of snapshots. However, in principle this is the most accurate description although this comes with a higher computational cost.

Here, I will explore the applicability of the PMC method which combines the accuracy of full QM/MM methods with the computational efficiency of perturbation theory demonstrated in Section 3.2. First, the computational approach will be explained. Second, a number of benchmarks systems will be studied and third, the method will be extended to solvent mixtures. This is a particular strength of explicit solvent models because implicit solvation models cannot account for these heterogeneous environments. Special sampling techniques are implemented to make the sampling of solvent mixtures very efficient.

6.1 The PMC Methodology - Absorption and Emission Spectra

A short introduction illustrating the theoretical basis of TD-DFT has been given in Section 2.6. Here I will explain how electronic spectra can be computed in the framework of PMC simulations applying the TD-DFT method. Equivalently, any other method being able to compute excited state densities in the context of QM/MM calculations may be used instead. However, I will focus on the TD-DFT method which strikes a good compromise between accuracy and computational costs.

The working equations formulated in a matrix notation are given by the eigenvalue equation

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^* & \mathbf{A}^* \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} = \omega \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \tag{6.1}$$

whereby matrices $\mathbf{X}$ and $\mathbf{Y}$ describe the linear response of the density matrix. The density matrix of the excited state is then approximated by summing up the ground state density and the unrelaxed difference density matrix in the following manner

$$\bar{D}_{ab}^{(n)} = \sum_i (\mathbf{X} + \mathbf{Y})_{ia} (\mathbf{X} + \mathbf{Y})_{ib} \tag{6.2}$$

$$\bar{D}_{ij}^{(n)} = \sum_a (\mathbf{X} + \mathbf{Y})_{ia} (\mathbf{X} + \mathbf{Y})_{ja} \tag{6.3}$$

with the assumption that the left and right solutions are identical. The density matrix $\mathbf{D}^{(n)}$ of the excited state n is then

$$\mathbf{D}^{(n)} = \mathbf{D}^{(0)} + \bar{\mathbf{D}}^{(n)} \tag{6.4}$$

which allows, after transformation into the atomic orbital basis, computing of the TD-DFT excitation energy $\omega_k^{(n)}$ for the PMC step k

$$\omega_k^{(n)} = \omega_0^{(n)} - \sum_{\mu\nu} \bar{D}_{\mu\nu}^{(n)} \left\langle \mu \left| \sum_{i\alpha \in m} q_\alpha \left(\frac{1}{r'_{i\alpha}} - \frac{1}{r_{i\alpha}} \right) \right| \nu \right\rangle \tag{6.5}$$

where $\omega^{(n)}$ corresponds to the excitation energy computed in the most recent update. For periodic systems the shifted force operator will be used instead and the excited state density is evaluated as well as the ground state density on a numerical grid.

The simulation of absorption spectra uses the ground state density in order to sample the configuration space while the respective excited state density may be used to compute emission spectra. Therefore, the computational costs for the perturbative steps increase by about a factor of two because in every step the electrostatic interaction of the target state's electron density with the environment has to be computed. The kernel, however, is identical and only the host program and the decision kernel have been modified for the implementation. A simple optimisation recomputes the excited state energy only if a MC move has been accepted and the solvent configuration changes. An interface to the Molpro program package has been implemented for TD-DFT calculations as well as Equations of Motion - Coupled Cluster with Singles and Doubles (EOM-CCSD) theory. The implementation of TD-DFT in Molpro uses density fitting for the electron repulsion integrals, approximates the exchange-correlation kernel by the adiabatic local density approximation and assumes that the matrices $\mathbf{A}$ and $\mathbf{B}$ commute which may cause some deviations of the excitation energies for larger systems and basis sets. [93]

6.2 Applications in Solute-Solvent Systems

6.2.1 Formaldehyde

Formaldehyde is an excellent benchmark system because of its reduced size. This allows employing high-level QM methods to describe the electronic excitations. However, the comparison with experiments is problematic because no experimental value can be measured in aqueous solution due to the formation of oligomers. Therefore, the experimental solvent shift of acetone in aqueous solution is used as a reference. Nevertheless, extensive theoretical results with many different combinations of QM methods and embedding approaches are available and allow evaluating at the very least which computational choices might have a strong influence on the results and which approximations can be made without further consequences.

The first three excitations of formaldehyde in aqueous solution and gas phase have been studied with the PMC approach. The solute geometries have been optimised with B3LYP-D3/def2-TZVP [17, 20–22, 111–114] while the simulations in PMC used the aug-cc-pVDZ [163] basis set, OPLS-AA vdW parameters, TIP3P water [29–31]and a temperature of 298 K. An additional classical MM MC simulation with the OPLS-AA force field has been carried out. The edge length of the box has been 40.0 Å and it contained 2132 water molecules. Snapshots of this simulation have been used for QM/MM TD-DFT calculations in a sequential QM/MM-type approach. The resulting absorption spectra are shown in Figure 6.1.

TABLE 6.1 Theoretical studies for the first n-π^* excitation of formaldehyde in gas phase ω^{gp} and the solvent shift $\Delta\omega$ in aqueous solution compared with PMC and sequential QM/MM simulations from this work both with B3LYP/aug-cc-pVDZ denoted by * and experimental results for formaldehyde in gas phase and the solvent shift for acetone as well as for highly concentrated formaldehyde.

Ref.	QM	Embedding	Sampling	ω^{gp}/eV	$\Delta\omega$/eV
[164]	CIS	QM MBE	——	4.47	0.35
[165]	CIS(D)	MLFMO	FMO-HF	4.08	0.14
[164]	TD-DFT(B3LYP)	QM MBE (30 H_2O)	——	3.81	0.13
[164]	TD-DFT(B3LYP)	QM MBE (81 H_2O)	——	3.81	0.16
[164]	EOM-CCSD	QM MBE	——	3.92	0.15
[166]	HF & EHP	QM Cluster	MM		0.39
[150]	CIS(INDO)	QM Cluster	MM		0.27
[167]	CASSCF(4,3)	PCM		4.03	0.12
[168]	CASSCF(6,5)	RISM		4.33	0.25
[169]	CASSCF(6,4)	QM/MM, ASEP	MM	4.04	0.18
[170]	CASSCF(12,10)	QM/MM-pol-vib	MM	4.53	0.33
[171]	RHF & ROHF	QM/MM	MM	3.47	0.24
[172]	AM1 & CIS	QM/MM-pol		4.07	0.14
[173]	MCSCF(4,3) & MRCI(4,3)	ASEP	QM/MM	4.09	0.19
[174]	CASSCF(12,10) & CASSI	QM/MM/cont.			0.16
[175]	LR-CCSD	QM/MM-pol	MM-pol	3.99	0.35
*	TD-DFT(B3LYP)	PMC		3.94	0.18
*	TD-DFT(B3LYP)	QM/MM	MM	3.94	0.10
[176]	Exp. (Formaldehyde)			4.07	
[177]	Exp. (Formaldehyde)				0.21
[178]	Exp. (Acetone)				0.16

Furthermore, the absolute excitation energy in gas phase as well as the solvent induced shift for the first n-π^* excitation have been compared with theoretical studies from the literature in Table 6.1 and experimental results for formaldehyde and acetone.

Taking a closer look at the first n-π^* excitation I first want to consider the differences between the results in gas phase due to the QM method. Ideally, this will allow in a second step discerning effects of the electronic structure method and the solvent model on the solvent shift. The reference for the gas phase excitation is the experimental value of 4.07 eV. [176] It is expected that the most reliable theoretical results are the MRCI(4,3) calculations on top of MCSCF(4,3) which gives a value of 4.09 eV that is in excellent agreement with the experimental results. [173]

Next, a number of CASSCF based studies are considered. Generally, the results should improve with the number of orbitals included in the active space and the number of electrons under consideration. However, it seems like the largest active space results in the largest deviation of about 0.5 eV in the case of the CASSCF calculation with twelve electrons in ten orbitals. [170] On the other side the results with four electrons in three orbitals are very close to the experimental results. [167] It is important to realise that CASSCF only accounts for static correlation and has be shown that the vertical excitation energies are strongly overestimated without further methods like CASPT2 or NEVPT2. These are based on top of CASSCF and account for the important effects of the dynamic correlation. [179, 180] With this in mind the results from the literature can be understood and the calculations with small active space rely on the error cancellation between the missing dynamic correlation and the inaccurate static correlation. However, it is questionable how reliable this error cancellation is under the influence of the solvent model.

Another group of methods including TD-DFT, EOM-CCSD and linear response (LR) CCSD is able to predict absolute excitation energies with an error of up to 0.3 eV. The largest error is found here in the case of the TD-DFT calculations. However, it is expected that this is a systematic error which has the same size in gas phase as well as in solution. DFT methods are generally able to reproduce the electronic density with good accuracy and the interaction with the environment is based on this quantity. Therefore, only a minor influence on the accuracy of solvent shifts is expected. Another point of interest are the results based on the semiempirical AM1 Hamiltonian. They suggest that a carefully chosen semiempirical method can give reliable results at hugely reduced computational costs. However, careful and extensive benchmarking is necessary in order to ascertain that the results are consistent and reliable for the properties of interest. It is expected that all these methods are intrinsically able to predict accurate solvent shifts without considering the influence of the solvent model.

Finally, the remaining studies have been carried out using the methods restricted open shell HF, the electron-hole potential in combination with HF and Configuration Interaction with Single Excitations (CIS). These methods predict the excitation energies qualitatively wrong and a considerable influence on the solvent shifts is expected as well. While the computational costs are clearly reduced, the loss in accuracy does not outweigh this gain in efficiency. Furthermore, even sophisticated solvent models cannot lead to reasonable results if they are based on qualitatively wrong methods describing the electronic excitation. Consequently, the influence of the solvent model and the electronic structure method on the solvent shifts cannot be discerned and they will not be included in the following considerations.

Next, I will consider the differences between the predicted solvent shifts. A difficulty is that no experimental value is available and furthermore there is no clearly superior approach among these theoretical studies. It is generally assumed that the solvent shift of acetone of 0.16 eV [178] can serve as a rough estimate for the shift of formaldehyde. A second reference is the experimental value measured in highly concentrated formaldehyde solution of 0.21 eV [177] although it is assumed that this corresponds to formaldehyde surrounded by formaldehyde molecules due to oligomerisation. Nevertheless, many of the theoretically predicted shifts are in the range of 0.12–0.25 eV.

The CASSCF calculations with larger active space overestimate the solvent shifts as well. This suggests that the consideration of dynamical correlation is important for the accurate prediction of solvent shifts. Because of the rearrangement of the electronic density upon excitation there is a difference in the dynamic correlation between ground state and excitation state. Therefore, the effect of the different solvent models PCM, RISM, sequential QM/MM and the Averaged Solvent Electrostatic Potential (ASEP) cannot be separated from the intrinsic error of the QM method.

The studies employing a QM description for the solute as well as the solvent have the most accurate description of the interaction between these two systems. [164, 165] Various many-body expansion (MBE) or fragmentation schemes can be used like the fragment molecular orbital approach. Nevertheless, the computational costs prohibit the sampling altogether or allow it only at a very low level like HF. These studies agree closely with their predictions of 0.14–0.16 eV. The effect by increasing the number of water molecules from thirty to eighty-one is about 0.3 eV which suggests that these calculations are not yet converged with regard to the number of solvent molecules.

The most accurate sampling is used in the study by Xu and Matsika [173] with QM/MM simulations from which an ASEP has been constructed. The prediction of 0.19 eV agrees closely with the PMC result of 0.18 eV which is based as well on a QM/MM sampling. The sequential QM/MM approach with equivalent computational settings allows to assess the effect of the sampling on the excitation energies. This is rather large with about 0.08 eV. In conclusion, sequential QM/MM studies require very careful benchmarking of the classical potentials which are used to generate the configurations used for high-level methods. Otherwise, the introduced error can be easily larger than any error due to the electronic structure method.

The complete spectrum for the first three excitations is shown in Figure 6.1. Also the second band is shifted compared to the sequential QM/MM approach but in this case to lower excitation energies. Otherwise, one can note that only with considerable sampling well converged spectra can be obtained as in the case of the PMC simulations. Even using a rather large number of configurations generated by classical simulations is not enough. Another important point is, that the oscillatory strength seems to be sensitive to the configuration space and that intensities of the bands relative to each other are different in the PMC approach. Last, the band shape is influenced as well and e.g. the band of the third excitation is slightly broadened which reflects as well the different configuration space.

The obtained results and the comparison with a number of theoretical and experimental results shows that TD-DFT can be used in combination with PMC simulations to obtain complete electronic spectra on the fly and including the information about the band shapes. This is possible at reduced computational costs compared to conventional QM/MM simulations. The comparison with sequential QM/MM simulations showed that the effect of the potential used for the sampling has a large effect on the electronic excitations. However, it became also clear that the computation of electronic spectra is far from trivial. Moreover, no dominant effect can be singled out on which should be in the focus of further developments. The biggest limitation in our approach is the neglect of the solvent's polarisation. It has been argued that this polarisation has also a large effect on the exploration of the configuration space and only a minor effect if it is applied in a sequential manner. [175]

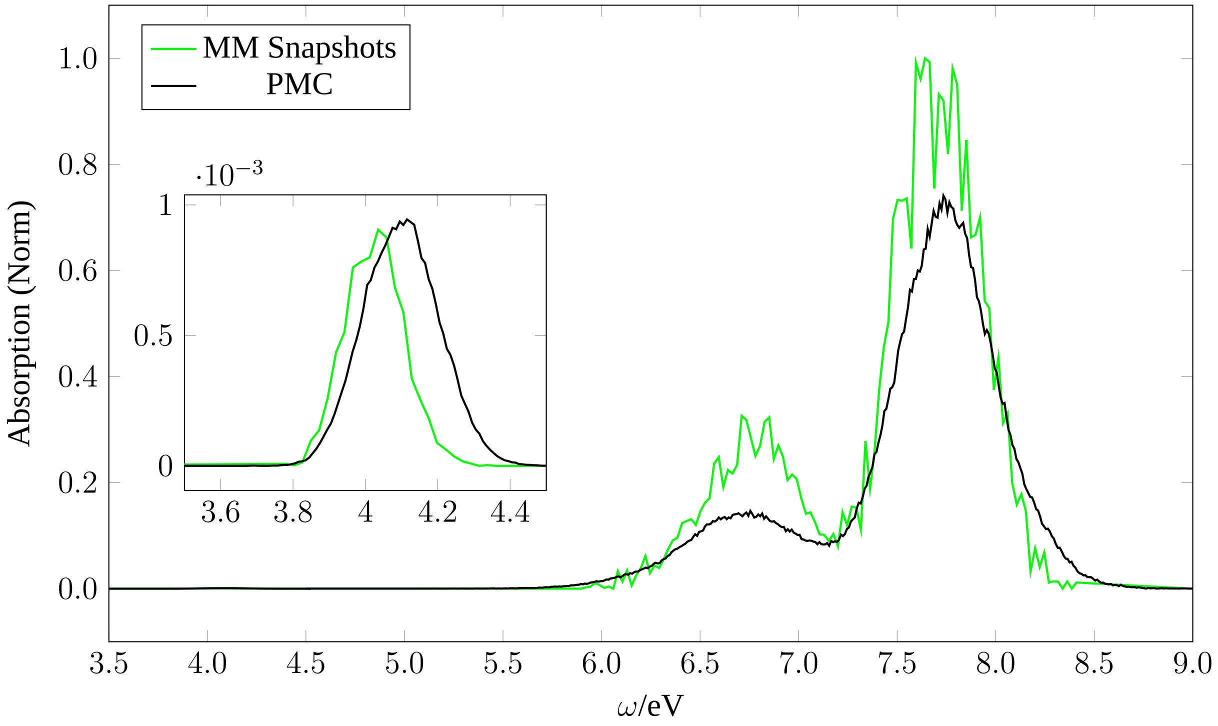


FIGURE 6.1 Electronic excitations for formaldehyde in TIP3P water computed for snapshots of MM MC (OPLS-AA) simulations and based on PMC (B3LYP/aug-cc-pVDZ and OPLS-AA vdW parameters) simulations.

6.2.2 Propenal

Propenal, better known as acrolein, is the simplest unsaturated aldehyde with the *trans* isomer being preferred at room temperature. Interestingly, the solvent shifts of the first two excitations are with opposing signs which is typical for n-π^* and π-π^* excitations. This illustrates that these excitations behave very differently under the influence of the solvent and pose a challenge for theoretical models aiming at predicting these effects.

The experimental value of the first n-π^* excitation in vacuum is 3.69–3.71 eV [181–184] and the second π-π^* excitation is well separated at 6.41–6.49 eV. [181, 185, 186] Both solvent shifts for aqueous solution have been determined experimentally as well. The first excitation in solution has been placed at 3.86–3.91 eV [187] or 3.94 eV [181, 188] which results in a solvent shift of 0.15–0.22 eV or 0.23–0.25 eV respectively. The second excitation has been measured at about 5.9 eV [181, 188] which gives a shift of -0.51 to -0.59.

The small size of the solute allows high-level theoretical calculations in order to study the complex electronic structure. This combined with the strong solvent effects aroused the interest of the theoretical chemistry community. A selection of theoretical studies will be discussed in the following text. A sequential QM/MM approach with TD-DFT with B3LYP has been carried out by Canuto and coworkers and they determined a solvent shift for the n-π^* excitation of 0.2 ± 0.1 eV [189] in reasonable agreement with the experimental findings. CASPT2 combined with PCM slightly overestimates the first excitations with 0.33 eV and underestimates strongly the red shift with only -0.1 eV. [190] However, the direction of the shifts is correctly reproduced. MRCISD+Q combined with the COSMO approach seems to improve upon this for aqueous solution with 0.21 eV and -0.43 eV respectively. [191] All these findings show that the solvent shift of the first excitation can be well described with QM/MM methods as well as continuum models.

On the other hand, the modelling of the solvent effects for the π-π^* excitation is more challenging. It has been found by Mikkelsen and coworkers by comparing TD-DFT and coupled cluster calculations that only the first excitation is well described by TD-DFT. Mainly electrostatic interactions cause the solvent shift of the first excitations while specific interactions are important for the shift of the second excitations. [192]

Both solvent shifts are sensitive to the number of solvent molecules considered in the theoretical treatments. Aguilar and coworkers found that the first solvation shell accounts only for about 35% of the first solvent shift. [193] The second excitation is even more sensitive in this regard and a number of studies confirmed that the shift converges slowly with the number of water molecules. [192, 194] It has been found by Mata [195] that up to fifty water molecules have to be included in the QM system in order to obtain converged results with about -0.56 eV with a many-body expansion based on EOM-CCSD. Canuto and coworkers carried out CIS(D) calculations on clusters generated by MC simulations. They attempted to extrapolate the solvent shift to the limit of a complete embedding and obtained a shift of -0.52 eV in good agreement with the experimental results. [196]

In this work PMC simulations with 160 M steps at 298 K in the canonical ensemble combined with TD-DFT have been carried out to study the first two excitations of propenal in gas phase and aqueous solution. The B3LYP-D3 functional [17, 20–22, 112–114] has been used with the aug-cc-pVDZ basis set, [163] OPLS-AA vdW parameters [29–31] and the TIP3P water model. [137] The obtained spectra is shown in Figure 6.2.

