Aalto University
School of Science

Hanne Antila

Electrostatics of a polarizable force field based on the Thole point dipole model

Master's Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Technology in the Degree Programme in Engineering Physics and Mathematics.

Espoo, 15.8.2011

Supervisor: Prof. Tapio Ala-Nissilä

Instructor: Emppu Salonen, Ph. D

AALTO UNIVERSITY
SCHOOL OF SCIENCE ABSTRACT OF THE MASTER'S THESIS

Author:	Hanne Antila
Thesis title:	Electrostatics of a polarizable force based on the Thole point dipole model
Title in Finnish:	Tholen malliin perustuvan polarisoituvan voimakentän sähköstatiikka
Degree program:	Degree Programme in Engineering Physics and Mathematics
Major subject:	Engineering Physics F3005
Minor subject:	Biomedical Engineering F3001
Chair:	Tfy-105 Computational Physics
Supervisor:	Prof. Tapio Ala-Nissilä **Instructor:** Emppu Salonen, Ph.D

Molecular dynamics (MD) simulations are widely used in the modelling of biomolecules because these models are able to provide information on those properties of biological systems which are hard to study by experimental means. The increase in computational power has provided the means to simulate more complex systems, but has also introduced both the possibility and the requirement to improve the force fields the simulations are based on.

At present, electrostatic interactions in the common MD force fields are represented as interactions between fixed partial charges. The downside is that these charges cannot accurately reflect the dependence of a charge distribution on the state of the system nor can they respond to fluctuations in the electric field due to molecular motion. For this, one should explicitly include the effect polarizability into the force field.

In this thesis, ways of parametrizing the electrostatics of a polarizable force field have been studied. It was examined how three different point charge fitting methods, MK, CHELPG, and RESP, and two multipole algorithms, DMA and GMM, perform when intramolecular polarizability contributions are self-consistently removed from the fitting done in the parametrization process. To this end, the different methods are combined with the induced point dipole model by Thole.

MK and RESP were determined to be the most promising candidates for polarizable force field parametrization at the moment. They provide a good compromise between accuracy and computational efficiency not to mention the ease of force field implementation. To our surprise, DMA multipoles up to octupoles were required to reach the same level of accuracy. The applicability of GMM is hindered by the convergence issues that arose when GMM was combined with the Thole model. Also, the functional forms of the electric interactions resulting from the GMM multipoles makes it less appealing for force field purposes.

Date: 15.8.2011 **Language:** English **Number of pages:** 70 + 28

Keywords: Polarizability, Force field, Parametrization, Multipole expansion

AALTO-YLIOPISTO
PERUSTIETEIDEN KORKEAKOULU DIPLOMITYÖN TIIVISTELMÄ

Tekijä:	Hanne Antila
Työn nimi:	Tholen malliin perustuvan polarisoituvan voimakentän sähköstatiikka
Title in English:	Electrostatics of a polarizable force field based on the Thole point dipole model
Tutkinto-ohjelma:	Teknillisen fysiikan ja matematiikan tutkinto-ohjelma
Pääaine:	Teknillinen fysiikka F3005
Sivuaine:	Lääketieteellinen tekniikka F3001
Opetusyksikkö:	Tfy-105 Laskennallinen fysiikka
Työn valvoja:	Prof. Tapio Ala-Nissilä **Työn ohjaaja:** FT Emppu Salonen

Molekyylidynamiikkasimulaatiot (MD) ovat nykyään laajalti käytössä biomolekyylien mallintamisessa, koska ne pystyvät antamaan tietoa niistä biologisten systeemien ominaisuuksista, joita on hankala tutkia kokeellisesti. Laskentakapasiteetin kasvaminen on mahdollistanut yhä monimutkaisempien systeemien simuloimisen, mutta myös luonut sekä tilaisuuden että tarpeen simulaatioiden perustana olevien voimakenttien kehittämiseen.

Tällä hetkellä sähköisiä vuorovaikutuksia mallinnetaan käytetyimmissä MD-voimakentissä pistevarauksilla. Nämä pistevaraukset eivät kuitenkaan pysty kuvaamaan oikein varausjakauman riippuvuutta systeemin tilasta, eivätkä ne pysty reagoimaan molekyylien liikkeestä johtuvaan sähkökentän vahteluun. Tämä voitaisiin saavuttaa lisäämällä voimakenttään erillinen kuvaus polarisoituvuudelle.

Tässä työssä on tutkittu miten polarisoituvan voimakentän sähköiset vuorovaikutukset tulisi parametrisoida. Tutkimuksessa yhdistettiin kolme erilaista menetelmää sovittaa pistevarauksia, MK, CHELPG ja RESP, ja kaksi multipolialgoritmia, DMA ja GMM, molekyylien polarisaatioita kuvaavaan Tholen malliin. Tämä tehtiin, jotta molekyylin sisäisen polarisoituvuuden osuus voitaisiin poistaa varausten/multipolien sovitusprosessista, ja nämä sähköiset termit esittää voimakentässä erikseen.

MK ja RESP todettiin sopivimmiksi menetelmiksi voimakenttien parametrisointiin. Ne tarjoavat hyvän kompromissin tarkkuuden ja tehokkuuden välillä, ja ovat suhteellisen helppoja soveltaa voimakenttiin. Yllättävä tulos oli se, että hyvin korkean asteen DMA-multipoleja tarvittiin, jotta päästiin näiden varausmenetelmien kanssa samaan tarkkuuteen. GMMn soveltuvuuden parametrisointiin vaarantavat suppenemisongelmat, joita kohdattiin kun GMM yhdistettiin Tholen malliin. Lisäksi GMM-multipolien sähköisten vuorovaikutuksien funktionaaliset muodot ovat hankalia voimakenttäsovelluksen kannalta.

Päivämäärä: 15.8.2011 **Kieli:** Englanti **Sivuja:** 70 + 28

Avainsanat: Polarisoituvuus, Voimakenttä, Parametrisointi, Multipolihajotelma

Preface

This Master's thesis was made in the Computational Soft Matter Group of the Aalto University School of Science and Technology. My greatest gratitude goes to my instructor Emppu Salonen for providing me the opportunity to work in his group and particularly on this project. What is more, he gave me the chance to attend the Biophysical Society meeting in Baltimore, which was truly a great learning experience. I thank him for his endless patience and encouragement which followed my frequent use of words "I don't understand".

I wish to thank Tapio Ala-Nissilä for acting as my supervisor. He guided me through the neverending bureaucracy one needs to go through in order to graduate.

Many thanks go to my roommates at work for the numerous good tips and many exhilarating conversations concerning research and non-research. I specifically thank you for the tea breaks.

I'm very grateful for all the friends I have made during my studies in Helsinki University of Technology. You were always there for me whether I needed encouragement, guidance or just to copy your homework. A special thank goes to Laura and Päivi, for all the lunch breaks and moral support during the time I have been working on my thesis.

Finally, I would like to give my most sincere thanks to my family. My brother Kari, for being the role model and the source of inspiration behind my venture into physics. My father Matti, for teaching me the value of common sense and hard work.

Otaniemi, 3.8.2011

Hanne Antila

Contents

Nomenclature

Abbreviations

MK	Merz-Kollman
CHELPG	Charges from Electrostatic Potentials using a Grid based method
RESP	Restrained electrostatic potential
vdW	van der Waals
ESP	Electrostatic potential
MD	Molecular dynamics
DMA	Distributed multipole analysis
GMM	Gaussian multipole model
QM	Quantum mechanical
RMS	Root mean square
RRMS	Relative root mean square

Symbols

e	Elementary charge	1.602176×10^{-9} C
a_0	Bohr radius	0.52918 Å
q	Point charge	e
V	Electrostatic potential	e/a_0
U	Potential energy	kJ/mol

U_{ind}	Induction energy	kJ/mol
$\boldsymbol{E}$	Electric field	$\mathrm{E_h/ea_0}$
$\boldsymbol{E^0}$	Electric field from static charges/multipoles	$\mathrm{E_h/ea_0}$
α	Polarizability	$\mathrm{a_0^3}$
$\boldsymbol{\alpha}_{\mathrm{mol}}$	Molecular polarizability tensor	$\mathrm{a_0^3}$
α_{C}	Elemental polarizability parameter of Thole model (for carbon)	$\mathrm{a_0^3}$
$\boldsymbol{\mu}$	Dipole moment	$\mathrm{ea_0^3}$
ρ	Charge density/distribution	-
$\boldsymbol{T}$	Interaction tensor	-
Q_{lm}	Multipole moment	-
R_{lm}	Regular solid harmonic function	-
Z_{eff}	Effective nuclear charge	e
λ	Exponent scaling parameter for GMM	$\mathrm{a_0^{-1}}$
$\lambda_3, \lambda_5, \lambda_7$	Thole damping factors for interaction tensor	-

List of Figures

List of Tables

Chapter 1

Introduction

Molecular dynamics (MD) simulations are nowadays widely in use in materials sciences and in the modelling of biomolecules. In molecular dynamic simulation the goal is to examine the motion of particles in a system over a time period. This information can be combined to obtain thermodynamic data, and eventually used to determine the relationships between molecular structure, movement, and function. To do this, one builds a model where atoms and molecules are allowed to interact by approximations of known physics. The result is a tool at the interface of experimental work and theory.