Computing the solvent shifts based upon the maxima of the peaks gives a value of 0.14 eV for the first excitation. This confirms that the main interactions are of electrostatic nature which are very well captured with the QM/MM approach and result only in a slight underestimation of the shift compared to the experimental findings. This remaining deviation may be due to polarisation effects of the solvent or dynamic effects including solute as well as solvent molecules. On the other hand, the red shift is strongly underestimated similar to previous studies with only -0.12 eV. This result is comparable to the findings of the PCM study although the continuum model captures different physical effects. Keeping in mind both the results from the continuum model and from the QM/MM simulations can give further insight into the physics of the solvent shift because they cover different solvent effects. Here it suggests that specific interactions as well as polarisation effects are equally important and have to be treated on an equal footing in order to obtain accurate results for this π-π^* and similar excitations.

Furthermore, PMC simulations pave the way to simulate complete electronic spectra in solution and give information about the band shapes as well. In particular the asymmetry of the n-π^* band cannot be easily predicted without extensive sampling of the configuration space. Furthermore, the configurations generated with the efficient PMC approach may serve as snapshots for more accurate methods. This is in the spirit of the sequential QM/MM approach but with a higher level of theory already used for the sampling.

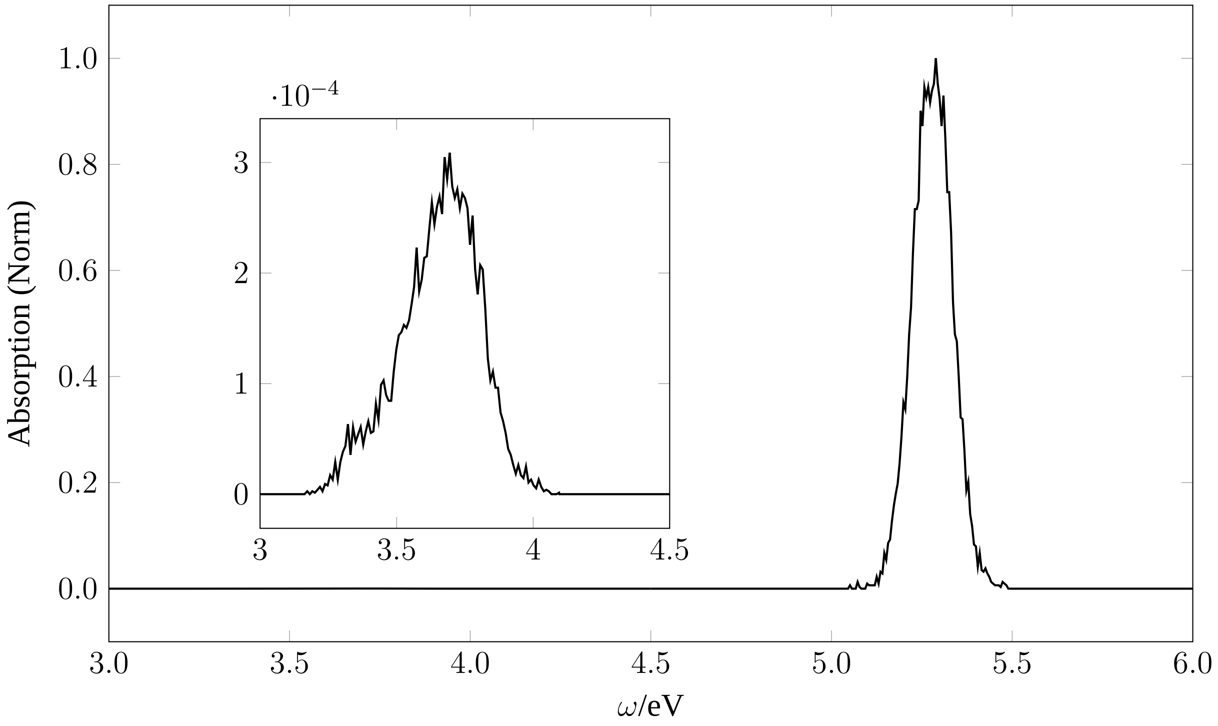


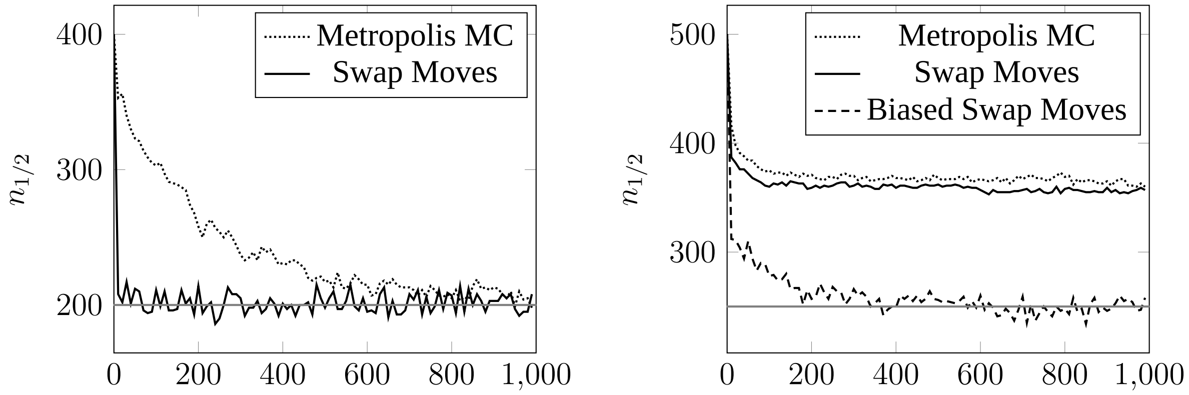
FIGURE 6.2 Electronic excitations for propenal in TIP3P water computed on-the-fly with PMC (B3LYP/aug-cc-pVDZ and OPLS-AA vdW parameters) simulations.

6.3 Applications in Binary Solvent Mixtures

In synthetic chemistry there is a considerable interest in solvent mixtures which may serve as tunable reaction media. Many properties e.g. transport properties, the dielectric constant or refractive indices can be conveniently adjusted simply by changing the composition of the involved solvents. [197–200] Solutes immersed in such a media can be very sensitive to the composition as in the case of bis-triazinyl-pyridine which has different equilibrium conformations depending on the methanol/water mixture. [201] Finally, interfaces formed between water and organic solvents show intriguing catalytic effects due to the very specific environment of dangling hydrogen bonds at the interface. [202] Simulations allow to study solvent mixtures and reveal the molecular mechanisms of the aforementioned phenomena which are difficult to be investigated experimentally. However, the long time scales on the order of nanoseconds of the diffusion controlled mixing processes pose their own challenge for any theoretical simulation. This issue will be addressed in the following Section by a specialised sampling procedure. Thereupon, this procedure will be applied to study the influence of binary solvent mixtures on electronic excitations.

6.3.1 Sampling of the Conformational Space

Based on a simple benchmark system I will investigate the mixing process of a binary solvent mixture by increasing stepwise the complexity of the system. All the following simulations have been carried out with a temperature of 298 K. The first system under consideration will be a Lennard-Jones (LJ) liquid. For simplicity I will start with a system of 400 physically identical



(a) 400 identical LJ particles (b) 1350 small and 500 large LJ particles

FIGURE 6.3 Simulations with conventional Metropolis MC and swap moves analyzed every 10 k steps. The number of particles in one half of the system $n_{1/2}$ is measured.

particles. However, the label A will be assigned to one half of the particles and B to the other. This allows measuring the mixing process between the particles A and B simply by counting the number of particles in each half of the simulation cell $n_{1/2}$. The process is only controlled by the diffusion because the particles A and B are identical and the free mixing energy is zero.

The MC approach allows — as opposed to MD simulations — introducing arbitrary steps in order to generate new configurations. These steps do not have to correspond to any physical process and only need to generate the correct ensemble. An obvious possibility is to use a MC move that swaps a particle A with a particle B. In fact, moves that exchange two particles have been used before usually in the context of low temperatures and therefore slow dynamics or in systems with inherently slow dynamical processes e.g. glasses or supercooled liquids. [203–205] In Figure 6.3a the standard protocol which includes only translation steps is compared with a modified one that attempts to swap two different particles in every 10^{th} step. For this particular system the swap is not associated with any change in the total energy because all particles are identical. Therefore any attempted swap is accepted and the system is mixed within the first 20 k steps. The standard protocol on the other side requires already about 10 M steps to reach equilibrium for only 800 LJ particles.

However, the assumption of identical particles cannot be extended to binary solvent mixtures. Even if the involved molecules have a similar volume the interactions can be very different considering water and organic solvents. Therefore, a more complex model of 1350 small LJ particles and 500 ones which are twice as large will be investigated. Conventional and swapping moves of a small with a large particle are compared in Figure 6.3b. Only a very minor improvement due to the swaps can be seen in the beginning of the simulation after which the swaps do not have any influence. This can be easily understood because only 0.005 % of the attempted swaps are accepted at all. While a smaller particle fits well into the cavity of the larger particle, the larger one rarely fits into the cavity of the small one. Every swap is combined with a large increase in the potential energy and is therefore rejected.

One way to improve the protocol may be to swap one large particle with two smaller ones. However, this approach still requires that the volume of the large particle is a multiple of the smaller one and is not generally applicable. Therefore, the low acceptance ratio will be addressed by introducing a bias. This biased MC method has been introduced first by Rosenbluth and Rosenbluth [206] which used it to generate new conformations of polymers which are rejected as well with a high probability due to overlapping atoms. The generalised configurational bias MC has been used to swap large particles with a number of smaller ones [207] and recently to study phase equilibria in the Gibbs Ensemble. [208]

The bias is introduced by the following procedure. Trial configurations are generated for the old configuration (before the swap) as well as for the new configuration (after the swap). The trials are generated by randomly translating both particles while molecules would be also rotated in this step. The weight of a configuration, denoted as the Rosenbluth weight, is defined as

$$W(n) = \sum_{j}^{k} \exp\left(-\beta V_j\right) \tag{6.6}$$

for k trials. Two random configurations are selected out of the new and old trials respectively with the probability

$$p_n = \frac{\exp\left(-\beta V_n\right)}{W_n}. \tag{6.7}$$

Finally, this randomly selected new trial is accepted with the probability $p(o \to n)$

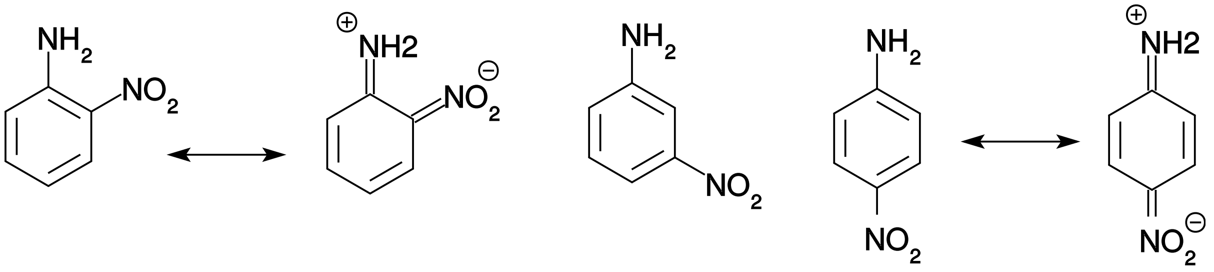
$$\frac{p(o \to n)}{p(n \to o)} = \frac{W(n)}{W(o)} \exp\left(-\beta(V_n - V_o)\right) \tag{6.8}$$

which shows that the acceptance of the new configuration is proportional to the weight while Equation 6.7 guarantees that energetically favourable trials are accepted with higher probability. The trials help to sample the local partition function. If any trial in the new configuration is energetically favourable the Rosenbluth weight and therefore the acceptance are increased. It is important to keep in mind that only the acceptance is biased but the underlying transition matrix and therefore the obtained distribution are not changed. The results are also independent of the number of trials. Increasing the number of trials generates, however, a better estimate of the local partition function. It is possible to use a different temperature which allows to accelerate mixing processes or a simplified potential for the trials reducing the computational costs of generating them.

In Figure 6.3b the results for a biased simulation are shown. Every 10 steps a biased swap with 10 trials is attempted. This increases the acceptance of the swaps to 7.3 % and the system converges in about 6 M steps to the equilibrium. The computational cost for a single trial is similar to a conventional MC step. The simulation time is therefore increased and in this case comparable to 11.4 M conventional steps. However, the convergence is still very much accelerated. The devised sampling procedure will be used in the following Section for realistic solvent applications.

6.3.2 Nitroaniline

Nitroanilines are prototypical examples of so called push-pull chromophores. They are composed of an electron-donating and -accepting groups connected by an aromatic system. The three different isomers 2-nitroaniline (2-NA), 3-nitroaniline (3-NA) and 4-nitroaniline (4-NA) are shown in Figure 6.4. However, only for 2-NA and 4-NA the shown mesomeric structures can be formulated where formally one electron is transferred from the amino group to the nitro group. This has an influence on the first electronic excitation and it has been found that the charge transfer character is considerably smaller for 3-NA. [209] Because the charge transfer states — especially of 2-NA and 4-NA — are very sensitive to the environment they may serve as probes for solute solvent interactions. Therefore, they received the attention from both experimental [154] and theoretical groups. [210–214]



2-NA **3-NA** **4-NA**

FIGURE 6.4 Lewis structures of the three isomers of nitroaniline — 2-nitroaniline (2-NA), 3-nitroaniline (3-NA) and 4-nitroaniline (4-NA) — and the mesomeric structures for 2-NA and 4-NA.

Two aspects will be investigated in the following. First, the solvent shift upon changing from an apolar solvent i.e. cyclohexane to a slightly more polar one i.e. tetrahydrofuran. The experimental shifts for 2-NA [215] and for 3-NA and 4-NA [216] are given in Table 6.2. Computations have been carried out with TD-DFT with the B3LYP functional [17, 20–22] and the aug-cc-pVDZ basis set. [163] The solvent effect has been modelled with COSMO [50] with $\epsilon(\mathrm{cyclohexan}) = 2.02$ and $\epsilon(\mathrm{cyclohexan}) = 7.43$ and additionally with PMC simulations with the TIP3P water model and updates of the density every 20 k steps. The simulations for 3-NA and 4-NA have been carried out by Johannes Kircher during his bachelor thesis [217] under my supervision.

The COSMO approach reproduces qualitatively the correct trend for 2-NA and 3-NA although both shifts are underestimated by about 0.08 eV. Furthermore, the shift of 4-NA is pre-

TABLE 6.2 The solvent shifts in eV upon changing from cyclohexane to tetrahydrofuran from experiments (2-NA, [215] 3-NA and 4-NA [216]), COSMO (B3LYP/aug-cc-pVDZ) and PMC (B3LYP/aug-cc-pVDZ and OPLS-AA vdW parameters) calculations.

Isomer	Experimental	COSMO	PMC
2-NA	−0.18	−0.10	0.01
3-NA	−0.31	−0.23	−0.46
4-NA	−0.59	0.04	0.03

dicted close to zero although experimentally this shift is the largest one. The PMC simulations describes the isomers with larger charge transfer character (2-NA and 4-NA) qualitatively wrong and predicts shifts close to zero. However, the shift of 3-NA is estimated at about -0.46 and overestimates even the experimental value. Especially for the charge transfer states inductive effects of the solvent are important which can only be captured by a polarisable force field. However, the drastic failure of both continuum model and QM/MM scheme including PMC suggest an underlying problem in the computational approach. It has been argued that TD-DFT even with the improved CAM-B3LYP functional fails for charge transfer states of 4-NA. The reason lies in the overestimation of the charge separation in the ground state and at the same time an underestimation in the excited state which is related to the response of TD-DFT to the polarisable embedding. [214]

Second, the preferential solvation for the solvent mixtures of cyclohexane and tetrahydrofuran have been studied. In an ideal solvent the dielectric constant of the mixture is given as a linear combination of the dielectric constant of the pure solvents weighted by the mole fractions. If the main interaction between the solute and the solvent can be described as the interaction of a dipole with a homogeneous continuum then the solvent shift should be linear with the mole fraction. However, even in ideal solvent mixtures non-linear effects are caused by the preferential solvation.

It has been suggested that a general mechanism independent of solute and solvent is one reason for the preferential solvation which is termed dielectric enrichment. Due to the different polarity of the solvents one of them prefers to interact with the solute. This stabilisation is balanced by an accompanying loss in the entropy of mixing of the solvents. [215, 218, 219] However, this explains only part of the physics. It has been found that if specific interactions between the solute and solvent are present they usually tend to dominate the preferential solvation. These interactions can include for example hydrogen bonds or electron donor-acceptor interactions. [215]

Explicit solvent models are in principle capable of describing these interactions and homogeneous environments that may occur in the solvent mixtures. Therefore, the solvent mixtures have been analysed as well with PMC simulations with B3LYP/aug-cc-pVDZ and OPLS-AA vdW parameters and the obtained results are shown in Figure 6.5. The edge length of the box ranges from 49.15 Å with 680 molecules cyclohexane to 46.98 Å with 745 molecules tetrahydrofuran. 3-NA is most relevant in this case because it has the largest shift as predicted by the simulations. One can see the effect of the preferential solvation when the shift deviates by about 0.1 eV from the linear shift. Interestingly, a saturation effect can be observed when the mole fraction of tetrahydrofuran is increased even further. This can be understood in terms of the potential interaction sites of 3-NA. Mostly the polar amino and nitro group can interact with tetrahydrofuran and only a limited number of solvent molecules can coordinate at these groups. If the number of polar solvent molecules is increased even further only a slight effect on the shift is observed. If the complete environment is replaced by the more polar solvent one can observe a slightly increased shift due to the long-range electrostatic interactions which have a small influence but accumulate for many solvent molecules.