As a tool, molecular dynamics is highly interdisciplinary since its theories stem from chemistry, physics, and mathematics, and it employs algorithms from computer science. However, the roots of the methods used in the molecular modelling lie in rising of modern physics in the beginning of the 20th century. For example, the first successful representation of a molecular structure was closely related to the development of nuclear physics [1]. Also, a group of scientists working in Los Alamos published a paper in 1953 titled "Equation of State Calculations by Fast Computing Machines." This work laid the groundwork for computer-based Monte Carlo methods, established the Metropolis algorithm (named after the first author) for simulated annealing, and was the predecessor of molecular dynamics calculations.

The concept of force fields, in relation to molecules, had its beginning in the development of vibrational spectroscopy, which studies the forces between a pair of atoms in a molecule or in a lattice. The idea of force fields did not spread beyond the physical chemistry community until 1946, when it was first suggested to use the concept for modelling molecules in a more quantitative way. The

new method was based on a combination of steric interactions and a Newtonian mechanical model of bond stretching, angle bending, and torsional vibrational modes. All together three research groups proposed their own versions of this method, which would later be know as the empirical force field or molecular mechanics method for modelling molecular structures [2].

The 20th and 21st centuries later saw a huge improvement in computational capacity and in the algorithms used in the complex optimization tasks in molecular simulations. However, the core of molecular dynamics has remained much the same as the force fields, which depict the potential energy in the system, still include mostly the same interactions as in the first half of the 20th century. For over 30 years, many attempts have been made to include the effects of polarization in simulations of molecular systems [3]. Despite these efforts polarizable force fields are still not in general use. This is probably partly due to the increase in computational capacity requirements and in simulation times that can be expected when a new kind of interaction is included in a force field model. In addition, a lot of work is required when implementing polarizability into a force field, as it will lead to the complete re-parametrization of the model.

Especially within the past decade, the development of polarizable force fields has become a topic of intense research. Many of the most commonly used force fields have a polarizable counterpart. Some of them have been developed as extensions of the existing non-polarizable parametrizations, others have included polarization from the first version onwards. The major part of these polarizable force fields are devoted to water models for liquid-phase simulations. The inclusion of polarizability in molecular simulations should increase the overall accuracy of biomolecular modelling, but it is particularly important for non-homogeneous systems. Being able to model the response of a molecule to a varying dielectrics of the environment would potentially make a great difference in studying, for example, RNA folding in an environment of divalent ions or membrane protein folding in a lipid environment [3].

Chapter 2

Theory

2.1 Purpose of study

The purpose of this study is to combine 5 different point charge/multipole assignment algorithms together with Thole's inducible point dipole model in order to determine which method would be the best choice for polarizable force field development. Different approaches will be compared based on how accurately they are able to reproduce the electrostatic potential around a molecule. In addition, it will be tested how much conformational changes of a molecule will affect the magnitude of charges/multipoles assigned with these methods and whether the parameters assigned based on the minimum energy conformation of these molecules can reproduce the electrostatic potential around other conformations.

2.2 Molecular simulations

Computer simulations have become increasingly popular in biology, biophysics, and biochemistry over the past few decades. These computational models are able to provide information on those properties of biological systems which are hard to study by experimental means. The gradual increase of computing power has provided means to study the large and complex data sets that are obtained from experiments, and this has in turn led to the formulation of simulation-friendly models for biomolecular processes. Nowadays, computation based models can complement experimental data and provide not only averaged data but also information about the distribution and time series of the

quantities of interest.

When modelling a biomolecular system, a few choices have to be made. One has to decide which atomic or molecular degrees of freedom are explicitly considered in the model and how the interactions between the components are represented. There are questions on how the degrees of freedom should be sampled and how the spatial boundaries and external forces are taken into account. Also, the time scale and spatial resolution have to be decided before modelling a biomolecular system [4].

One of the main obstacles in biomolecular simulations is the fact that the behaviour of a biomolecular system is governed by statistical mechanics. That is, the system cannot be characterized only by the global energy minimum configuration. Instead, statistical mechanics brings in the concept of entropy, which together with the energy of the system determines the free energy of the system. The state of the system is not characterized by single a configuration, but by an ensemble of systems.

The importance of entropy also makes the modelling of the interactions between atoms and molecules more complicated. This is because the internal energy and entropy effects can work together or against each other in non-bonded interactions. Another difficulty is that the free energy differences between states can be relatively small, and systems generally consist of many atom pairs having mutual interactions contributing to the energy by summation. To reach the desired accuracy in the free energy for the system, the accuracy of the summation terms has to be even higher, and this naturally poses a challenge to the force interaction model [4].

As mentioned above, a biomolecular system is generally characterized by a very large number of degrees of freedom (around $10^4 - 10^6$ is routinely accessible by simulation) [4]. The motion along these degrees of freedom is usually very complex since they show a variety of characteristics from highly harmonic to anharmonic, chaotic and diffusive. What is more, there are correlations over a wide scale in time and space. The potential energy surface of this kind of a system is very complex. Therefore, a great challenge in biomolecular modelling is to develop means to search this complex surface for regions of low energy. A variety of methods are available, each with its own particular advantages and disadvantages.

Even though simulations are becoming more and more important in the study of biological systems, experimental work remains at the core of the field. In fact,

experimental data plays an essential role even in biomolecular modelling as it forms the basis on which the classical force fields (see below) are built. Quantum mechanical (QM) theoretical data alone is not sufficient for building a force field, and there is a vast variety of biomolecular compounds for which force field parameters should be derived. If the force field parameters in the model are even somewhat transferable between atoms or groups of atoms in different molecules, some of this workload of parametrizing can be avoided. In addition, the methodology and force field used in simulations cannot be validated without comparison between simulated and experimental data.

Some problems arise from the important role of experimental data. Almost every experiment involves averaging over time and space or molecules, and therefore, does not contain direct information on all configurations constituting a simulation trajectory. Also, the experimental data is often scarce relative to the vast amount of degrees of freedom available. Hence, there is a conceptual gap between simulation data and experimental data which makes validating the simulation unsure when actually multiple ensembles can produce the same experimental data. In reality, the experimental data can also be of insufficient accuracy in order to validate or discredit some simulation results [4].

2.3 Force Fields

The core of any force field is the potential energy function used to connect the configuration and structure to the energy of the system being simulated. A typical potential for a force field is [5]

$$\begin{aligned} U = &\sum_{bonds} K_b(r-r_0)^2 + \sum_{angles} K_\theta(\theta-\theta_0)^2 + \\ &\sum_{dihedrals}\sum_{n}\frac{V_n}{2}(1-\cos(n\phi-\gamma)) + \\ &\sum_{nonbonded\,pairs}\left\{4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}-\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]+\frac{q_iq_j}{4\pi\epsilon_0 r_{ij}}\right\}, \end{aligned} \tag{2.1}$$

where r is the bond length, with force constant K_b and equilibrium bond length r_0. The bond angle is denoted as θ, with force constant K_θ, and equilibrium angle θ_0. There is also the dihedral angle ϕ, with force constant V_n and equilibrium angles γ. The last part of the formula is the familiar Coulomb interaction

between charged particles. The second to last part is called the Lennard-Jones interaction

$$U_{\mathrm{LJ}} = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{2.2}$$

in which ϵ_{ij} and σ_{ij} are the parameters depicting the energy and distance scale of the interaction, and r_{ij} is the distance between non-bonded atoms i and j. The r^{-12} part is the short range repulsive interaction, and the long range attraction is described by the term proportional to r^{-6}. The attractive part has the same distance dependence as dipole-dipole London dispersion energy, which for two particles with polarizability α is proportional to $-\alpha^2/r^6$. The Lennard-Jones parameters are not typically assigned using known values of α, but this interaction is one way in which polarizability, in an averaged sense, is included in the model.

The form of potential function described in eq. (2.1) is common for majority of the force fields currently in use, including CHARMM [6, 7], AMBER [8], GROMOS [9], and OPLS [10], among others. That said, there are force fields which use alternative or additional terms for eq. (2.1). These terms include, for example, higher order terms to treat the bond and valence angle terms and/or cross terms between the bonds and valence angles or valence angles and dihedrals. One purpose of these additional terms is to increase the ability of the force field to reproduce conformational energies far away from the minimum conformation [11]. Alternative forms for the van der Waals (vdW) interaction, described with the Lennard-Jones potential in eq. (2.1), have been implemented in some force fields. One of these alternatives is the Buckingham potential which replaces the repulsion term in Lennard-Jones with the more realistic exponential term to describe the repulsion associated with the Pauli exclusion principle.