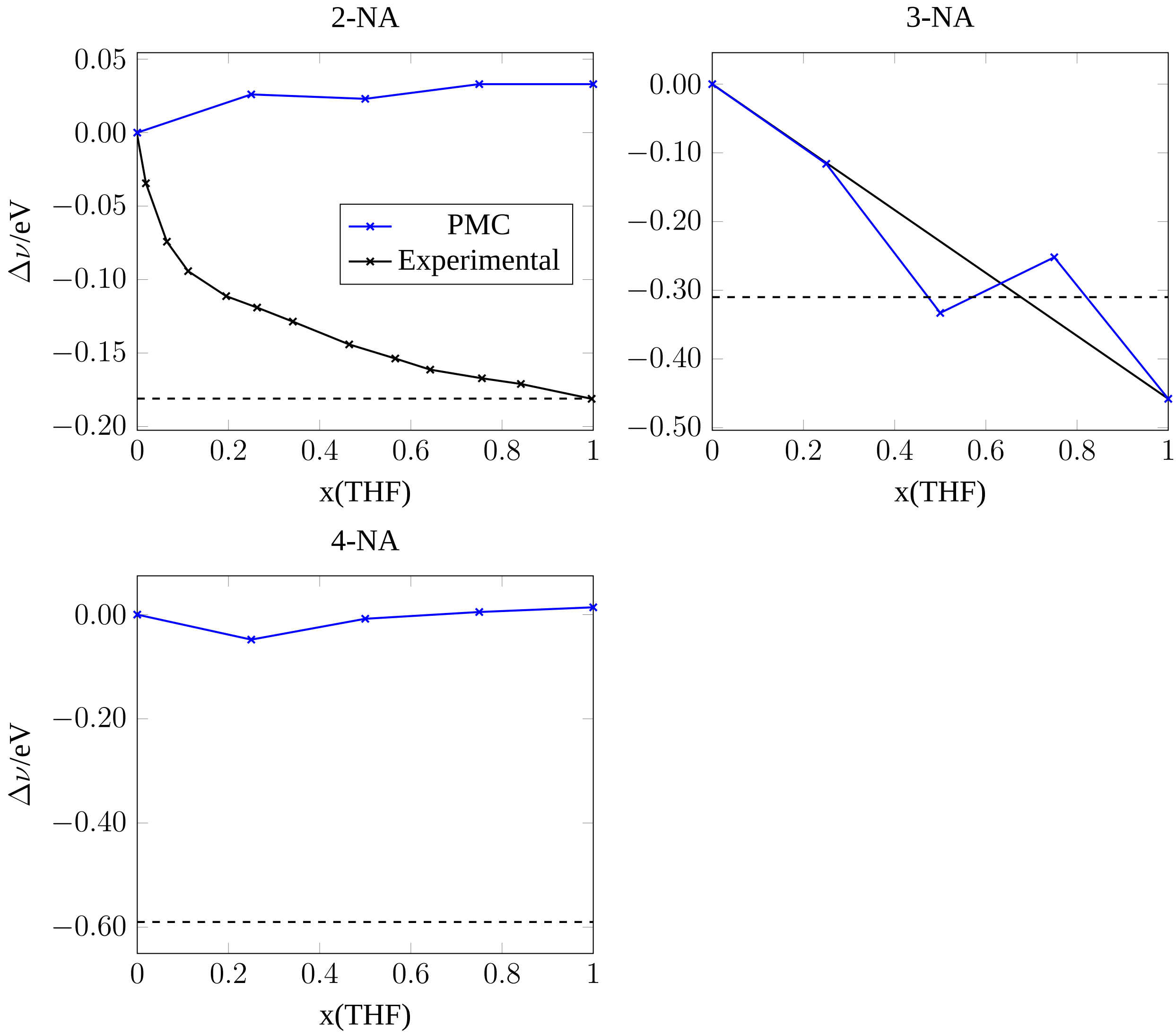


FIGURE 6.5 The PMC (B3LYP/aug-cc-pVDZ) computed solvent shifts in the binary solvent mixture of cyclohexane and tetrahydrofuran of 2-nitroaniline with experimental results [215] (top left), 3-nitroaniline (top right) and 4-nitroaniline (bottom). The dashed lines denote the experimental shift for x(THF)=1. For 3-NA the linear shift has been plotted as well with the continuous black line.

These results show that the devised sampling protocol can be used to describe electronic excitations in solvent mixtures. The largest limitation is certainly a static partial charge force field which does not capture important inductive effects for charge transfer states. Furthermore, the electronic structure might be insufficiently described by the TD-DFT method. The PMC simulations allow using more accurate and costly QM methods because of the efficient protocol that reduces the actual number of electronic structure calculations that have to be carried out. An interface for the EOM-CCSD level has been implemented which will be used to explore this aspect in future studies.

CHAPTER 7

Differential Solvation Effect of the Uracil Base

A Deoxyribonucleic Acid (DNA) molecule in its most typical form consists of two single strands which are coiled around each other forming a double helix. The basic units are the nucleotides which consist of a sugar, denoted deoxyribose, a phosphate group and one of four nucleobases. The nucleobases Guanine (G), Adenine (A), Thymine (T) and Cytosine (C) are shown in Figure 7.1. The helical structure of the DNA is maintained by the pairing of complementary nucleobases into the well-known Watson-Crick base pairs (see Figure 7.2). Hydrogen bonds comprise the main interaction between the nucleobases of different strands and control therefore the correct pairing between the nucleobases A-T and G-C.

The replication of a DNA molecule is illustrated in Figure 7.3. A plethora of enzymes is involved in this process. First the double helix is unwound by a helicase and topoisomerase. For each single strand a DNA polymerase creates the complementary sequence through complementary base pairing. That means the old strand determines the sequence of the new strand. Finally, a DNA ligase combines the separate fragments that have been created by the polymerase.

Ribonucleic Acid (RNA), which is closely related to DNA, contains Uracil (U) instead of T. In position 5 Uracil has a hydrogen instead of the methyl group that is found in thymine. However, because U and T are very similar U can be selected during the DNA replication process instead of the correct T. Thereby, it will be incorporated into the DNA sequence which is

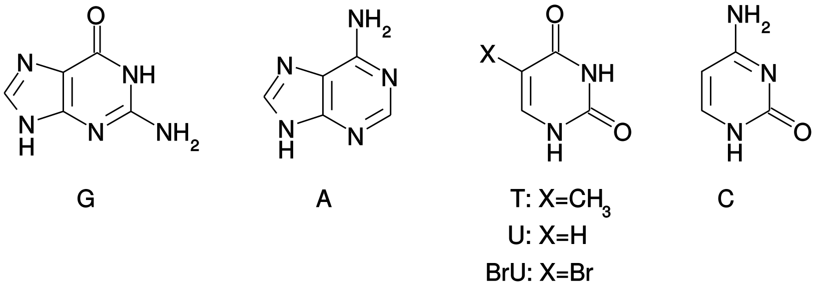


FIGURE 7.1 Lewis structures of the nucleobases guanine (G), adenine (A), thymine (T) or uracil (U) and cytosine (C).

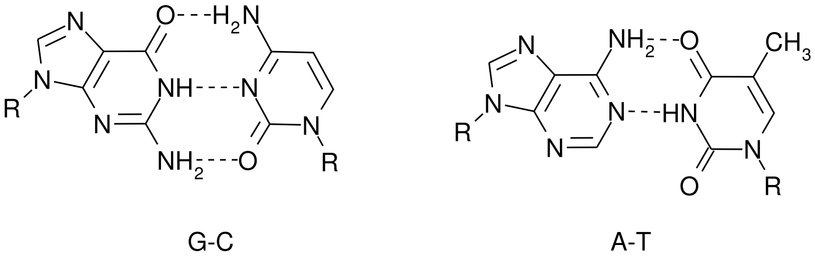


FIGURE 7.2 Lewis structures of the Watson-Crick base pairs as occurring in DNA with the backbone indicated by R. Hydrogen bonds are shown in dashed lines.

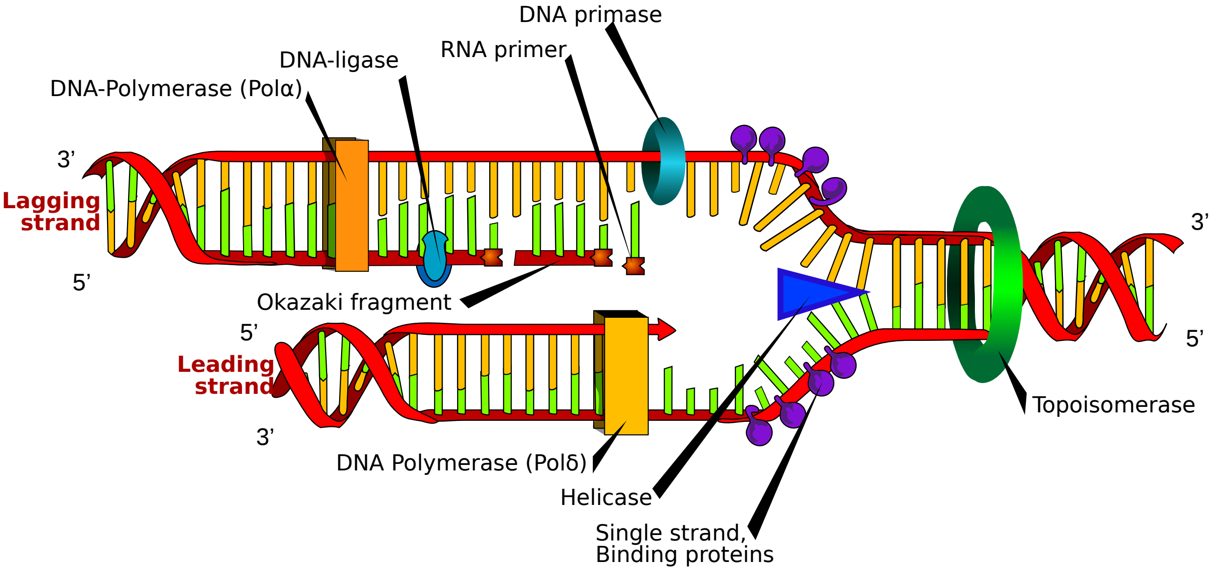


FIGURE 7.3 Schematic representation of the replication of DNA. This work has been created by Mariana Ruiz. [220]

one of the simplest modifications that can occur as confirmed by experimental studies. [221,222] This does not disturb the structure of the DNA. In fact, bacteriophages exist which have DNA that contains only U instead of T. [223] This mismatch is of special interest for this study because instead of U also its brominated species 5-Bromouracil (BrU) can be incorporated which may ultimately alter the sequence of the DNA.

It has been first suggested by Watson and Crick [224] that mutations of the DNA strand could be caused by rare tautomeric forms of the nucleobases. The probability for these imino or enol tautomers — estimated to be about $\approx 0.01\%$ — is very low but they could be essential for mutagenic properties of certain nucleobases. Recent experimental findings confirm based on X-ray crystal structures the rare tautomer hypothesis. [225] In this context halogenated nucleobases, which exhibit mutagenic properties, have received considerable attention leading to a number of experimental findings. 5-Chlorouracil, which is mutagenic, and also BrU — a well-known mutagenic agent [226] — are biologically very relevant and can actually be generated in human cells. Inflammatory conditions can lead to the production of hypochloric or hypobromic acid which can in turn attack the DNA double helix and finally lead to the formation of halogenated Uracil. Certain peroxidases can also form nitryl chloride which can chlorinate nucleotides as well. [227] The fluorinated species has been suggested to have mutagenic properties but recent experimental studies show no measurable effect. [228] Nevertheless, very little is known about the origin of the mutagenic properties and the exact mechanism leading to point mutations of the DNA.

The focus of this study will be on BrU which has been studied experimentally and theoretically more than any other halogenated uracil. Different models (ionisation model, tautomerisation model, wobble model) have been proposed in order to explain the mechanism of the mutagenic action. [229–232] However, the experimental evidence is ambiguous and the theoretical studies are contradictory at best. It is in most cases assumed that a certain unknown activation mechanism transforms U when the DNA is a single strand which in turn leads to the mismatch in the complementary base pairing as illustrated in Figure 7.4. These proposed models will be investigated using PMC simulations and described in more detail in the following Sections.

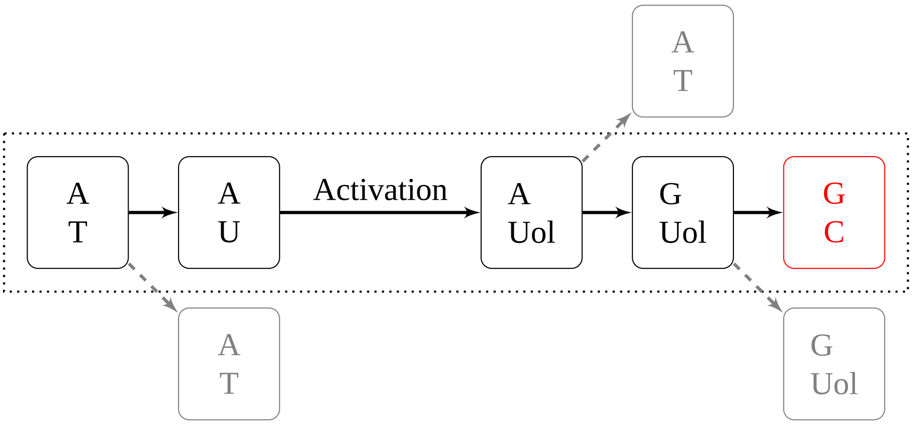


FIGURE 7.4 Schematic representation of the assumed mechanism that leads to point mutations during the DNA replication. Each branching represents a DNA replication and the activated enol Uol is one of the intermediates shown in Figure 7.5.

7.1 Computational Details

Theoretical studies of the photoionisation of DNA show that the aqueous environment very efficiently screens the stacking interactions between nucleotides. Therefore, nucleotides can be treated separately and depend only weakly on the particular DNA sequence. [233] However, local electrostatic interactions in DNA polymerases have been suggested to influence the occurrence of rare tautomers [225] and that the isolated treatment of a base pair in solution might neglect important effects of the environment. In this Chapter all QM calculations have been carried out with B3LYP-D3/def2-SVP [17, 20–22, 111–114] in combination with COSMO [41] ($\epsilon = 80.0$). Furthermore PMC simulations have been carried out with B3LYP-D3/def2-SVP in combination with the TIP3P water force field. [137] The Lennard-Jones parameters for the solute have been taken from the OPLS-AA force field. [29, 30] The simulations have been carried out with a temperature of 298 K and a box of 3243 water molecules with an edge length of about 46 Å.

As a correction for the rigid solutes used in the PMC simulations thermodynamic corrections have been derived from the partition function using the rigid rotor/harmonic oscillator/ideal gas approximation. The solute geometries for the PMC simulations have been taken from the COSMO optimizations. All frequencies have been considered for the computation of thermodynamic frequencies. This could cause generally a significant error but in this case only the differential thermal corrections from gas phase to solution are of interest. Therefore, this is a good approximation since even the relative geometry relaxation effect from gas phase to solution is only about 0.2 kcal/mol. [234] Finally, the difference between the continuum and the explicit solvent is expected to have an even smaller effect on the thermal corrections and is negligible.

All nucleobases are methylated in the 1-position in order to model the influence of the DNA backbone. The works by Hobza [232] and Tsukamoto [229, 235] and their respective coworkers use a hydrogen instead of a methyl group. However, the comparison shows that this influences the results very little.

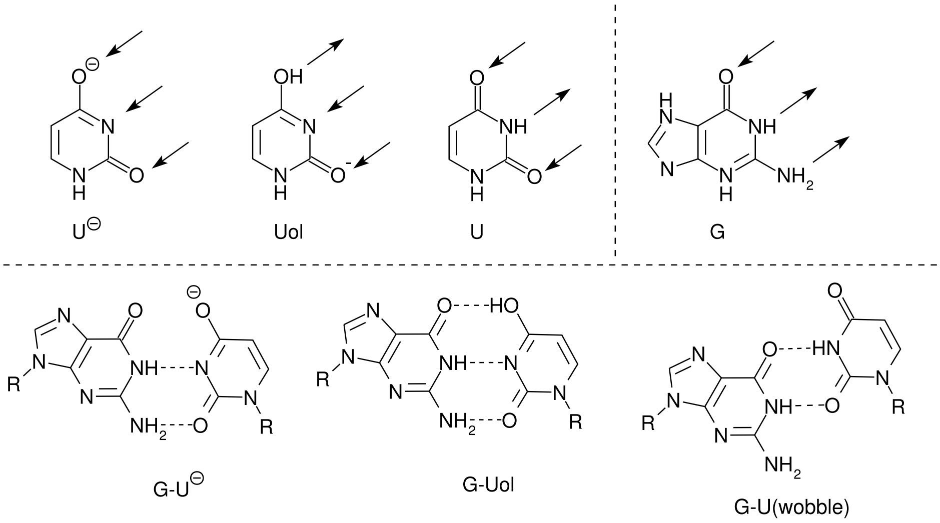


FIGURE 7.5 Top: Lewis structures of the intermediates corresponding to the ionisation (U^-), tautomerisation (Uol) as well as U and G with hydrogen bond donor and acceptor capabilities indicated by arrows. Bottom: The resulting base pairs.

7.2 Ionisation Model

The first of the investigated models is the ionisation model which proposes that uracil is deprotonated at the nitrogen in 3-position leading to U^-. This changes the donating NH hydrogen bond site to an accepting one (N) while the oxygen in 4-position becomes formally negatively charged. The resulting Lewis structure is shown in Figure 7.5. Experimental results confirm that the acidity of BrU is increased with a $\mathrm{p}K_\mathrm{a}$ of 7.84 ± 0.03 compared to 9.71 for U. [236] However, theoretical gas phase calculations suggest that the pairs U^--G and BrU^--G prefer a twisted conformation with propeller twist and buckle angles of about 43°and 5–13°. [237] Only two instead of three hydrogen bonds can be formed and the two unpaired donating sites are repelling each other. This makes the planar conformation — which would be enforced in the DNA helix — energetically even less favourable.

The acidity of U and BrU has been investigated with PMC simulations. Hydrogen bonds are of large importance for the structures and an explicit description of the solvent is necessary. However, the charge of the deprotonated species is especially challenging for QM/MM approaches with force fields that do not include any explicit polarisation terms. Therefore, this poses an interesting challenge for our implementation of the PMC approach. The $\mathrm{p}K_\mathrm{a}$ values can be derived from the free solvation energies of the involved species considering the thermodynamic cycle shown in Figure 7.6. For U — and on the same way for BrU — the free dissociation energy in solution $\Delta G^\mathrm{aq}_\mathrm{diss}(\mathrm{U})$ can be computed from the free solvation energies ΔG_solv and the dissociation in gas phase

$$\Delta G^\mathrm{aq}_\mathrm{diss}(\mathrm{U}) = -\Delta G_\mathrm{solv}(\mathrm{U}) + \Delta G^\mathrm{gp}_\mathrm{diss}(\mathrm{U}) + \Delta G_\mathrm{solv}(\mathrm{U}^-) + \Delta G_\mathrm{solv}(\mathrm{H}^+). \tag{7.1}$$

This allows the direct computation of the $\mathrm{p}K_\mathrm{a}$ values

$$\begin{tikzcd}
{[\mathrm{U}]} \arrow[r, "\Delta G^{\mathrm{aq}}_{\mathrm{diss}}(\mathrm{U})"] \arrow[d, "-\Delta G_{\mathrm{solv}}(\mathrm{U})"'] & {[\mathrm{U}^-] + [\mathrm{H}^+]} \\
\mathrm{U} \arrow[r, "\Delta G^{\mathrm{gp}}_{\mathrm{diss}}(\mathrm{U})"'] & \mathrm{U}^- + \mathrm{H}^+ \arrow[u, "\Delta G_{\mathrm{solv}}(\mathrm{U}^-) + \Delta G_{\mathrm{solv}}(\mathrm{H}^+)"']
\end{tikzcd}$$

FIGURE 7.6 Thermodynamic cycle with dissociation and solvation energies. Square brackets denote a system in aqueous solution.

$$\mathrm{p}K_{\mathrm{a}} = \Delta G^{aq}_{\mathrm{diss}}/(\ln(10)RT). \tag{7.2}$$

However, the computation of the free solvation energy of a proton is challenging and the associated uncertainty may be so large that even qualitative conclusions cannot be drawn. Using an experimental value of -1105 kJ/mol [238] leads to the values of -2.1 and -7.4 for U and BrU, respectively. This allows qualitative conclusions and shows that the acidity is clearly increased for the brominated species. However, this difficulty can be avoided altogether by computing the change of the $\Delta\mathrm{p}K_{\mathrm{a}}$ upon going from U to BrU which is given by

$$\begin{aligned}
\Delta G^{\mathrm{aq}}_{\mathrm{diss}}(\mathrm{BrU}) - \Delta G^{\mathrm{aq}}_{\mathrm{diss}}(\mathrm{U}) = & -\Delta G_{\mathrm{solv}}(\mathrm{BrU}) + \Delta G_{\mathrm{solv}}(\mathrm{U}) \\
& + \Delta G^{\mathrm{gp}}_{\mathrm{diss}}(\mathrm{BrU}) - \Delta G^{\mathrm{gp}}_{\mathrm{diss}}(\mathrm{U}) \\
& + \Delta G_{\mathrm{solv}}(\mathrm{BrU}^-) - \Delta G_{\mathrm{solv}}(\mathrm{U}^-).
\end{aligned} \tag{7.3}$$

The free energy term $\Delta G_{\mathrm{solv}}(\mathrm{H}^+)$ does not have to be computed at all with this approach and no experimental values are required. The obtained $\Delta\mathrm{p}K_{\mathrm{a}}$ is -5.3 which is about two times larger than the experimental value of -2. The results are in qualitative agreement but the acidity is overestimated with the QM/MM approach. This may be caused by the description of the solvent with fixed partial charges. Especially for the charged species U^- and BrU^- polarisation effects will become more relevant and they are insufficiently captured by the average polarisation of the effective potential. Furthermore, one can only partly benefit from error cancellation because the error is larger in the charged species. The deviation in the $\Delta\mathrm{p}K_{\mathrm{a}}$ value from the experimental value is equivalent to an overestimation of the dissociation energy of 4.5 kcal/mol.

The base pairing between the ionised forms and G has been revisited under consideration of solvent effects with COSMO. It has been confirmed that the twisted confirmation is still the minimum. The propeller-twist and the buckle angles are $-41°$ and $14°$ respectively with the definition of the angles given in Figure 7.7. For BrU^--G the angles are $-48.2°$ and $16°$. This agrees qualitatively with the gas phase studies, although the propeller-twist angle is slightly larger for BrU and the buckle angle is larger for both bases by about $10°$.

A relaxed surface scan of the dihedral angle shown in Figure 7.7 reveals that the barrier to the planar transition state is only about 1.2 kcal/mol for both U^--G and BrU^--G. The bromine atom does not influence the potential. The base pairs have to be forced into a planar conformation in a DNA helix but the energetic penalty is very low. Furthermore, experimentally propeller-twist angles of about $25°$ are known for A-T base pairs [239] and for example the protein L7Ae [240]

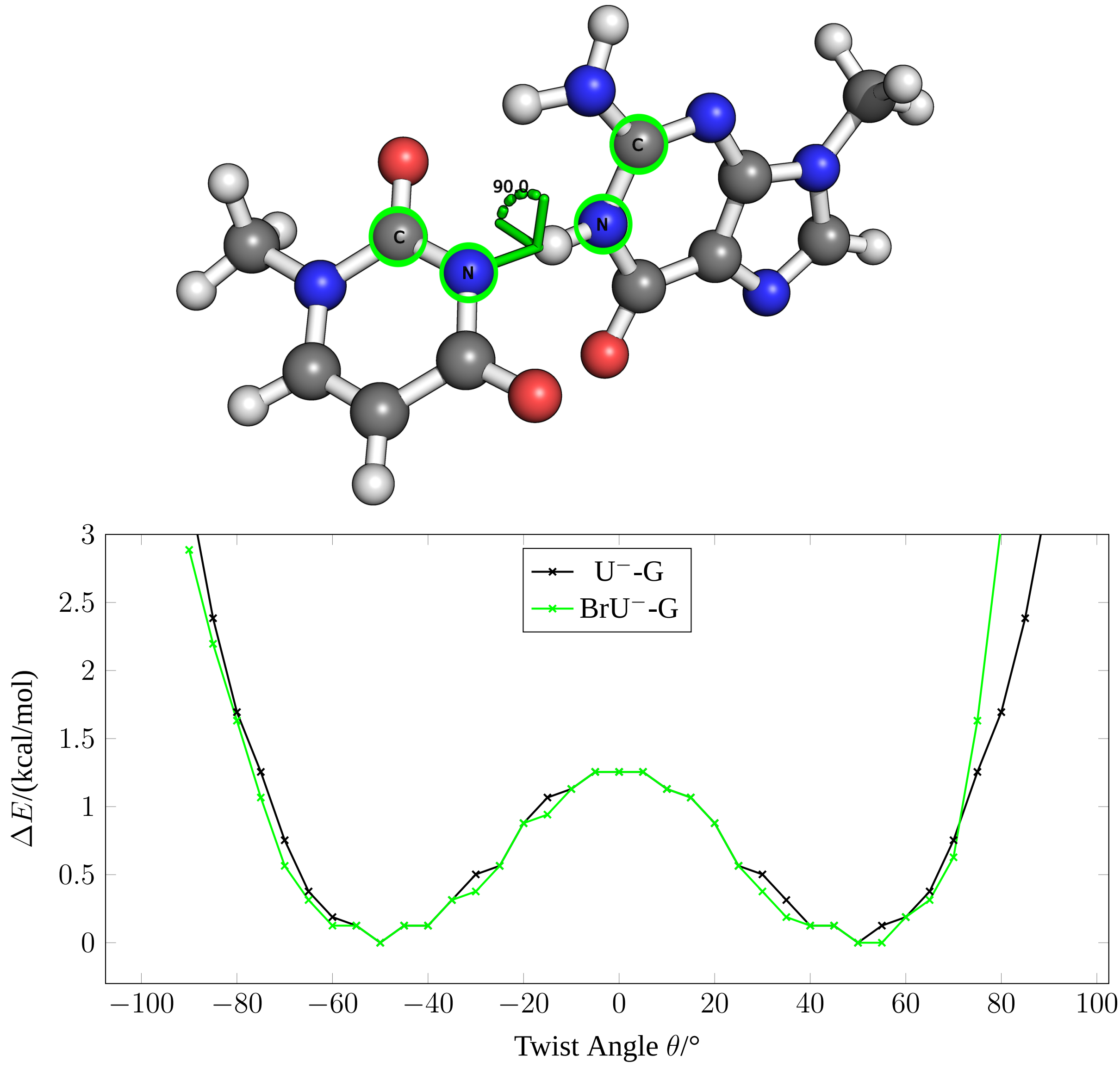


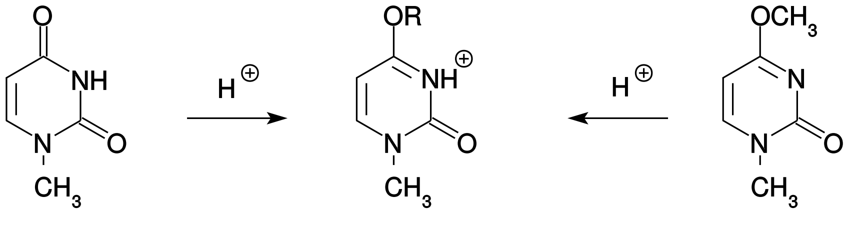
FIGURE 7.7 Top: Definition of the dihedral angle with the involved atoms marked in green. Bottom: Potential energy computed with B3LYP-D3/def2-SVP and COSMO ($\epsilon = 80.0$) as a function of the dihedral angle for the U$^-$-G and BrU$^-$-G base pairs.

includes the non-canonical base pair A-G with a buckle angle of about $33°$. This suggest that the occurrence of the (Br)U$^-$-G pair might be sensitive to the DNA context. Suitable neighbouring base pairs could alleviate the energetic penalty of the large propeller-twist and buckle angles.

7.3 Tautomerisation Model

The tautomerisation model assumes that a non-negligible amount of the keto-enol form of U (Uol, see Figure 7.5) is present in the equilibrium between the diketon U and Uol. The keto-enol form matches very well with G forming all three hydrogen bonds and is therefore the ideal analogue to C.

An exact experimental determination of the tautomerisation equilibrium constants K_{taut} is difficult because one of the tautomers dominates greatly. The method of choice is in that case the basicity method which determines the ratio of the acid constants of species (1) and (3) as



(1) (2) (3)

FIGURE 7.8 Schematic representation of the experimental approach to determine the ration of the tautomers of uracil.

shown in Figure 7.8. Assuming that the influence of the methyl groups is negligible the ratio of the tautomers of U is approximately equal to the ratio of the acid constants. For weakly basic amines one can define the concentration of the free base A and the conjugate acid HA^+ as a function of the Hammett H_0 constant

$$\log \frac{[\text{A}]}{[\text{HA}]^+} = H_0 + \text{constant} \tag{7.4}$$

which leads to values for K_taut of $1.0 \cdot 10^{-4}$ and $5.0 \cdot 10^{-4}$ for U and BrU respectively. Another approach is to plot $\log \frac{[\text{A}]}{[\text{HA}]^+}$ against H_0 — which should be close to linear — and choose an H_0 in the middle between the two pK_a values. This gives the values $5.0 \cdot 10^{-4}$ and $2.0 \cdot 10^{-2}$ for U and BrU respectively. [236] These analyses agree that the amount of keto-enol is increased for the brominated species even if still very low.

Most theoretical results agree with the finding that the diketo form is energetically much more stable. However, the difference between U and BrU is very much under debate. Gas phase and PCM computations show no difference at all, [231, 232] classical simulations hint at an increased tautomerisation of U [234] and QM cluster calculations with 50 and 100 water molecules show a very strong preference for the keto-enol form in BrU. [229,235] Microsolvation studies with 1–2 water molecules reveal a sensitivity of the equilibrium to the position of the water molecule. [231] Furthermore, the presence of sodium ions which are relevant in biological environments might have a significant influence on the equilibrium. [230] Unfortunately, we did not investigate this aspect in this work.

It becomes clear that the description of solvent effects is crucial for this step. Microsolvation studies are strongly biased because the position of the water molecules is chosen manually. Continuum models lack the explicit solvent structure description, which is especially crucial for hydrogen bonds, and entropic effects are not very well described. QM cluster models are the most accurate approach in describing the electronic structure but miss completely any sampling of the configuration space while classical simulations are at the other end of the spectrum with sufficient sampling but empirical potential energy functions. An investigation with PMC simulations of the tautomerisation model combines the sampling at a rather high level of theory with an explicit description of solvent molecules.

The tautomerisation in aqueous solution has been studied by PMC simulations. A thermodynamic cycle (Figure 7.9) — similar to the cycle shown in the ionisation model — has been

$$\begin{tikzcd}
{[\mathrm{U}]} \arrow[r, "\Delta G^{\mathrm{aq}}_{\mathrm{taut}}"] \arrow[d, "-\Delta G_{\mathrm{solv}}(\mathrm{U})"'] & {[\mathrm{Uol}]} \\
\mathrm{U} \arrow[r, "\Delta G^{\mathrm{gp}}_{\mathrm{taut}}"'] & \mathrm{Uol} \arrow[u, "\Delta G_{\mathrm{solv}}(\mathrm{Uol})"']
\end{tikzcd}$$

FIGURE 7.9 Thermodynamic cycle with tautomerisation and solvation free energies. Square brackets denote a system in aqueous solution.

used leading to the computation of free solvation energies as well as the tautomerisation in gas phase. The determining factor for the reactivity in solution is the difference between the free solvation energies of reactant U and product Uol. The results are summarised in comparison with theoretical studies from the literature in Table 7.1.

The gas phase results show that the free tautomerisation energies are rather insensitive to the QM description. HF in comparison with QCISD(T) shows an error of about 3 kcal/mol which can be attributed to correlation. However, even HF reproduces correctly the energetic ordering – the difference between U and BrU. This ordering is consistent for all methods employed and is about 0.5–0.7 kcal/mol except for QCISD with only 0.3 kcal/mol. All these calculations underestimate the difference as confirmed by LCCSD(T) [241] calculations with basis set extrapolation to the complete basis set limit from aug-cc-pVTZ and aug-cc-pVQZ basis sets. [163,242] This results in a difference of about 0.9 kcal/mol. The deviation of the QCISD results can be understood in terms of the remaining finite basis set effect and the geometries that have been only optimised with HF. The DFT approach with the functional B3LYP and D3 dispersion corrections used in this work is a good compromise between accuracy and computational costs for the simulations. The functional gives an accurate description of the electron density which is decisive for the interactions between the solute and solvent in the simulations. The results from the PMC simulations can be combined with the coupled cluster gas phase results which show overall the smallest deviation from the experimental results for the absolute values. However, the preference of BrU over U is little influenced by the gas phase results and solvent effects dominate the difference between U and BrU.

The determining difference between the studies in solution is the treatment of the solvent effects. In this case the experimental results, which predict a stabilisation of about 1–2 kcal/mol of BrUol compared to Uol, are the reference for any theoretical study. The continuum model stabilises slightly the keto-enol form but it does not distinguish between U and BrU emphasizing again that the explicit description of the solvent is necessary. The force field-based simulations predict even a destabilisation of the brominated species which shows on the other hand that the QM description of the solute is fundamental in this particular case to obtain accurate free solvation energies. This is in agreement with the findings of the hydrogen bonding in the case of methylchloride in Section 5.2 where it has been found that the classical simulations cannot reproduce the hydrogen bond between the halogen and water. Therefore, only the PMC and QM cluster model computations predict indeed an energetic order in agreement with the experiment.

TABLE 7.1 Free tautomerisation energy in gas phase and solution kcal/mol and $K_{\text{taut}} = \frac{[\text{Enol}]}{[\text{Ketone}]}$. The gas phase calculations from the work of Luque and coworkers [234] used the 6-311+G(d,p) basis set and HF/6-311+G(d,p) optimized geometries. For the PMC (B3LYP-D3/def2-SVP and OPLS-AA vdW parameters) simulations the level of theory for the gas phase reaction is specified in the Table.

		ΔG_{taut}		K_{taut}	
		U	BrU	U	BrU
Gas Phase	B3LYP-D3	12.7	13.2	$5.0 \cdot 10^{-10}$	$2.1 \cdot 10^{-10}$
	HF [234]	13.7	14.3	$9.0 \cdot 10^{-11}$	$3.3 \cdot 10^{-11}$
	MP2 [234]	10.5	11.2	$2.0 \cdot 10^{-8}$	$6.1 \cdot 10^{-9}$
	MP4(SDTQ) [234]	11.2	11.8	$6.1 \cdot 10^{-9}$	$2.2 \cdot 10^{-9}$
	QCISD [234]	11.5	11.8	$3.7 \cdot 10^{-9}$	$2.2 \cdot 10^{-9}$
	QCISD(T) [234]	10.9	–	$1.0 \cdot 10^{-8}$	–
	LCCSD(T)/CBS[3,4]	9.6	10.5	$9.0 \cdot 10^{-8}$	$2.1 \cdot 10^{-8}$
Solution	COSMO	11.8	11.8	$2.3 \cdot 10^{-9}$	$2.1 \cdot 10^{-9}$
	PMC + B3LYP-D3	10.9	9.2	$9.9 \cdot 10^{-9}$	$9.2 \cdot 10^{-7}$
	PMC + LCCSD(T)/CBS[3,4]	7.8	6.5	$1.8 \cdot 10^{-6}$	$1.8 \cdot 10^{-5}$
	MM MC [234]	9.8	11.0	$6.5 \cdot 10^{-8}$	$8.6 \cdot 10^{-9}$
	Cluster Model [229]	9.3	-4.5	$1.5 \cdot 10^{-7}$	$1.9 \cdot 10^{+3}$
	Experiment [236]	5.5	4.5	$1.0 \cdot 10^{-4}$	$5.0 \cdot 10^{-4}$
		4.5	2.3	$5.0 \cdot 10^{-4}$	$2.0 \cdot 10^{-2}$

However, the cluster model results overestimate strongly the stability of BrUol by predicting even the keto-enol form as preferred over the diketo form. While the cluster model is the most accurate approach to compute the interactions it uses a single optimised geometry. This neglects temperature and ensemble effects and the results cannot be directly compared to experimental values. Interestingly, it shows that for a certain solvent configuration the equilibrium is shifted strongly towards the enol. However, sufficient sampling, explicit solvent and a reasonable high-level QM description have to be combined to obtain quantitative results. Nevertheless, the computation of absolute equilibrium constants remains challenging and the error is about two to three orders of magnitude. Due to the constants depending exponentially on the free energy even the smallest deviations have a large influence on the final result. Nevertheless, the PMC simulations combined with the best gas phase results predict indeed a stabilisation of 1.3 kcal/mol which is in good agreement with the experimental findings.

7.4 Wobble Model

Opposed to the above presented models, the wobble model proposes not a direct transformation of U but rather a different binding mode of U with G (see Figure 7.5). If U just changes slightly its geometry inside the double helix two of the hydrogen bond sites can pair with G and form a mismatching pair. Experimentally these pairs have been observed in RNA molecules. [243, 244] In DNA a pH dependent equilibrium between the wobble pair and the

TABLE 7.2 Interaction energies computed with B3LYP-D3/def2-SVP and COSMO ($\epsilon = 80$) in kcal/mol.

Dimer	H	Br
U-G(Wobble)	-13.1	-13.3
Uol-G	-23.4	-23.3
U^--G(twisted)	-17.1	-16.1
U^--G(planar)	-15.9	-15.0
C-G	-20.8	

ionic pair has been found for BrU as well as the fluorinated species. [245, 246] However, it has been suggested that the possibility to wobble is intrinsic to U and does not depend on the halogen. [235] This is in agreement with the experimental findings.

Interaction energies have been computed for the wobble pair and the pairs of the ionisation and tautomerisation model as well as the canonical C-G pair to establish a reference. Comparing the interaction energies of the different models the Uol-G pair shows the largest interaction energy, even larger than the one of the canonical pair C-G by about 2.5 kcal/mol. The other mismatched pairs show lower interaction energies up to 8 kcal/mol. The wobble model with the mismatched U-G pair has as expected the weakest interaction. However, all of them form stable dimers compared to the separated monomers. Furthermore, if Uol is present it will form a stable dimer with G and will not be replaced by the canonical base C. The overall shape of the dimer has been experimentally excluded as the decisive factor. Different sites have been identified in DNA polymerases which recognise mismatched pairs. However, these do not recognise the mismatch by the shape but through specific interactions. [225]

A comparison between U and BrU reveals that the strength of the base pair interaction does not depend on the bromine atom in 5-position. This excludes the wobble model as a possible pathway leading to mutations. Interestingly, also the other models show little difference between U and BrU. This strongly suggests a separate activation step before the actual base pairing as in the ionisation and tautomerisation model.