The potential function alone does not make a force field. Instead, it is the combination of the potential function and the parameters of that function that can be called a force field. The search for these parameters, that is, the parametrization of a force field often begins with quantum mechanical *ab initio* calculations for small molecules. These *ab initio* calculations include the optimization of the molecular structure, calculation of partial charges, and conformational energy calculations among others. Also experimental spectroscopy data can be utilised to find out properties like the force constants for bonds. Usually adjustments are made to the *ab initio* results in order to reproduce target data of condensed

phases. Often special attention is paid so that solute-solute, solvent-solvent, and solute-solvent interactions correspond to those observed experimentally. After the small molecule results are satisfactory, the parameters are tested by formulating the potential function suitable for simulation of a larger assembly (for example a lipid bilayer), and again simulations are carried out to compare with appropriate target data [12].

There are few more things that should be considered when parametrizing a force field. For example, it is necessary realize that the *ab initio* calculations are based on the gas phase QM wave function, and the result may not be consistent with the condensed phase. Also, the correlation among parameters both makes the parametrization process a complicated task and limits the applicability of parameters. For example, The Lennard-Jones parameters are highly correlated with the partial atomic charges, which means that the Lennard-Jones parameters determined for a given set of charges are typically not appropriate for charges determined using different methodology. What is more, the energy surface of conformational rotation, typically dominated by the dihedral term, will also contain contributions from the electrostatics and the Lennard-Jones term [11].

2.4 Polarizability in molecular simulations

2.4.1 Polarizability

Polarization means the redistribution of the electron density of a particle in space due to an electric field. This electric field can be applied in experiment, or may be due to the molecular environment. There are three different mechanism for polarizability (Fig. 2.1): 1. *Electronic polarizability* is caused by a redistribution of electrons over the atom, or atoms in a molecule. 2. *Geometric polarizability*, which is due to changes in the molecular geometry. 3. *Orientation polarizability*, which is caused by a realignment of a molecule by an electric field [13].

In terms of molecular interactions, polarization leads to non-additivity, since a molecule polarized by another molecule will interact differently with a third molecule than it would if it was not polarized. Hence, polarization has a significant effect to the energetics of a molecular system (estimated to be around 10-20% of total interaction energy at the van der Waals minimum distance [3]).

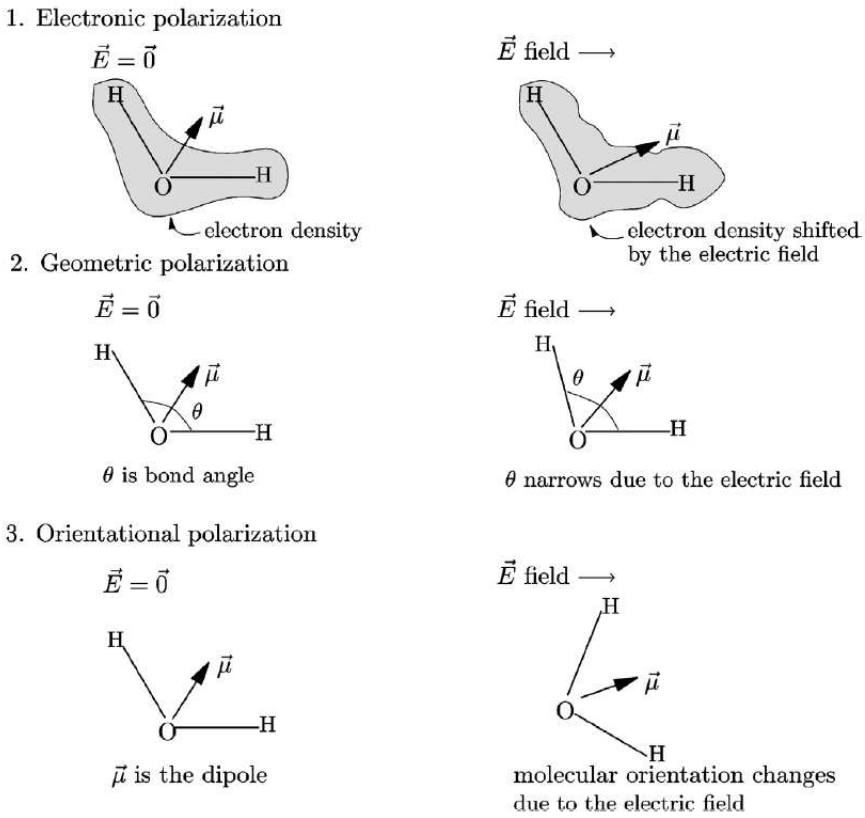


Figure 2.1: Three molecular polarization mechanisms illustrated for a water molecule [13].

2.4.2 Polarizable force fields

Currently, the majority of force fields in general use treat the electrostatic interactions using the Coulomb interaction and partial charges (eq. (2.1)). This means that most computer simulation studies of biomolecular systems do not treat polarizability explicitly. Instead, polarizability is implicitly included by choosing the partial charges so that they are enhanced from the values that would be consistent with the gas-phase dipole moment, or the values that would best reproduce the electrostatic potential from gas phase *ab initio* calculations [5]. This overestimation is designed to approximate electrostatic interactions that occur in the aqueous, condensed phase environment common to biomolecules. The downside in the effective partial charge method is that these charges can not accurately reflect the dependence of the charge distribution on the state of the system, nor can they respond dynamically to fluctuations in the electric field due to molecular motion.

Particularly, partial charges are (once assigned to the molecule) constant under conformational changes of the molecule and cannot alone correctly model the dependency of electrostatics on the geometry of the molecule. That is, a single set of fixed charges or multipoles is generally not applicable to the variety of conformations present in a flexible biomolecule. One possible solution to the problem could be the addition of polarizable potential, which is dependent on the local geometry and captures the correct intramolecular polarization behaviour in terms of electrostatic potential and energy [14].

As an example of the deficiency of current force fields one can mention the work by Rasmussen *et al.* [15]. They showed that conventional force fields are not able to predict the conformational energies of molecules correctly. In fact, the more polar the molecule is, the larger the error becomes. Rasmussen *et al.* were able to improve the correlation between force field and *ab initio* calculation results by inclusion of polarizability into the simulation. That said, their results also indicated that addition of higher permanent multipole moments is equally important and one should also consider including them in the force field development at the same stage.

Against this background it is easy to see that the explicit inclusion of polarizability will be the next major step in improving the current biomolecular force fields. The energy of induced dipoles can be divided into three parts [5]

$$U_{\text{ind}} = U_{\text{stat}} + U_{\mu\mu} + U_{\text{pol}} \,. \tag{2.3}$$

U_{stat} is the interaction energy of N induced dipoles $\boldsymbol{\mu_i}$ in a static electric field $\boldsymbol{E^0}$. The $U_{\mu\mu}$ is the interaction energy between induced dipoles

$$U_{\mu\mu} = \sum_{i=1}^{N} \sum_{i \neq j} \boldsymbol{\mu_i} \boldsymbol{T_{ij}} \boldsymbol{\mu_j} \,, \tag{2.4}$$

where $\boldsymbol{T_{ij}}$ is the interaction tensor (see below) and where $\boldsymbol{\mu_i}$ is the induced dipole moment of atom i. The energy required to distort the electron distribution and create the dipole reads

$$U_{\text{pol}} = \frac{1}{2} \sum_{i} \boldsymbol{\mu_i} \cdot \boldsymbol{E_i} \,, \tag{2.5}$$

where $\boldsymbol{E_i}$ is the electric field at the location of atom i. Combining the three en-

ergy terms gives

$$U_{\text{ind}} = \sum_{i=1}^{N} \boldsymbol{\mu_i} \left[-\boldsymbol{E_i^0} + \frac{1}{2} \sum_{i \neq j} \boldsymbol{T_{ij}} \boldsymbol{\mu_j} + \frac{1}{2} \boldsymbol{E_i} \right]. \tag{2.6}$$

The total electric field at i can be presented as a combination of the field from induced dipoles and the static field $\boldsymbol{E^0}$ resulting from the permanent charges (or even multipoles) in the system

$$\boldsymbol{E_i} = \boldsymbol{E_i^0} - \sum_{i \neq j} \boldsymbol{T_{ij}} \boldsymbol{\mu_j} \tag{2.7}$$

and using this eq. (2.6) can be simplified to

$$U_{\text{ind}} = -\frac{1}{2} \sum_{i=1}^{N} \boldsymbol{\mu_i} \cdot \boldsymbol{E_i^0}. \tag{2.8}$$

The addition of the polarizability contribution will lead to the complete reparametrization of the force field because polarizability is closely connected to the partial charges assigned to the molecule and, as mentioned above, partial charges are correlated with the rest of the parameters in the force field.

To date, majority of work on the polarizable force fields has been concentrating on water models. These water models have given encouraging results by accurately treating both gas and condensed phase properties [11]. However, development in the field of biomacromolecules has been more limited. This is probably partly due to the increase in computational capacity requirements and simulation times that can be expected when a new kind of interaction is included in a force field model. Also, the parametrization of a force field is very time consuming task in itself, and the addition of polarizability makes the problem even more complicated. For example, it is not clear how gas-phase molecular polarizabilities should be treated when used in the parametrization of condensed phase polarizable force field, but it is believed that directly applying the gas phase polarizabilities would cause a tendency towards overpolarization in condensed phase simulations [11].