Finally, it has been found that U and BrU behave differently in the triplet state and that this might be the reason for the mutagenic properties. [117] However, the mutation reaction requires no light or radiation and the involvement of excited states is rather unlikely. [235]

7.5 Deactivation Pathways

In the theoretical study by Hu and coworkers [231] a deactivation mechanism has been suggested which might explain the different mutagenic properties of U and BrU. It has been found that the mismatching pair G-Uol may react to Gol-U by a simple proton transfer reaction as shown in Figure 7.10. This reaction followed by another DNA replication step would in the end result only in the original base pair A-T. This pathway is represented schematically in Figure 7.11. It had been found that for both U and the bromated species the reaction is without barrier. However, for U the free reaction energy is -5 kcal/mol while it is about 0 kcal/mol for BrU. This means that the equilibrium is strongly shifted towards the deactivated species for U

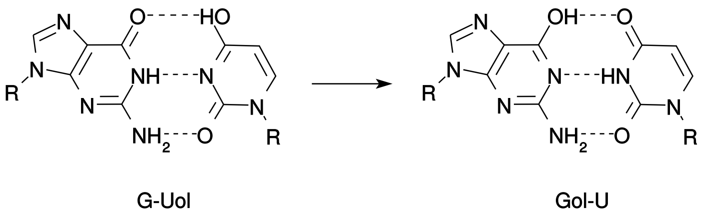


FIGURE 7.10 Lewis structure of the deactivation step from G-Uol to Gol-U.

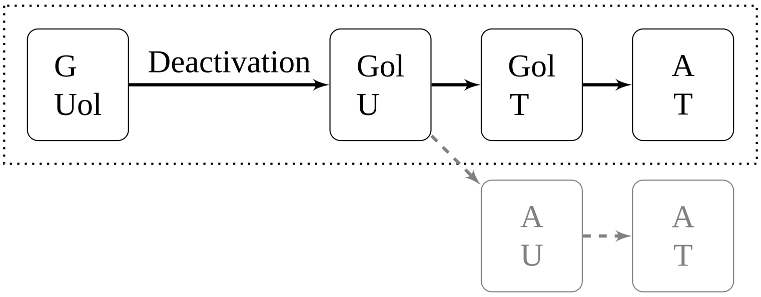


FIGURE 7.11 Schematic representation of the proposed deactivation mechanism. Each branching represents a DNA replication.

but that the bromated species follows only by about 50% the deactivation pathway. Nevertheless, the theoretical studies had been carried out only at the HF/STO-3G level of theory in gas phase so that another look at these reactions is warranted.

Two-dimensional surface scans have been carried out with B3LYP-D3/def2-SVP [17, 20–22, 111–114] with the CPCM model ($\epsilon = 80$). The coordinates of the scan denote the O-H bond of the Uol that is broken and the N-H bond of G which is formed. The resulting contour plot is shown in Figure 7.12 with the reaction proceeding from the bottom right corner to the top left corner following the minimum energy path. Looking at the free reaction energies they are about 0 kcal/mol for both species. This stands in contrast to the previous results which is due to the higher accuracy of the QM method. However, it also does not show any preference for the deactivation of either species. On the other hand, the reaction barriers are about 6 kcal/mol and 5 kcal/mol for Uol-G and BrUol-G, respectively. This shows a slight preference for the deactivation of BrUol-G assuming the reaction is kinetically controlled. Furthermore, one can notice that the reaction of the BrUol-G system proceeds through a very shallow intermediate state at about d(NH) = 1.63 Å and d(OH) = 1.39 Å which is not present for Uol-G. This means that BrUol-G is in equilibrium with this energetically very close intermediate and consequently is more likely to follow from the deactivation pathway over the lower and less wide barrier.

Our findings of the deactivation pathway under consideration of correlation effects by DFT with the B3LYP functional and D3 dispersion corrections hint at the possibility of a preferred deactivation for the bromated species. However, our findings suggest a kinetic control of the reaction rather then a strong thermodynamic preference. Therefore, these findings hint at another factor which contributes to the difference in the mutagenic properties of U and BrU.

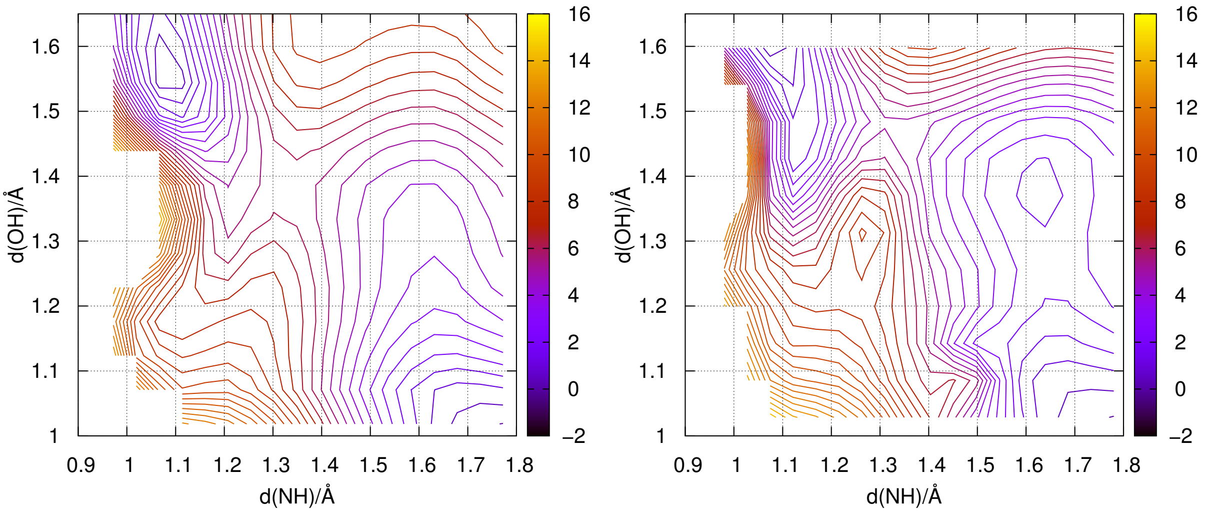


FIGURE 7.12 The potential energy surface computed with B3LYP-D3/def2-SVP and COSMO ($\epsilon = 80$) for the proton transfer between Uol and G (left) and BrUol and G (right). The colour scale is in kcal/mol.

CHAPTER 8

Summary and Future Work

In this thesis a new method for simulations in condensed phase has been brought forward. The PMC is a hybrid QM/MM approach describing the system of interest with the high accuracy of QM methods and the environment with an explicit representation and the low computational cost of MM methods. The key step has been to apply first order perturbation theory to reduce the costs of the energy calculations during the MC simulations. This allows exploring the configuration space at the QM/MM level and goes thereby beyond the accuracy of commonly used sequential QM/MM approaches. It has been demonstrated that PMC allows to study chemistry in solution and aspects have been identified where the description of the physics by this approach is not sufficient. However, chemistry in solution is an active field of research in the computational chemistry community and the PMC method is a good starting point for further developments. Some of these possibilities will be considered in the following text.

The choice of the reference wave function for the perturbation theory makes our approach more efficient than alternative developments. [87, 88] The convergence of the wave function with embedding allows to restrict the perturbation to first order because the perturbation due to a MC move can be treated accurately. Further work in this direction should explore alternative reference wave functions. It might be advantageous to use a reference from a continuum calculation which takes the solvent through the average description of the continuum into account. Therefore, it might be possible to avoid the updates of the wave function altogether because the wave function is already polarised for solution but not biased towards a certain solvent configuration. Another way would be to construct the ASEP on the fly and use this to compute the wave function. This goes one step further than the continuum model because the explicit solvent representation is considered.

Furthermore, a specialised procedure for simulations under PBC conditions has been presented. Avoiding the computational costs and problems of the Ewald summation, the direct summation with a shifted force operator has been proposed. This requires the numerical integration of the electrostatic interaction between the QM and MM system. However, a very efficient implementation utilising the computing power of GPUs has been developed which benefits from the simple energy expressions and the high data parallelism. There remains space for improvement because of inefficiencies on the GPU due to branch divergence. This occurs because of the cut-off of the long-range interactions. Particles that are within the cut-off and particles that are too distant from each other are computed at the same time. Using a grid ordered according to the coordinates would already reduce this problem. Additionally, separating the grid into cells based on a neighbour list as they are commonly used in large scale simulations would allow to reduce the computational costs as well. [247]

The parameters of the proposed approach have been thoroughly tested and guidelines have been devised for future applications. The number of updates allows to balance between accuracy and computational costs. These show a dependence on the solute and the solvent and can be drastically reduced or increased as required. For example, in apolar solvents only few updates are required while in water with its strong hydrogen bonding interactions more updates have to be carried out. Further developments can replace this static and empirical criteria with a dynamic approach in order to reduce the computational costs without sacrificing accuracy. A first way would be a simple geometric or distance based criteria. Another possibility would be to evaluate the deviation from the exact energy on the fly and increase or decrease the number of updates as required.

Furthermore, the accuracy of the numerical integration has been evaluated which is used to compute the electrostatic interaction between the QM and MM system in the perturbation steps. It has been found that the integration can be well converged towards the exact solution by increasing the number of grid points. The large amount of grid points does not slow down the simulations significantly because the architecture of graphics cards is well suited for this purpose.

Benchmark calculations computing free solvation energies have been presented. These confirmed that the physics of solute-solvent interactions is well captured by our PMC method. However, the results also raised concerns regarding the compatibility of classical force fields with QM methods. Especially the simulations in chloroform showed a strong dependence on the force field. This issue has been confirmed when studying the influence of different classical water models on the solvent structure. Improving a force field for purely classical simulations is not necessarily accompanied by an improvement in QM/MM calculations. The TIP5P water model should clearly be avoided in this context because of the charge sites without repulsive vdW interactions that become embedded in the electronic density. Furthermore, the correct definition of vdW parameters during reactions is not clear. While the overall change of free energy is independent of this choice, the shape of the reaction barrier depends clearly on it. Conventionally, a simple linear interpolation is used for reactions. This shows that force fields have to be developed with the explicit focus on the use in QM/MM schemes.

The computation of electronic spectra in solution is a very challenging task. However, it has been shown that the PMC approach in combination with TD-DFT can give valuable insight into the physics of solvent shifts. However, given certain types of excitations (e.g. charge transfer) the current approach shows severe restrictions. On one hand, the description of the solvent without considering more than the average effect of the polarisation can lead to large errors in absolute excitation energies as well as solvent shifts. On the other hand, TD-DFT lacks the required accuracy for charge transfer excitations in the context of embedding methods. To improve the former different ways can be pursued in future work to account for polarisation based developments allowing for a rapid evaluation of polarisation effects in MC simulations [248–250] or based on generalised polarisabilities. [88] To overcome the limits of TD-DFT more accurate methods may be used in future applications which is facilitated by the reduced number of QM calculations needed in the PMC approach.

Additionally, it has been shown recently that especially for emission spectra the effect of the geometry optimisation is very important. [251] While quantum MC methods emerge with recent advances in the evaluation of forces as the most accurate method for the optimisation of excited state geometries even for larger systems, [252,253] solvent effects have been considered mostly with continuum models in this context. Geometry optimisations in PMC could be carried out in different ways. First, the ASEP can be evaluated and the geometry can be optimised in the usual self-consistent way. Second, the averages of the forces can be evaluated directly from the simulations followed by an optimisation step. The influence of these approaches on the geometries has to be assessed in future work. In a completely different approach one would simply sample the degrees of freedom of the QM system as well and avoid the problem of the rigid geometry altogether. This would give access to vibrational properties of the solute under consideration of solvent effects.

The application of the PMC approach to binary solvent mixtures showed the potential of improved MC moves. Also for pure solvents more efficient MC moves can be devised. [254–256] Furthermore, not only the generation of the configurations can be improved but also the formulation of the observables itself. The zero-variance principle allows deriving renormalised observables with greatly reduced variance. [257] This means that less MC steps are required to obtain the same statistical accuracy.

Finally, the application of the PMC approach in order to understand the processes that may damage the DNA showed the value of this newly developed approach. Especially in biochemistry solvent effects are ubiquitous and efficient and accurate methods need to be developed in order to deepen our understanding of these complex systems. This thesis contributes towards this and will certainly help in future applications to understand the complex interplay of interactions in solution.

Bibliography

[1] M. Orozco and F. J. Luque, “Theoretical Methods for the Description of the Solvent Effect in Biomolecular Systems,” *Chemical Reviews*, vol. 100, pp. 4187–4226, Nov. 2000.

[2] D. Voet, J. G. Voet, and C. W. Pratt, *Fundamentals of Biochemistry: Life at the Molecular Level*. Hoboken, NJ: Wiley, 4 edition ed., Jan. 2012.

[3] C. N. Matthews, “The origin of proteins: Heteropolypeptides from hydrogen cyanide and water,” *Origins of Life*, vol. 6, pp. 155–162, Apr. 1975.

[4] L. D. Barron, L. Hecht, and G. Wilson, “The Lubricant of Life: A Proposal That Solvent Water Promotes Extremely Fast Conformational Fluctuations in Mobile Heteropolypeptide Structure,” *Biochemistry*, vol. 36, pp. 13143–13147, Oct. 1997.

[5] https://www.marketsandmarkets.com/Market-Reports/solvent-market 1325.html, “Solvent Market by Type, Application, Source & by Geography - 2021 | MarketsandMarkets,” 07.02.2018.

[6] C. Jimenez-Gonzalez, C. S. Ponder, Q. B. Broxterman, and J. B. Manley, “Using the Right Green Yardstick: Why Process Mass Intensity Is Used in the Pharmaceutical Industry To Drive More Sustainable Processes,” *Organic Process Research & Development*, vol. 15, pp. 912–917, July 2011.

[7] T. Welton, “Solvents and sustainable chemistry,” *Proceedings. Mathematical, Physical, and Engineering Sciences / The Royal Society*, vol. 471, Nov. 2015.

[8] K. Griebenow, M. Vidal, C. Baéz, A. M. Santos, and G. Barletta, “Nativelike Enzyme Properties Are Important for Optimum Activity in Neat Organic Solvents,” *Journal of the American Chemical Society*, vol. 123, pp. 5380–5381, June 2001.

[9] A. Warshel, “Electrostatic basis of structure-function correlation in proteins,” *Accounts of Chemical Research*, vol. 14, pp. 284–290, Sept. 1981.

[10] A. Warshel, “Energetics of enzyme catalysis.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 75, pp. 5250–5254, Nov. 1978.

[11] https://www.solvation.de/, “RESOLV - Cluster of Excellence - EXC 1069,” 12.02.2018.

[12] F. Jensen, *Introduction to Computational Chemistry*. Chichester, England ; Hoboken, NJ: Wiley, 2 edition ed., 2007.

[13] P. Hohenberg and W. Kohn, “Inhomogeneous Electron Gas,” *Physical Review*, vol. 136, pp. B864–B871, Nov. 1964.

[14] F. Bloch, “Bemerkung zur Elektronentheorie des Ferromagnetismus und der elektrischen Leitfähigkeit,” *Zeitschrift für Physik*, vol. 57, pp. 545–555, July 1929.

[15] P. a. M. Dirac, “Note on Exchange Phenomena in the Thomas Atom,” *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 26, pp. 376–385, July 1930.

[16] W. Kohn and L. J. Sham, “Self-Consistent Equations Including Exchange and Correlation Effects,” *Physical Review*, vol. 140, pp. A1133–A1138, Nov. 1965.

[17] S. H. Vosko, L. Wilk, and M. Nusair, “Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis,” *Canadian Journal of Physics*, vol. 58, pp. 1200–1211, Aug. 1980.

[18] A. D. Becke, “Density-functional exchange-energy approximation with correct asymptotic behavior,” *Physical Review A*, vol. 38, pp. 3098–3100, Sept. 1988.

[19] B. Miehlich, A. Savin, H. Stoll, and H. Preuss, “Results obtained with the correlation energy density functionals of becke and Lee, Yang and Parr,” *Chemical Physics Letters*, vol. 157, pp. 200–206, May 1989.

[20] C. Lee, W. Yang, and R. G. Parr, “Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density,” *Physical Review B*, vol. 37, pp. 785–789, Jan. 1988.

[21] A. D. Becke, “Density functional thermochemistry. III. The role of exact exchange,” *The Journal of Chemical Physics*, vol. 98, pp. 5648–5652, Apr. 1993.

[22] 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,” *The Journal of Physical Chemistry*, vol. 98, pp. 11623–11627, Nov. 1994.

[23] A. D. Becke, “A multicenter numerical integration scheme for polyatomic molecules,” *The Journal of Chemical Physics*, vol. 88, pp. 2547–2553, Feb. 1988.

[24] G. Voronoi, “Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Deuxième mémoire. Recherches sur les parallélloèdres primitifs.,” *Journal für die reine und angewandte Mathematik (Crelle’s Journal)*, vol. 1908, no. 134, pp. 198–287, 1908.

[25] E. Wigner and F. Seitz, “On the Constitution of Metallic Sodium,” *Physical Review*, vol. 43, pp. 804–810, May 1933.

[26] V. Lebedev, “Values of the nodes and weights of ninth to seventeenth order gauss-markov quadrature formulae invariant under the octahedron group with inversion,” *USSR Computational Mathematics and Mathematical Physics*, vol. 15, pp. 44–51, Jan. 1975.

[27] M. E. Mura and P. J. Knowles, “Improved radial grids for quadrature in molecular density functional calculations,” *The Journal of Chemical Physics*, vol. 104, pp. 9848–9858, June 1996.

[28] M. P. Allen and D. J. Tildesley, *Computer simulation of liquids*. Oxford [England]; New York: Clarendon Press ; Oxford University Press, 1989.

[29] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, “Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids,” *Journal of the American Chemical Society*, vol. 118, pp. 11225–11236, Jan. 1996.

[30] W. L. Jorgensen and J. Tirado-Rives, “The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin,” *Journal of the American Chemical Society*, vol. 110, pp. 1657–1666, Mar. 1988.

[31] E. Harder, W. Damm, J. Maple, C. Wu, M. Reboul, J. Y. Xiang, L. Wang, D. Lupyan, M. K. Dahlgren, J. L. Knight, J. W. Kaus, D. S. Cerutti, G. Krilov, W. L. Jorgensen, R. Abel, and R. A. Friesner, “OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins,” *Journal of Chemical Theory and Computation*, vol. 12, pp. 281–296, Jan. 2016.

[32] P. P. Ewald, “Die Berechnung optischer und elektrostatischer Gitterpotentiale,” *Annalen der Physik*, vol. 369, pp. 253–287, Jan. 1921.

[33] J. S. Hub, B. L. de Groot, H. Grubmüller, and G. Groenhof, “Quantifying Artifacts in Ewald Simulations of Inhomogeneous Systems with a Net Charge,” *Journal of Chemical Theory and Computation*, vol. 10, pp. 381–390, Jan. 2014.

[34] F. Pullara and I. J. General, “Population reversal driven by unrestrained interactions in molecular dynamics simulations: A dialanine model,” *AIP Advances*, vol. 5, p. 107235, Oct. 2015.

[35] W. Weber, P. H. Hünenberger, and J. A. McCammon, “Molecular Dynamics Simulations of a Polyalanine Octapeptide under Ewald Boundary Conditions: Influence of Artificial Periodicity on Peptide Conformation,” *The Journal of Physical Chemistry B*, vol. 104, pp. 3668–3675, Apr. 2000.

[36] P. H. Hünenberger and J. A. McCammon, “Effect of artificial periodicity in simulations of biomolecules under Ewald boundary conditions: a continuum electrostatics study,” *Biophysical Chemistry*, vol. 78, pp. 69–88, Apr. 1999.