Polarizable force fields hold great potential to increase the overall accuracy of biomolecular simulations, and they have been speculated to make a great difference in in studying, for example, RNA folding in an environment of divalent

ions [3]. Specifically, polarizable force fields may prove to be essential for studying the electrical properties of lipid membranes, and all the membrane functions related to those properties, such as membrane protein folding [3, 12, 16]. Unfortunately, it is only after a highly refined force field with polarizability included is developed, that one is able to compare the accuracy and applicability of such models compared to their non-polarizable, additive counterparts [11].

Polarization models currently in use in force field development can be divided into three categories: 1) point dipole models 2) shell model a.k.a. Drude model 3) electronegativity equalization model. Each type has its distinct advantages and disadvantages which will be addressed more in depth below.

2.4.3 Models for polarizability

Point dipole model

One way to account for polarizability in molecular models is the point dipole method, which has been applied to a wide variety of systems ranging from noble gases to proteins. In this method one adds ideal point dipoles to selected sites in a molecule, most commonly to the atomic sites.

In the most general case, all the point dipoles assigned to a molecule will interact through the dipole field tensor $\boldsymbol{T}_{ij}$ (see below). One of the first point dipole methods by Applequist *et al.* [17] uses this approach for calculating molecular polarizabilities. The downside of the method is that coupling all the dipoles can lead to a polarization catastrophe. This means that the molecular polarization, and therefore also the induced dipole moment, may become infinite at short distances. This is due to the fact that when two inducible dipoles come spatially too close to each other, the dipolar interaction between them will mutually enhance their induced dipoles to infinite magnitudes. The polarization catastrophe can be avoided by screening the dipole-dipole interaction at short distances as in the point dipole model by Thole [18]. The screening can be physically justified by the fact that the electronic distribution is not well represented by point charges and point dipoles at short distances. Among other things, the electronic distributions change shape when atoms come close enough to each other.

A good feature in point dipole models is that the assignment of electrostatic potential parameters is more straightforward than for non-polarizable models. Charges can be assigned, for example, based on experimental dipole moments or *ab initio* electrostatic potential [5].

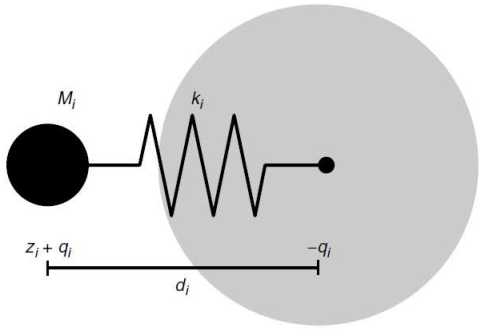


Figure 2.2: The shell model. A core charge $z_i + q_i$ is attached by a harmonic spring with spring constant k_i to a shell charge $-q_i$. For a neutral atom $z_i = 0$. The center of mass is at or near the core charge, but the short-range interactions (*e.g.* the Lennard-Jones interactions) are centered on the shell charge [5].

Shell model

As opposed to the point dipole models, the shell models depict polarization by using dipoles of finite size. In the shell model each polarizable unit is represented by a pair of point charges separated by a variable distance. These charges consist of a positive core charge located at the site of the nucleus and a negative shell charge (Fig. 2.2) connected to the positive core by a harmonic spring. Although these charges can to some extent be interpreted as an effective, shielded nuclear charge and a corresponding valence shell charge, the charges are typically treated more as adjustable parameters for the model than true shielded values. The magnitudes of the shell and core charges are fixed. Hence, polarization in this model is due to the relative displacement of the charges.

The electrostatic interaction between different atoms is simply the sum of charge-charge interactions between the four charge sites (two shell-core pairs). The advantage is that no new interaction types, such as the dipole tensor T_{ij} in the case of the point dipole, are required. The computational advantage is nevertheless nullified in practice by the fact that one has to calculate four times as many charge-charge interactions.

In a way, the point dipole model is an idealized version of the shell model, and it can be argued that the shell model is more physically realistic with its finite length dipoles. That said, both models include additional approximations, that may have more influence on the results than ignoring the finite electronic displacement in polarization. Among these approximations are the assumption of isotropic electrostatic polarizability (in the shell model) and the assumption that

the electrostatic interactions can be truncated after the dipole-dipole interaction, not including the multipole moments [5].