[37] D. Wolf, P. Keblinski, S. R. Phillpot, and J. Eggebrecht, “Exact method for the simulation of Coulombic systems by spherically truncated, pairwise r^{-1} summation,” *The Journal of Chemical Physics*, vol. 110, pp. 8254–8282, Apr. 1999.

[38] C. J. Fennell and J. D. Gezelter, “Is the Ewald summation still necessary? Pairwise alternatives to the accepted standard for long-range electrostatics,” *The Journal of Chemical Physics*, vol. 124, no. 23, p. 234104, 2006.

[39] O. Acevedo and W. L. Jorgensen, “Quantum and molecular mechanical Monte Carlo techniques for modeling condensed-phase reactions,” *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 4, pp. 422–435, Sept. 2014.

[40] B. W. McCann and O. Acevedo, “Pairwise Alternatives to Ewald Summation for Calculating Long-Range Electrostatics in Ionic Liquids,” *Journal of Chemical Theory and Computation*, vol. 9, pp. 944–950, Feb. 2013.

[41] A. Klamt and G. Schüürmann, “COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient,” *Journal of the Chemical Society, Perkin Transactions 2*, pp. 799–805, Jan. 1993.

[42] S. Miertuš, E. Scrocco, and J. Tomasi, “Electrostatic interaction of a solute with a continuum. A direct utilizaion of AB initio molecular potentials for the prevision of solvent effects,” *Chemical Physics*, vol. 55, pp. 117–129, Feb. 1981.

[43] A. V. Marenich, C. J. Cramer, and D. G. Truhlar, “Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions,” *The Journal of Physical Chemistry B*, vol. 113, pp. 6378–6396, May 2009.

[44] G. Alagona, C. Ghio, R. Cammi, and J. Tomasi, “A Reappraisal of the Hydrogen Bonding Interaction Obtained by Combining Energy Decomposition Analyses and Counterpoise Corrections,” in *Molecules in Physics, Chemistry, and Biology*, Topics in Molecular Organization and Engineering, pp. 507–559, Springer, Dordrecht, 1988.

[45] J. Tomasi, “Thirty years of continuum solvation chemistry: a review, and prospects for the near future,” *Theoretical Chemistry Accounts*, vol. 112, pp. 184–203, Sept. 2004.

[46] W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson, “Semianalytical treatment of solvation for molecular mechanics and dynamics,” *Journal of the American Chemical Society*, vol. 112, pp. 6127–6129, Aug. 1990.

[47] L. Onsager, “Electric Moments of Molecules in Liquids,” *Journal of the American Chemical Society*, vol. 58, pp. 1486–1493, Aug. 1936.

[48] J. Tomasi, B. Mennucci, and R. Cammi, "Quantum Mechanical Continuum Solvation Models," *Chemical Reviews*, vol. 105, pp. 2999–3094, Aug. 2005.

[49] A. Klamt, V. Jonas, T. Bürger, and J. C. W. Lohrenz, "Refinement and Parametrization of COSMO-RS," *The Journal of Physical Chemistry A*, vol. 102, pp. 5074–5085, June 1998.

[50] A. Klamt, "Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena," *The Journal of Physical Chemistry*, vol. 99, pp. 2224–2235, Feb. 1995.

[51] L. S. Ornstein and F. Zernike, "Accidental deviations of density and opalescence at the critical point of a single substance," *Koninklijke Nederlandse Akademie van Wetenschappen Proceedings Series B Physical Sciences*, vol. 17, pp. 793–806, 1914.

[52] A. Kovalenko, "Molecular theory of solvation: Methodology summary and illustrations," *Condensed Matter Physics*, vol. 18, p. 32601, Sept. 2015.

[53] T. Kloss, J. Heil, and S. M. Kast, "Quantum Chemistry in Solution by Combining 3d Integral Equation Theory with a Cluster Embedding Approach," *The Journal of Physical Chemistry B*, vol. 112, pp. 4337–4343, Apr. 2008.

[54] F. Hoffgaard, J. Heil, and S. M. Kast, "Three-Dimensional RISM Integral Equation Theory for Polarizable Solute Models," *Journal of Chemical Theory and Computation*, vol. 9, pp. 4718–4726, Nov. 2013.

[55] J.-P. Hansen and I. R. McDonald, "Theory of Simple Liquids," in *Theory of Simple Liquids (Fourth Edition)*, p. i, Oxford: Academic Press, 2013.

[56] P. H. Fries and G. N. Patey, "The solution of the hypernetted chain approximation for fluids of nonspherical particles. A general method with application to dipolar hard spheres," *The Journal of Chemical Physics*, vol. 82, pp. 429–440, Jan. 1985.

[57] A. Kovalenko and F. Hirata, "Self-consistent description of a metal–water interface by the Kohn–Sham density functional theory and the three-dimensional reference interaction site model," *The Journal of Chemical Physics*, vol. 110, pp. 10095–10112, May 1999.

[58] S. M. Kast and T. Kloss, "Closed-form expressions of the chemical potential for integral equation closures with certain bridge functions," *The Journal of Chemical Physics*, vol. 129, p. 236101, Dec. 2008.

[59] V. Shapovalov, T. N. Truong, A. Kovalenko, and F. Hirata, "Liquid structure at metal oxide–water interface: accuracy of a three-dimensional RISM methodology," *Chemical Physics Letters*, vol. 320, pp. 186–193, Mar. 2000.

[60] J. S. Perkyns and B. Montgomery Pettitt, "A dielectrically consistent interaction site theory for solvent - electrolyte mixtures," *Chemical Physics Letters*, vol. 190, pp. 626–630, Mar. 1992.

[61] S. Gusarov, B. S. Pujari, and A. Kovalenko, “Efficient treatment of solvation shells in 3d molecular theory of solvation,” *Journal of Computational Chemistry*, vol. 33, pp. 1478–1494, June 2012.

[62] E. Brunk and U. Rothlisberger, “Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States,” *Chemical Reviews*, vol. 115, pp. 6217–6263, June 2015.

[63] H. M. Senn and W. Thiel, “QM/MM Methods for Biomolecular Systems,” *Angewandte Chemie International Edition*, vol. 48, pp. 1198–1229, Feb. 2009.

[64] K. Senthilkumar, J. I. Mujika, K. E. Ranaghan, F. R. Manby, A. J. Mulholland, and J. N. Harvey, “Analysis of polarization in QM/MM modelling of biologically relevant hydrogen bonds,” *Journal of The Royal Society Interface*, vol. 5, pp. 207–216, Dec. 2008.

[65] K. E. Shaw, C. J. Woods, and A. J. Mulholland, “Compatibility of Quantum Chemical Methods and Empirical (MM) Water Models in Quantum Mechanics/Molecular Mechanics Liquid Water Simulations,” *The Journal of Physical Chemistry Letters*, vol. 1, pp. 219–223, Jan. 2010.

[66] S. Sumner, P. Söderhjelm, and U. Ryde, “Effect of Geometry Optimizations on QM-Cluster and QM/MM Studies of Reaction Energies in Proteins,” *Journal of Chemical Theory and Computation*, vol. 9, pp. 4205–4214, Sept. 2013.

[67] L. Hu, P. Söderhjelm, and U. Ryde, “Accurate Reaction Energies in Proteins Obtained by Combining QM/MM and Large QM Calculations,” *Journal of Chemical Theory and Computation*, vol. 9, pp. 640–649, Jan. 2013.

[68] H. Hu and W. Yang, “Free Energies of Chemical Reactions in Solution and in Enzymes with Ab Initio Quantum Mechanics/Molecular Mechanics Methods,” *Annual Review of Physical Chemistry*, vol. 59, no. 1, pp. 573–601, 2008.

[69] D. A. Yarne, M. E. Tuckerman, and G. J. Martyna, “A dual length scale method for plane-wave-based, simulation studies of chemical systems modeled using mixed ab initio/empirical force field descriptions,” *The Journal of Chemical Physics*, vol. 115, pp. 3531–3539, Aug. 2001.

[70] T. Laino, F. Mohamed, A. Laio, and M. Parrinello, “An Efficient Real Space Multigrid QM/MM Electrostatic Coupling,” *Journal of Chemical Theory and Computation*, vol. 1, pp. 1176–1184, Nov. 2005.

[71] T. Laino, F. Mohamed, A. Laio, and M. Parrinello, “An Efficient Linear-Scaling Electrostatic Coupling for Treating Periodic Boundary Conditions in QM/MM Simulations,” *Journal of Chemical Theory and Computation*, vol. 2, pp. 1370–1378, Sept. 2006.

[72] W. G. Hoover, “Canonical dynamics: Equilibrium phase-space distributions,” *Physical Review A*, vol. 31, pp. 1695–1697, Mar. 1985.

[73] G. S. Grest and K. Kremer, “Molecular dynamics simulation for polymers in the presence of a heat bath,” *Physical Review A*, vol. 33, pp. 3628–3631, May 1986.

[74] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, “Equation of State Calculations by Fast Computing Machines,” *The Journal of Chemical Physics*, vol. 21, p. 1087, June 1953.

[75] G. O. Roberts, A. Gelman, and W. R. Gilks, “Weak convergence and optimal scaling of random walk Metropolis algorithms,” *The Annals of Applied Probability*, vol. 7, pp. 110–120, Feb. 1997.

[76] M. Bédard, “Optimal acceptance rates for Metropolis algorithms: Moving beyond 0.234,” *Stochastic Processes and their Applications*, vol. 118, pp. 2198–2222, Dec. 2008.

[77] J. K. Shah, E. Marin-Rimoldi, R. G. Mullen, B. P. Keene, S. Khan, A. S. Paluch, N. Rai, L. L. Romanielo, T. W. Rosch, B. Yoo, and E. J. Maginn, “Cassandra: An open source Monte Carlo package for molecular simulation,” *Journal of Computational Chemistry*, vol. 38, pp. 1727–1739, July 2017.

[78] K. Coutinho and S. Canuto, “DICE: A Monte Carlo program for molecular liquid simulation,” 1997. see `http://fig.if.usp.br/~kaline/`.

[79] R. W. Zwanzig, “High Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases,” *The Journal of Chemical Physics*, vol. 22, pp. 1420–1426, Aug. 1954.

[80] C. H. Bennett, “Efficient estimation of free energy differences from Monte Carlo data,” *Journal of Computational Physics*, vol. 22, pp. 245–268, Oct. 1976.

[81] E. Runge and E. K. U. Gross, “Density-Functional Theory for Time-Dependent Systems,” *Physical Review Letters*, vol. 52, pp. 997–1000, Mar. 1984.

[82] J. Tomasi, E. Cancès, C. S. Pomelli, M. Caricato, G. Scalmani, M. J. Frisch, R. Cammi, M. V. Basilevsky, G. N. Chuev, and B. Mennucci, “Modern Theories of Continuum Models,” in *Continuum Solvation Models in Chemical Physics* (B. Mennucci and R. Cammi, eds.), pp. 1–123, John Wiley & Sons, Ltd, 2007.

[83] Sebastião Miranda, Jonas Feldt, Frederico Pratas, Ricardo A. Mata, Nuno Roma, and Pedro Tomás, “Efficient parallelization of perturbative Monte Carlo QM/MM simulations in heterogeneous platforms,” *The International Journal of High Performance Computing Applications*, vol. 31, pp. 499–516, Nov. 2017.

[84] J. Gao, “An Automated Procedure for Simulating Chemical Reactions in Solution. Application to the Decarboxylation of 3-Carboxybenzisoxazole in Water,” *Journal of the American Chemical Society*, vol. 117, pp. 8600–8607, Aug. 1995.

[85] 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," *Journal of Computational Chemistry*, vol. 17, pp. 19–29, Jan. 1996.

[86] T. N. Truong and E. V. Stefanovich, "Development of a perturbative approach for Monte Carlo simulations using a hybrid ab initio QM/MM method," *Chemical Physics Letters*, vol. 256, pp. 348–352, June 1996.

[87] E. Cubero, F. J. Luque, M. Orozco, and J. Gao, "Perturbation Approach to Combined QM/MM Simulation of Solute−Solvent Interactions in Solution," *The Journal of Physical Chemistry B*, vol. 107, pp. 1664–1671, Feb. 2003.

[88] T. Janowski, K. Wolinski, and P. Pulay, "Ultrafast Quantum Mechanics/Molecular Mechanics Monte Carlo simulations using generalized multipole polarizabilities," *Chemical Physics Letters*, vol. 530, pp. 1–9, Mar. 2012.

[89] T. N. Truong and E. V. Stefanovich, "Microsolvation of Cl anion by water clusters: Pertubative Monte Carlo simulations using a hybrid HF/MM potential," *Chemical Physics*, vol. 218, pp. 31–36, May 1997.

[90] T. J. Evans and T. N. Truong, "Optimizing efficiency of perturbative Monte Carlo method," *Journal of Computational Chemistry*, vol. 19, pp. 1632–1638, Nov. 1998.

[91] S. F. Boys, "Electronic Wave Functions. I. A General Method of Calculation for the Stationary States of Any Molecular System," *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 200, pp. 542–554, Feb. 1950.

[92] A. K. H. Weiss and C. Ochsenfeld, "A rigorous and optimized strategy for the evaluation of the Boys function kernel in molecular electronic structure theory," *Journal of Computational Chemistry*, vol. 36, pp. 1390–1398, July 2015.

[93] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, *et al.*, "MOLPRO, version 2012.1, a package of ab initio programs," 2012. see `www.molpro.net`.

[94] J. W. Ponder and F. M. Richards, "An efficient newton-like method for molecular mechanics energy minimization of large molecules," *Journal of Computational Chemistry*, vol. 8, pp. 1016–1024, Oct. 1987.

[95] J. Feldt, *Entwicklung einer störungstheoretischen QM/MM Monte Carlo Methode für die Studie von Molekülen in Lösung*. Master Thesis, Georg-August-Universität Göttingen, Göttingen, 2013.

[96] D. A. Pearlman, "A Comparison of Alternative Approaches to Free Energy Calculations," *The Journal of Physical Chemistry*, vol. 98, pp. 1487–1493, Feb. 1994.

[97] G. J. Rocklin, D. L. Mobley, and K. A. Dill, "Separated topologies - A method for relative binding free energy calculations using orientational restraints," *The Journal of Chemical Physics*, vol. 138, Feb. 2013.

[98] J. W. Pitera and W. F. v. Gunsteren, “A Comparison of Non-Bonded Scaling Approaches for Free Energy Calculations,” *Molecular Simulation*, vol. 28, pp. 45–65, Jan. 2002.

[99] “Valgrind Home.”

[100] R. H. Gamma, Ralph Johnson & John Vlissidess Erich, *Design Patterns: Elements Of Reusable Object-Oriented Software.* Pearson India, 2015.

[101] S. Ryazanov, “Member Function Pointers and the Fastest Possible C++ Delegates - CodeProject,” 16.01.2018. see `https://www.codeproject.com/Articles/7150/Member-Function-Pointers-and-the-Fastest-Possible`.

[102] S. A. Kryukov, “The Impossibly Fast C++ Delegates, Fixed - CodeProject,” 16.01.2018. see `https://www.codeproject.com/Articles/1170503/The-Impossibly-Fast-Cplusplus-Delegates-Fixed`.

[103] M. Matsumoto and T. Nishimura, “Mersenne Twister: A 623-dimensionally Equidistributed Uniform Pseudo-random Number Generator,” *ACM Trans. Model. Comput. Simul.*, vol. 8, pp. 3–30, Jan. 1998.

[104] M. Saito and M. Matsumoto, “Variants of Mersenne Twister Suitable for Graphic Processors,” *ACM Trans. Math. Softw.*, vol. 39, pp. 1–20, Feb. 2013.

[105] T. F. Chan, G. H. Golub, and R. J. LeVeque, “Algorithms for Computing the Sample Variance: Analysis and Recommendations,” *The American Statistician*, vol. 37, pp. 242–247, Aug. 1983.

[106] M. R. Shirts and J. D. Chodera, “Statistically optimal analysis of samples from multiple equilibrium states,” *The Journal of Chemical Physics*, vol. 129, p. 124105, Sept. 2008.

[107] J. D. Chodera, W. C. Swope, J. W. Pitera, C. Seok, and K. A. Dill, “Use of the Weighted Histogram Analysis Method for the Analysis of Simulated and Parallel Tempering Simulations,” *Journal of Chemical Theory and Computation*, vol. 3, no. 1, pp. 26–41, 2006.

[108] M. Lindner, D. Marjamäki, A. Tytula, G. Herteg, M. Renaud, and J. Tallon, “Libconfig,” 16.01.2018. see `http://hyperrealm.github.io/libconfig/`.

[109] J. Feldt, S. Miranda, F. Pratas, N. Roma, P. Tomás, and R. A. Mata, “Optimization and benchmarking of a perturbative Metropolis Monte Carlo quantum mechanics/molecular mechanics program,” *The Journal of Chemical Physics*, vol. 147, p. 244105, Dec. 2017.

[110] J. P. Perdew, K. Burke, and M. Ernzerhof, “Generalized Gradient Approximation Made Simple,” *Physical Review Letters*, vol. 77, pp. 3865–3868, Oct. 1996.

[111] F. Weigend and R. Ahlrichs, “Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy,” *Physical Chemistry Chemical Physics*, vol. 7, pp. 3297–3305, Aug. 2005.

[112] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, “A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu,” *The Journal of Chemical Physics*, vol. 132, p. 154104, Apr. 2010.

[113] S. Grimme, S. Ehrlich, and L. Goerigk, “Effect of the damping function in dispersion corrected density functional theory,” *Journal of Computational Chemistry*, vol. 32, pp. 1456–1465, May 2011.

[114] E. R. Johnson and A. D. Becke, “A post-Hartree–Fock model of intermolecular interactions,” *The Journal of Chemical Physics*, vol. 123, p. 024101, July 2005.

[115] D. Shivakumar, J. Williams, Y. Wu, W. Damm, J. Shelley, and W. Sherman, “Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field,” *Journal of Chemical Theory and Computation*, vol. 6, pp. 1509–1519, May 2010.

[116] M. Wang, P. Li, X. Jia, W. Liu, Y. Shao, W. Hu, J. Zheng, B. R. Brooks, and Y. Mei, “Efficient Strategy for the Calculation of Solvation Free Energies in Water and Chloroform at the Quantum Mechanical/Molecular Mechanical Level,” *Journal of Chemical Information and Modeling*, Sept. 2017.

[117] G. D. Hawkins, C. J. Cramer, and D. G. Truhlar, “Universal Quantum Mechanical Model for Solvation Free Energies Based on Gas-Phase Geometries,” *The Journal of Physical Chemistry B*, vol. 102, pp. 3257–3271, Apr. 1998.

[118] V. N. Levchuk, I. I. Sheikhet, and B. Y. Simkin, “Calculation of the properties of liquid chloroform by the Monte Carlo method,” *Theoretical and Experimental Chemistry*, vol. 25, pp. 66–68, Jan. 1989.

[119] L. Robbin Martin and D. E. Damschen, “Aqueous oxidation of sulfur dioxide by hydrogen peroxide at low pH,” *Atmospheric Environment (1967)*, vol. 15, pp. 1615–1621, Jan. 1981.

[120] Awwarf, B. Langlais, and A. W. W. R. Fund, *Ozone in Water Treatment: Application and Engineering*. Chelsea, Mich: Lewis Publ, 1991.

[121] F. Antunes and E. Cadenas, “Estimation of H_2O_2 gradients across biomembranes,” *FEBS Letters*, vol. 475, pp. 121–126, June 2000.

[122] G. P. Bienert, J. K. Schjoerring, and T. P. Jahn, “Membrane transport of hydrogen peroxide,” *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 1758, pp. 994–1003, Aug. 2006.

[123] H. S. Marinho, C. Real, L. Cyrne, H. Soares, and F. Antunes, “Hydrogen peroxide sensing, signaling and regulation of transcription factors,” *Redox Biology*, vol. 2, pp. 535–562, Jan. 2014.

[124] B. D'Autréaux and M. B. Toledano, "ROS as signalling molecules: mechanisms that generate specificity in ROS homeostasis," *Nature Reviews Molecular Cell Biology*, vol. 8, p. 813, Oct. 2007.

[125] M. T. C. Martins-Costa and M. F. Ruiz-López, "Molecular dynamics of hydrogen peroxide in liquid water using a combined quantum/classical force field," *Chemical Physics*, vol. 332, pp. 341–347, Feb. 2007.

[126] D.-m. Du, A.-p. Fu, and Z.-y. Zhou, "Theoretical study of the rotation barrier of hydrogen peroxide in hydrogen bonded structure of HOOH-H_2O complexes in gas and solution phase," *Journal of Molecular Structure: THEOCHEM*, vol. 717, pp. 127–132, Mar. 2005.

[127] M. C. Caputo, P. F. Provasi, L. Benitez, H. C. Georg, S. Canuto, and K. Coutinho, "Monte Carlo–Quantum Mechanics Study of Magnetic Properties of Hydrogen Peroxide in Liquid Water," *The Journal of Physical Chemistry A*, vol. 118, pp. 6239–6247, Aug. 2014.

[128] T. Kitayama, H. Kiyonaga, K. Morihashi, O. Takahashi, and O. Kikuchi, "Ab initio spin–orbit coupling SCF calculation of parity-violating energy of chiral molecules," *Journal of Molecular Structure: THEOCHEM*, vol. 589-590, pp. 183–193, Aug. 2002.

[129] P. K. Chattaraj, P. Fuentealba, P. Jaque, and A. Toro-Labbé, "Validity of the Minimum Polarizability Principle in Molecular Vibrations and Internal Rotations: An ab Initio SCF Study," *The Journal of Physical Chemistry A*, vol. 103, pp. 9307–9312, Nov. 1999.

[130] M. Tyblewski, T. Ha, R. Meyer, A. Bauder, and C. E. Blom, "Microwave and millimeter wave spectra, electric dipole moment, and internal rotation effects of methyl hydroperoxide," *The Journal of Chemical Physics*, vol. 97, pp. 6168–6180, Nov. 1992.

[131] J. Koput, "On the r_0^* structure and the torsional potential function of hydrogen peroxide," *Journal of Molecular Spectroscopy*, vol. 115, pp. 438–441, Feb. 1986.

[132] M. W. Mahoney and W. L. Jorgensen, "A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions," *The Journal of Chemical Physics*, vol. 112, pp. 8910–8922, May 2000.

[133] Y.-H. Chung, J. Xia, and C. J. Margulis, "Diffusion and Residence Time of Hydrogen Peroxide and Water in Crowded Protein Environments," *The Journal of Physical Chemistry B*, vol. 111, pp. 13336–13344, Nov. 2007.

[134] D. G. Fedorov, Y. Sugita, and C. H. Choi, "Efficient Parallel Implementations of QM/MM-REMD (Quantum Mechanical/Molecular Mechanics-Replica-Exchange MD) and Umbrella Sampling: Isomerization of H_2O_2 in Aqueous Solution," *The Journal of Physical Chemistry B*, vol. 117, pp. 7996–8002, July 2013.

[135] D. Nocito and G. J. O. Beran, “Averaged Condensed Phase Model for Simulating Molecules in Complex Environments,” *Journal of Chemical Theory and Computation*, Feb. 2017.

[136] M. Pagliai, S. Raugei, G. Cardini, and V. Schettino, “Car–Parrinello molecular dynamics on the S_N2 reaction Cl^-+CH_3Br in water,” *Journal of Molecular Structure: THEOCHEM*, vol. 630, pp. 141–149, July 2003.

[137] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, “Comparison of simple potential functions for simulating liquid water,” *The Journal of Chemical Physics*, vol. 79, pp. 926–935, July 1983.

[138] B. J. C. Cabral, “Dynamics, magnetic properties, and electron binding energies of H_2O_2 in water,” *The Journal of Chemical Physics*, vol. 146, p. 234502, June 2017.

[139] C. Yu, L. Gong, and Z. Yang, “Theoretical study on the hydration of hydrogen peroxide in terms of ab initio method and atom-bond electronegativity equalization method fused into molecular mechanics,” *Frontiers of Chemistry in China*, vol. 6, pp. 287–299, Jan. 2012.

[140] S. T. Moin, T. S. Hofer, B. R. Randolf, and B. M. Rode, “An ab initio quantum mechanical charge field molecular dynamics simulation of hydrogen peroxide in water,” *Computational and Theoretical Chemistry*, vol. 980, pp. 15–22, Jan. 2012.

[141] M. J. Gillan, D. Alfè, P. J. Bygrave, C. R. Taylor, and F. R. Manby, “Energy benchmarks for water clusters and ice structures from an embedded many-body expansion,” *The Journal of Chemical Physics*, vol. 139, p. 114101, Sept. 2013.

[142] E. Miliordos and S. S. Xantheas, “An accurate and efficient computational protocol for obtaining the complete basis set limits of the binding energies of water clusters at the MP2 and CCSD(T) levels of theory: Application to $(H_2O)_m$, m = 2–6, 8, 11, 16, and 17,” *The Journal of Chemical Physics*, vol. 142, p. 234303, June 2015.

[143] D. Yuan, Y. Li, Z. Ni, P. Pulay, W. Li, and S. Li, “Benchmark Relative Energies for Large Water Clusters with the Generalized Energy-Based Fragmentation Method,” *Journal of Chemical Theory and Computation*, vol. 13, pp. 2696–2704, June 2017.

[144] J. P. Furtado, A. P. Rahalkar, S. Shanker, P. Bandyopadhyay, and S. R. Gadre, “Facilitating Minima Search for Large Water Clusters at the MP2 Level via Molecular Tailoring,” *The Journal of Physical Chemistry Letters*, vol. 3, pp. 2253–2258, Aug. 2012.

[145] T. N. Truong and E. V. Stefanovich, “Hydration effects on reaction profiles: an ab initio dielectric continuum study of the S_N2 Cl^- + CH_3Cl reaction,” *The Journal of Physical Chemistry*, vol. 99, pp. 14700–14706, Oct. 1995.

[146] C. K. Regan, S. L. Craig, and J. I. Brauman, “Steric Effects and Solvent Effects in Ionic Reactions,” *Science*, vol. 295, pp. 2245–2247, Mar. 2002.

[147] M. Cossi, C. Adamo, and V. Barone, “Solvent effects on an S_N2 reaction profile,” *Chemical Physics Letters*, vol. 297, pp. 1–7, Nov. 1998.

[148] B. Ensing, E. J. Meijer, P. E. Blöchl, and E. J. Baerends, “Solvation Effects on the S_N2 Reaction between CH_3Cl and Cl^- in Water,” *The Journal of Physical Chemistry A*, vol. 105, pp. 3300–3310, Apr. 2001.

[149] K. Coutinho and S. Canuto, “Solvent Effects from a Sequential Monte Carlo - Quantum Mechanical Approach,” in *Advances in Quantum Chemistry* (P.-O. Löwdin, J. R. Sabin, M. C. Zerner, J. Karwowski, and M. Karelson, eds.), vol. 28, pp. 89–105, Academic Press, Jan. 1997.

[150] S. Canuto and K. Coutinho, “From hydrogen bond to bulk: Solvation analysis of the n-π* transition of formaldehyde in water,” *International Journal of Quantum Chemistry*, vol. 77, pp. 192–198, Jan. 2000.

[151] A. Marini, A. Muñoz-Losa, A. Biancardi, and B. Mennucci, “What is Solvatochromism?,” *The Journal of Physical Chemistry B*, vol. 114, pp. 17128–17135, Dec. 2010.

[152] C. Reichardt, “Solvatochromic Dyes as Solvent Polarity Indicators,” *Chemical Reviews*, vol. 94, pp. 2319–2358, Dec. 1994.

[153] C. Reichardt, *Solvents and solvent effects in organic chemistry*. VCH, 1988.

[154] T. Fujisawa, M. Terazima, and Y. Kimura, “Solvent Effects on the Local Structure of p-Nitroaniline in Supercritical Water and Supercritical Alcohols,” *The Journal of Physical Chemistry A*, vol. 112, pp. 5515–5526, June 2008.

[155] Y. Shiraishi, T. Inoue, and T. Hirai, “Local Viscosity Analysis of Triblock Copolymer Micelle with Cyanine Dyes as a Fluorescent Probe,” *Langmuir*, vol. 26, pp. 17505–17512, Nov. 2010.

[156] V. Cavalli, D. C. d. Silva, C. Machado, V. G. Machado, and V. Soldi, “The Fluorosolvatochromism of Brooker’s Merocyanine in Pure and in Mixed Solvents,” *Journal of Fluorescence*, vol. 16, pp. 77–86, Jan. 2006.

[157] F. M. Testoni, E. A. Ribeiro, L. A. Giusti, and V. G. Machado, “Merocyanine solvatochromic dyes in the study of synergistic effects in mixtures of chloroform with hydrogen-bond accepting solvents,” *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 71, pp. 1704–1711, Jan. 2009.

[158] C. J. Cramer and D. G. Truhlar, “Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics,” *Chemical Reviews*, vol. 99, pp. 2161–2200, Aug. 1999.

[159] T. Sakata, Y. Kawashima, and H. Nakano, “Solvent effect on the absorption spectra of coumarin 120 in water: A combined quantum mechanical and molecular mechanical study,” *The Journal of Chemical Physics*, vol. 134, p. 014501, Jan. 2011.

[160] N. A. Murugan, P. C. Jha, Z. Rinkevicius, K. Ruud, and H. Ågren, “Solvatochromic shift of phenol blue in water from a combined Car–Parrinello molecular dynamics hybrid quantum mechanics-molecular mechanics and ZINDO approach,” *The Journal of Chemical Physics*, vol. 132, p. 234508, June 2010.

[161] C. M. Isborn, A. W. Götz, M. A. Clark, R. C. Walker, and T. J. Martínez, “Electronic Absorption Spectra from MM and ab Initio QM/MM Molecular Dynamics: Environmental Effects on the Absorption Spectrum of Photoactive Yellow Protein,” *Journal of Chemical Theory and Computation*, vol. 8, pp. 5092–5106, Dec. 2012.

[162] M. Bergeler, H. Mizuno, E. Fron, and J. N. Harvey, “QM/MM-Based Calculations of Absorption and Emission Spectra of LSSmOrange Variants,” *The Journal of Physical Chemistry B*, vol. 120, pp. 12454–12465, Dec. 2016.

[163] T. H. D. Jr, “Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen,” *The Journal of Chemical Physics*, vol. 90, pp. 1007–1023, Jan. 1989.

[164] S. Hirata, M. Valiev, M. Dupuis, S. S. Xantheas, S. Sugiki, and H. Sekino, “Fast electron correlation methods for molecular clusters in the ground and excited states,” *Molecular Physics*, vol. 103, pp. 2255–2265, Aug. 2005.

[165] Y. Mochizuki, Y. Komeiji, T. Ishikawa, T. Nakano, and H. Yamataka, “A fully quantum mechanical simulation study on the lowest n–π state of hydrated formaldehyde,” *Chemical Physics Letters*, vol. 437, pp. 66–72, Mar. 2007.

[166] H. Fukunaga and K. Morokuma, “Cluster and solution simulation of formaldehyde-water complexes and solvent effect on formaldehyde $^1(n,\pi^*)$ transition,” *The Journal of Physical Chemistry*, vol. 97, pp. 59–69, Jan. 1993.

[167] B. Mennucci, R. Cammi, and J. Tomasi, “Excited states and solvatochromic shifts within a nonequilibrium solvation approach: A new formulation of the integral equation formalism method at the self-consistent field, configuration interaction, and multiconfiguration self-consistent field level,” *The Journal of Chemical Physics*, vol. 109, pp. 2798–2807, Aug. 1998.

[168] K. Naka, A. Morita, and S. Kato, “Effect of solvent fluctuation on the electronic transitions of formaldehyde in aqueous solution,” *The Journal of Chemical Physics*, vol. 110, pp. 3484–3492, Feb. 1999.

[169] M. E. Martín, M. L. Sánchez, F. J. Olivares del Valle, and M. A. Aguilar, “A multiconfiguration self-consistent field/molecular dynamics study of the $(\mathrm{n} \to \pi^*)^1$ transition of carbonyl compounds in liquid water,” *The Journal of Chemical Physics*, vol. 113, pp. 6308–6315, Oct. 2000.

[170] Y. Kawashima, M. Dupuis, and K. Hirao, “Monte Carlo microsolvation simulations for excited states using a mixed-Hamiltonian model with polarizable and vibrating waters: Applications to the blueshift of the H_2CO $^1(\pi^* \leftarrow n)$ excitation,” *The Journal of Chemical Physics*, vol. 117, pp. 248–257, June 2002.

[171] J. T. Blair, K. Krogh-Jespersen, and R. M. Levy, “Solvent effects on optical absorption spectra: the $^1A_1 \rightarrow {}^1A_2$ transition of formaldehyde in water,” *Journal of the American Chemical Society*, vol. 111, pp. 6948–6956, Aug. 1989.

[172] M. A. Thompson, “QM/MMpol: A Consistent Model for Solute/Solvent Polarization. Application to the Aqueous Solvation and Spectroscopy of Formaldehyde, Acetaldehyde, and Acetone,” *The Journal of Physical Chemistry*, vol. 100, pp. 14492–14507, Jan. 1996.

[173] Z. Xu and S. Matsika, “Combined Multireference Configuration Interaction/ Molecular Dynamics Approach for Calculating Solvatochromic Shifts: Application to the $n_O \rightarrow \pi^*$ Electronic Transition of Formaldehyde,” *The Journal of Physical Chemistry A*, vol. 110, pp. 12035–12043, Nov. 2006.

[174] A. Öhrn and G. Karlström, “A theoretical study of the solvent shift to the transition in formaldehyde with an effective discrete quantum chemical solvent model including non-electrostatic perturbation,” *Molecular Physics*, vol. 104, pp. 3087–3099, Oct. 2006.

[175] J. Kongsted, A. Osted, K. V. Mikkelsen, P.-O. Åstrand, and O. Christiansen, “Solvent effects on the $n \rightarrow \pi^*$ electronic transition in formaldehyde: A combined coupled cluster/-molecular dynamics study,” *The Journal of Chemical Physics*, vol. 121, pp. 8435–8445, Oct. 2004.

[176] D. E. Freeman and W. Klemperer, “Electric Dipole Moment of the $^1A_2$ Electronic State of Formaldehyde,” *The Journal of Chemical Physics*, vol. 45, pp. 52–57, July 1966.

[177] M. Robin, *Higher Excited States of Polyatomic Molecules*. Elsevier, Dec. 2012.

[178] P. Suppan, “Invited review solvatochromic shifts: The influence of the medium on the energy of electronic states,” *Journal of Photochemistry and Photobiology A: Chemistry*, vol. 50, pp. 293–330, Jan. 1990.

[179] M. Merchán and B. O. Roos, “A theoretical determination of the electronic spectrum of formaldehyde,” *Theoretica chimica acta*, vol. 92, pp. 227–239, Oct. 1995.

[180] C. Angeli, S. Borini, L. Ferrighi, and R. Cimiraglia, “Ab initio n-electron valence state perturbation theory study of the adiabatic transitions in carbonyl molecules: Formaldehyde, acetaldehyde, and acetone,” *The Journal of Chemical Physics*, vol. 122, p. 114304, Mar. 2005.

[181] A. F. Moskvin, O. P. Yablonskii, and L. F. Bondar, “An experimental investigation of the effect of alkyl substituents on the position of the K and R absorption bands in acrolein derivatives,” *Theoretical and Experimental Chemistry*, vol. 2, pp. 469–472, Sept. 1966.

[182] K. Inuzuka, "Near Ultraviolet Absorption Spectra of Acrolein and Crotonaldehyde," *Bulletin of the Chemical Society of Japan*, vol. 33, pp. 678–680, May 1960.

[183] F. E. Blacet, W. G. Young, and J. G. Roof, "Studies of Absorption Spectra. I. Crotonaldehyde and Acrolein," *Journal of the American Chemical Society*, vol. 59, pp. 608–614, Apr. 1937.

[184] A. C. P. Alves, J. Christoffersen, and J. M. Hollas, "Near ultra-violet spectra of the s-trans and a second rotamer of acrolein vapour," *Molecular Physics*, vol. 20, pp. 625–644, Jan. 1971.

[185] J. F. Horwood and J. R. Williams, "Vapour phase carbonyl absorption in the far ultra-violet," *Spectrochimica Acta*, vol. 19, pp. 1351–1362, Aug. 1963.

[186] A. D. Walsh, "The absorption spectra of acrolein, crotonaldehyde and mesityl oxide in the vacuum ultra-violet," *Transactions of the Faraday Society*, vol. 41, pp. 498–505, Jan. 1945.

[187] A. M. Buswell, E. C. Dunlop, W. H. Rodebush, and J. B. Swartz, "Change of the Ultra-violet Absorption Spectrum of Acrolein with Time," *Journal of the American Chemical Society*, vol. 62, pp. 325–328, Feb. 1940.

[188] G. Mackinney and O. Temmer, "The Deterioration of Dried Fruit. IV. Spectrophotometric and Polarographic Studies," *Journal of the American Chemical Society*, vol. 70, pp. 3586–3590, Nov. 1948.

[189] H. C. Georg, K. Coutinho, and S. Canuto, "A sequential Monte Carlo quantum mechanics study of the hydrogen-bond interaction and the solvatochromic shift of the n–π* transition of acrolein in water," *The Journal of Chemical Physics*, vol. 123, p. 124307, Sept. 2005.

[190] F. Aquilante, V. Barone, and B. O. Roos, "A theoretical investigation of valence and Rydberg electronic states of acrolein," *The Journal of Chemical Physics*, vol. 119, pp. 12323–12334, Dec. 2003.

[191] S. A. d. Monte, T. Müller, M. Dallos, H. Lischka, M. Diedenhofen, and A. Klamt, "Solvent effects in electronically excited states using the continuum solvation model COSMO in combination with multireference configuration interaction with singles and doubles (MR-CISD)," *Theoretical Chemistry Accounts*, vol. 111, pp. 78–89, Mar. 2004.

[192] K. Aidas, A. Møgelhøj, E. J. K. Nilsson, M. S. Johnson, K. V. Mikkelsen, O. Christiansen, P. Söderhjelm, and J. Kongsted, "On the performance of quantum chemical methods to predict solvatochromic effects: The case of acrolein in aqueous solution," *The Journal of Chemical Physics*, vol. 128, p. 194503, May 2008.

[193] M. E. Martín, A. Muñoz Losa, I. Fdez.-Galván, and M. A. Aguilar, "A theoretical study of solvent effects on the $^1(n \rightarrow \pi^*)$ electron transition in acrolein," *The Journal of Chemical Physics*, vol. 121, pp. 3710–3716, Aug. 2004.

[194] R. A. Mata and B. J. Costa Cabral, “Chapter 4 - QM/MM Approaches to the Electronic Spectra of Hydrogen-Bonding Systems with Connection to Many-Body Decomposition Schemes,” in *Advances in Quantum Chemistry* (J. R. Sabin and E. Brändas, eds.), vol. 59 of *Combining Quantum Mechanics and Molecular Mechanics. Some Recent Progresses in QM/MM Methods*, pp. 99–144, Academic Press, Jan. 2010.

[195] R. A. Mata, “Assessing the accuracy of many-body expansions for the computation of solvatochromic shifts,” *Molecular Physics*, vol. 108, pp. 381–392, Feb. 2010.

[196] C. Bistafa, L. Modesto-Costa, and S. Canuto, “A complete basis set study of the lowest n–π* and π–π* electronic transitions of acrolein in explicit water environment,” *Theoretical Chemistry Accounts*, vol. 135, p. 129, May 2016.

[197] M. G. Müller, E. H. Hardy, P. S. Vogt, C. Bratschi, B. Kirchner, H. Huber, and D. J. Searles, “Calculation of the deuteron quadrupole relaxation rate in a mixture of water and dimethyl sulfoxide,” *Journal of the American Chemical Society*, vol. 126, pp. 4704–4710, Apr. 2004.

[198] J. Catalán, C. Díaz, and F. García-Blanco, “Characterization of binary solvent mixtures of DMSO with water and other cosolvents,” *The Journal of Organic Chemistry*, vol. 66, pp. 5846–5852, Aug. 2001.

[199] S. J. Suresh, “Detailed Molecular Model for Dielectric Constant of Multicomponent, Associating Liquids,” *The Journal of Physical Chemistry B*, vol. 108, pp. 715–720, Jan. 2004.

[200] L.-J. Yang, X.-Q. Yang, K.-M. Huang, G.-Z. Jia, and H. Shang, “Dielectric Properties of Binary Solvent Mixtures of Dimethyl Sulfoxide with Water,” *International Journal of Molecular Sciences*, vol. 10, pp. 1261–1270, Mar. 2009.

[201] M. Trumm, C. Adam, C. Koke, M. Maiwald, S. Höfener, A. Skerencak-Frech, P. J. Panak, and B. Schimmelpfennig, “The influence of polarity in binary solvent mixtures on the conformation of bis-triazinyl-pyridine in solution,” *Molecular Physics*, vol. 116, pp. 507–514, Feb. 2018.

[202] K. Karhan, R. Z. Khaliullin, and T. D. Kühne, “On the role of interfacial hydrogen bonds in “on-water” catalysis,” *The Journal of Chemical Physics*, vol. 141, p. 22D528, Dec. 2014.

[203] A. Cavagna, T. S. Grigera, and P. Verrocchio, “Dynamic relaxation of a liquid cavity under amorphous boundary conditions,” *The Journal of Chemical Physics*, vol. 136, p. 204502, May 2012.

[204] D. Gazzillo and G. Pastore, “Equation of state for symmetric non-additive hard-sphere fluids: An approximate analytic expression and new Monte Carlo results,” *Chemical Physics Letters*, vol. 159, pp. 388–392, July 1989.

[205] H. Ikeda, F. Zamponi, and A. Ikeda, "Mean field theory of the swap Monte Carlo algorithm," *The Journal of Chemical Physics*, vol. 147, p. 234506, Dec. 2017.

[206] M. N. Rosenbluth and A. W. Rosenbluth, "Monte Carlo Calculation of the Average Extension of Molecular Chains," *The Journal of Chemical Physics*, vol. 23, pp. 356–359, Feb. 1955.

[207] T. Biben, P. Bladon, and D. Frenkel, "Depletion effects in binary hard-sphere fluids," *Journal of Physics: Condensed Matter*, vol. 8, no. 50, p. 10799, 1996.

[208] P. Bai and J. I. Siepmann, "Assessment and Optimization of Configurational-Bias Monte Carlo Particle Swap Strategies for Simulations of Water in the Gibbs Ensemble," *Journal of Chemical Theory and Computation*, vol. 13, pp. 431–440, Feb. 2017.

[209] K. R. Popov and N. V. Platonova, "Polarization of electronic transitions and nature of the excitation of electronic states in the o- and m-nitroaniline molecules," *Journal of Applied Spectroscopy*, vol. 32, pp. 393–397, Apr. 1980.

[210] B. J. C. Cabral, "Electron binding energies and the fundamental gap of a push-pull dye in a polar environment: p-nitroaniline in liquid water," *Chemical Physics Letters*, vol. 667, pp. 332–336, Jan. 2017.

[211] B. J. C. Cabral, K. Coutinho, and S. Canuto, "A First-Principles Approach to the Dynamics and Electronic Properties of p-Nitroaniline in Water," *The Journal of Physical Chemistry A*, vol. 120, pp. 3878–3887, June 2016.

[212] H. Nakano and H. Sato, "An Ab Initio QM/MM-Based Approach to Efficiently Evaluate Vertical Excitation Energies in Condensed Phases Including the Nonequilibrium Solvation Effect," *The Journal of Physical Chemistry B*, vol. 120, pp. 1670–1678, Mar. 2016.

[213] A. V. Marenich, C. J. Cramer, and D. G. Truhlar, "Electronic Absorption Spectra and Solvatochromic Shifts by the Vertical Excitation Model: Solvated Clusters and Molecular Dynamics Sampling," *The Journal of Physical Chemistry B*, vol. 119, pp. 958–967, Jan. 2015.

[214] J. J. Eriksen, S. P. A. Sauer, K. V. Mikkelsen, O. Christiansen, H. J. A. Jensen, and J. Kongsted, "Failures of TDDFT in describing the lowest intramolecular charge-transfer excitation in para-nitroaniline," *Molecular Physics*, vol. 111, pp. 1235–1248, July 2013.

[215] R. Cattana, J. J. Silber, and J. Anunziata, "Dielectric enrichment in binary solvent mixtures. The intramolecular hydrogen bond in N-alkyl-substituted o-nitroanilines. Substituent effects," *Canadian Journal of Chemistry*, vol. 70, pp. 2677–2682, Oct. 1992.

[216] H. Boggetti, J. D. Anunziata, R. Cattana, and J. J. Silber, "Solvatochromic study on nitroanilines. Preferential solvation vs dielectric enrichment in binary solvent mixtures," *Spectrochimica Acta Part A: Molecular Spectroscopy*, vol. 50, pp. 719–726, Jan. 1994.

[217] J. Kircher, *Präferierte Solvatation in binären Lösungsmittelgemischen*. Bachelor Thesis, Georg-August-Universität Göttingen, Göttingen, 2017.

[218] P. Suppan, “Local polarity of solvent mixtures in the field of electronically excited molecules and exciplexes,” *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases*, vol. 83, pp. 495–509, Jan. 1987.

[219] D. V. Matyushov and M. D. Newton, “Solvent-Induced Shift of Spectral Lines in Polar–Polarizable Solvents,” *The Journal of Physical Chemistry A*, Feb. 2017.

[220] https://commons.wikimedia.org/wiki/File:DNA_replication_en.svg, “Dna replication,” 09.05.2018.

[221] B. G. Vértessy and J. Tóth, “Keeping Uracil Out of DNA: Physiological Role, Structure and Catalytic Mechanism of dUTPases,” *Accounts of chemical research*, vol. 42, pp. 97–106, Jan. 2009.

[222] R. Olinski, M. Jurgowiak, and T. Zaremba, “Uracil in DNA—Its biological significance,” *Mutation Research/Reviews in Mutation Research*, vol. 705, pp. 239–245, Dec. 2010.

[223] B. K. Duncan and H. R. Warner, “Metabolism of uracil-containing DNA: degradation of bacteriophage PBS2 DNA in Bacillus subtilis,” *Journal of Virology*, vol. 22, pp. 835–838, June 1977.

[224] F. H. C. Crick and J. D. Watson, “Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid,” *Nature*, vol. 171, p. 737, Apr. 1953.

[225] W. Wang, H. W. Hellinga, and L. S. Beese, “Structural evidence for the rare tautomer hypothesis of spontaneous mutagenesis,” *Proceedings of the National Academy of Sciences*, vol. 108, pp. 17644–17648, Oct. 2011.

[226] J. P. Henderson, J. Byun, D. M. Mueller, and J. W. Heinecke, “The Eosinophil Peroxidase-Hydrogen Peroxide-Bromide System of Human Eosinophils Generates 5-Bromouracil, a Mutagenic Thymine Analogue,” *Biochemistry*, vol. 40, pp. 2052–2059, Feb. 2001.

[227] V. Valinluck, P. Liu, J. I. Kang, A. Burdzy, and L. C. Sowers, “5-Halogenated pyrimidine lesions within a CpG sequence context mimic 5-methylcytosine by enhancing the binding of the methyl-CpG-binding domain of methyl-CpG-binding protein 2 (MeCP2),” *Nucleic Acids Research*, vol. 33, pp. 3057–3064, Jan. 2005.

[228] B. Szikriszt, A. Póti, O. Pipek, M. Krzystanek, N. Kanu, J. Molnár, D. Ribli, Z. Szeltner, G. E. Tusnády, I. Csabai, Z. Szallasi, C. Swanton, and D. Szüts, “A comprehensive survey of the mutagenic impact of common cancer cytotoxics,” *Genome Biology*, vol. 17, p. 99, May 2016.

[229] T. v. Mourik, V. I. Danilov, V. V. Dailidonis, N. Kurita, H. Wakabayashi, and T. Tsukamoto, “A DFT study of uracil and 5-bromouracil in nanodroplets,” *Theoretical Chemistry Accounts*, vol. 125, pp. 233–244, Mar. 2010.

[230] X. Hu, H. Li, L. Zhang, and S. Han, “Tautomerism of Uracil and 5-Bromouracil in a Microcosmic Environment with Water and Metal Ions. What Roles Do Metal Ions Play?,” *The Journal of Physical Chemistry B*, vol. 111, pp. 9347–9354, Aug. 2007.

[231] X. Hu, H. Li, J. Ding, and S. Han, “Mutagenic Mechanism of the A-T to G-C Transition Induced by 5-Bromouracil: An ab Initio Study,” *Biochemistry*, vol. 43, pp. 6361–6369, June 2004.

[232] M. Hanus, M. Kabeláč, D. Nachtigallová, and P. Hobza, “Mutagenic Properties of 5-Halogenuracils: Correlated Quantum Chemical ab Initio Study,” *Biochemistry*, vol. 44, pp. 1701–1707, Feb. 2005.

[233] E. Pluhařová, P. Slavíček, and P. Jungwirth, “Modeling Photoionization of Aqueous DNA and Its Components,” *Accounts of Chemical Research*, vol. 48, pp. 1209–1217, May 2015.

[234] M. Orozco, B. Hernández, and F. J. Luque, “Tautomerism of 1-Methyl Derivatives of Uracil, Thymine, and 5-Bromouracil. Is Tautomerism the Basis for the Mutagenicity of 5-Bromouridine?,” *The Journal of Physical Chemistry B*, vol. 102, pp. 5228–5233, June 1998.

[235] V. I. Danilov, T. van Mourik, N. Kurita, H. Wakabayashi, T. Tsukamoto, and D. M. Hovorun, “On the Mechanism of the Mutagenic Action of 5-Bromouracil: A DFT Study of Uracil and 5-Bromouracil in a Water Cluster,” *The Journal of Physical Chemistry A*, vol. 113, pp. 2233–2235, Mar. 2009.

[236] A. R. Katritzky and A. J. Waring, “299. Tautomeric azines. Part I. The tautomerism of 1-methyluracil and 5-bromo-1-methyluracil,” *Journal of the Chemical Society (Resumed)*, pp. 1540–1544, Jan. 1962.

[237] T. van Mourik, “Abstracts: Albany 2007, The 15th Conversation,” *Journal of Biomolecular Structure and Dynamics*, vol. 24, pp. 609–776, June 2007.

[238] A. V. Marenich, J. Ho, M. L. Coote, C. J. Cramer, and D. G. Truhlar, “Computational electrochemistry: prediction of liquid-phase reduction potentials,” *Physical Chemistry Chemical Physics*, vol. 16, pp. 15068–15106, July 2014.

[239] M. A. El Hassan and C. R. Calladine, “Propeller-Twisting of Base-pairs and the Conformational Mobility of Dinucleotide Steps in DNA,” *Journal of Molecular Biology*, vol. 259, pp. 95–103, May 1996.

[240] T. Hamma and A. R. Ferré-D'Amaré, "Structure of Protein L7ae Bound to a K-Turn Derived from an Archaeal Box H/ACA sRNA at 1.8 Å Resolution," *Structure*, vol. 12, pp. 893–903, May 2004.

[241] H.-J. Werner and M. Schütz, "An efficient local coupled cluster method for accurate thermochemistry of large systems," *The Journal of Chemical Physics*, vol. 135, p. 144116, Oct. 2011.

[242] A. K. Wilson, D. E. Woon, K. A. Peterson, and T. H. Dunning, "Gaussian basis sets for use in correlated molecular calculations. IX. The atoms gallium through krypton," *The Journal of Chemical Physics*, vol. 110, pp. 7667–7676, Apr. 1999.

[243] C. Cheong, I. T. Jr, S. R. Holbrook, and S.-H. Kim, "Crystal structure of an RNA double helix incorporating a track of non-Watson–Crick base pairs," *Nature*, vol. 353, p. 579, Oct. 1991.

[244] K. Shi, M. Wahl, and M. Sundaralingam, "Crystal structure of an RNA duplex r(GGGCGCUCC)$_2$ with non-adjacent G·U base pairs," *Nucleic Acids Research*, vol. 27, pp. 2196–2201, Jan. 1999.

[245] L. C. Sowers, M. F. Goodman, R. Eritja, B. Kaplan, and G. V. Fazakerley, "Ionized and wobble base-pairing for bromouracil-guanine in equilibrium under physiological conditions: A nuclear magnetic resonance study on an oligonucleotide containing a bromouracil-guanine base-pair as a function of pH," *Journal of Molecular Biology*, vol. 205, pp. 437–447, Jan. 1989.

[246] L. C. Sowers, R. Eritja, B. Kaplan, M. F. Goodman, and G. V. Fazakerly, "Equilibrium between a wobble and ionized base pair formed between fluorouracil and guanine in DNA as studied by proton and fluorine NMR," *The Journal of Biological Chemistry*, vol. 263, pp. 14794–14801, Oct. 1988.

[247] L. Verlet, "Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules," *Physical Review*, vol. 159, pp. 98–103, July 1967.

[248] M. Předota, P. T. Cummings, and A. A. Chialvo, "Pair approximation for polarization interaction: efficient method for Monte Carlo simulations of polarizable fluids," *Molecular Physics*, vol. 99, pp. 349–354, Feb. 2001.

[249] M. W. Mahoney and W. L. Jorgensen, "Rapid estimation of electronic degrees of freedom in Monte Carlo calculations for polarizable models of liquid water," *The Journal of Chemical Physics*, vol. 114, pp. 9337–9349, May 2001.

[250] 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," *The Journal of Chemical Physics*, vol. 108, pp. 3383–3385, Mar. 1998.

[251] D. Jacquemin, “What is the Key for Accurate Absorption and Emission Calculations, Energy or Geometry?,” *Journal of Chemical Theory and Computation*, Jan. 2018.

[252] R. Guareschi, F. M. Floris, C. Amovilli, and C. Filippi, “Solvent Effects on Excited-State Structures: A Quantum Monte Carlo and Density Functional Study,” *Journal of Chemical Theory and Computation*, vol. 10, pp. 5528–5537, Dec. 2014.

[253] R. Guareschi and C. Filippi, “Ground- and Excited-State Geometry Optimization of Small Organic Molecules with Quantum Monte Carlo,” *Journal of Chemical Theory and Computation*, vol. 9, pp. 5513–5525, Dec. 2013.

[254] T. Neumann, D. Danilov, and W. Wenzel, “Multiparticle moves in acceptance rate optimized monte carlo,” *Journal of Computational Chemistry*, vol. 36, pp. 2236–2245, Nov. 2015.

[255] K. M. Bal and E. C. Neyts, “On the time scale associated with Monte Carlo simulations,” *The Journal of Chemical Physics*, vol. 141, p. 204104, Nov. 2014.

[256] E. C. Neyts, B. J. Thijsse, M. J. Mees, K. M. Bal, and G. Pourtois, “Establishing Uniform Acceptance in Force Biased Monte Carlo Simulations,” *Journal of Chemical Theory and Computation*, vol. 8, pp. 1865–1869, June 2012.

[257] R. Assaraf and M. Caffarel, “Zero-Variance Principle for Monte Carlo Algorithms,” *Physical Review Letters*, vol. 83, pp. 4682–4685, Dec. 1999.

Curriculum Vitae

Name: Jonas Feldt
Address: Ulmenweg 2b
37077 Göttingen
Germany
E-Mail: jfeldt@gwdg.de
Place of birth: Neustadt am Rübenberge, Germany
Date of birth: 17.04.1988
Nationality: German

Education

2014 - today	PhD student in Chemistry under the supervision of Prof. Dr. Ricardo Mata, Institute of Physical Chemistry, Georg-August-Universität Göttingen, Germany.
2011 - 2014	Master of Science in Chemistry at Faculty of Chemistry, Georg-August-Universität Göttingen, Germany.
2007 - 2011	Bachelor of Science in Chemistry at Faculty of Chemistry, Georg-August-Universität Göttingen, Germany.

Oral Presentations

Dec 2015	*Perturbative Monte Carlo: Absolute Solvation Free Energies* Talk at the 31st Winterschool in Theoretical Chemistry, University of Helsinki, Finland.
Aug 2017	*Perturbative Monte Carlo Simulations - A Hybrid QM/MM Approach* Talk at the Coding Solvation Workshop, Livorno, Italy.

Publications

- J. Feldt, S. Miranda, F. Pratas, N. Roma, P. Tomás, R. A. Mata, *Optimization and benchmarking of a perturbative Metropolis Monte Carlo Quantum Mechanics/Molecular Mechanics program*, J. Chem. Phys. **147**, 244105 (2017).
- S. Miranda, J. Feldt, R. A. Mata, N. Roma, P. Tomás, *Efficient parallelization of perturbative Monte Carlo QM/MM simulations in heterogeneous platforms*, Int. J. High Perform. Comput. Appl. **31**, 499-516 (2016).
- J. C. A. Oliveira, J. Feldt, N. Galamba and R. A. Mata, *Study of Specific Ion-Amino Acid Interactions through the Use of Local Correlation Methods*, J. Phys. Chem A **116**, 5464-5471 (2012).
- J. Feldt, R. A. Mata and J. M. Dieterich, *Atomdroid: A computational chemistry tool for mobile platforms*, J. Chem. Inf. Model **52**, 1072-1078 (2012).