Electronegativity equalization

Polarizability can also be included into standard potentials by allowing the values of the partial charges to respond to the electric field of their environment. Again, this method introduces polarizability without any additional interaction types, and unlike the shell model, this can be done without the additional charge-charge interactions [5]. However, this model does need a shorter time step in molecular dynamics (MD) simulations, and this leads to additional computational cost [19]. In addition, electronegativity equalization model does not reproduce off-plane polarization for plane-like structures like aromatic rings.

The instantaneous values of the partial charges are solved by minimizing the electrostatic energy of the system. In the equation for electrostatic energy, the so-called Mulliken electronegativity and absolute atomic hardness are optimised to reproduce molecular dipoles, interactions with water and the molecular polarization response, typically determined from QM calculations [19]. The energy minimization process can be portrayed as charge flow between atomic sites. Charge neutrality can be introduced into this model in two ways: a charge neutrality constraint can be applied to the entire system, allowing charge to flow from atomic site to atomic site until the electronegativities are equal on all the atoms of the system. Alternatively, charge can be constrained independently on each molecule (or part of a molecule), so that charge flows only between atoms in the same molecule until the electronegativities are equalized within the molecule. In most cases, the latter method is preferred, and there is no charge transfer between molecules. Some models only allow charge transfer along bonded atoms, which guarantees the charge conservation in each set of bonded atoms. However, sometimes charge transfer is an essential part of interaction energy, and this constraint has to be removed [5].

2.4.4 The Thole model

As mentioned above, Thole's model [18] belongs to the category of polarizable point dipole models. The model is based on the work of Silberstein [20] and Applequist [17], but the difference between Thole's model and the preceding work is the modified dipole interaction tensor, which in Thole's model is used to avoid

the polarization catastrophe.

In point dipole models, the molecule is considered an arrangement of N atoms each of which has a polarizability. The induced dipole moment at atom p, $\boldsymbol{\mu_p}$, can be calculated as function of the applied electric field $\boldsymbol{E_p^0}$

$$\boldsymbol{\mu_p} = \boldsymbol{\alpha_p} \left[\boldsymbol{E_p^0} - \sum_{q \neq p}^{N} \boldsymbol{T_{pq}} \boldsymbol{\mu_q} \right] , \tag{2.9}$$

where $\boldsymbol{\alpha_p}$ is the atomic polarizability tensor of atom p and $\boldsymbol{T_{pq}}$ is the dipole field tensor

$$\boldsymbol{T_{pq}} = r_{pq}^{-3} \boldsymbol{I} - 3(r_{pq}^{-5}) \begin{bmatrix} x^2 & xy & xz \\ yx & y^2 & yz \\ zx & zy & z^2 \end{bmatrix} . \tag{2.10}$$

Here $\boldsymbol{I}$ is the unit tensor, r_{pq} is the distance between atoms p an q, and x, y, and z are the cartesian components of the vector connecting atoms p and q. Equation (2.9) can be rearranged to a matrix equation

$$\tilde{\boldsymbol{A}} \tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{E}} , \tag{2.11}$$

where $\tilde{\boldsymbol{A}}$ is a $3N \times 3N$ matrix containing the inverse of the atom polarizability tensors along the 3×3 diagonal. $\tilde{\boldsymbol{E}}$ and $\tilde{\boldsymbol{\mu}}$ are $3N \times 1$ vectors where dipole moments $\boldsymbol{\mu_i}$ and electric fields $\boldsymbol{E_i}$ at each atom site i are placed one after another in the corresponding order as in $\tilde{\boldsymbol{A}}$. That is

$$\begin{bmatrix} \boldsymbol{\alpha_1^{-1}} & \boldsymbol{T_{12}} & \dots & \boldsymbol{T_{1N}} \\ \boldsymbol{T_{21}} & \boldsymbol{\alpha_2^{-1}} & \dots & \boldsymbol{T_{2N}} \\ \vdots & & \ddots & \vdots \\ \boldsymbol{T_{N1}} & \boldsymbol{T_{N2}} & \dots & \boldsymbol{\alpha_N^{-1}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\mu_1} \\ \boldsymbol{\mu_2} \\ \vdots \\ \boldsymbol{\mu_N} \end{bmatrix} = \begin{bmatrix} \boldsymbol{E_1} \\ \boldsymbol{E_2} \\ \vdots \\ \boldsymbol{E_N} \end{bmatrix} \tag{2.12}$$

Inverting $\tilde{A}$ results in

$$\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{B}} \tilde{\boldsymbol{E}} \tag{2.13}$$

$$\tilde{\boldsymbol{B}} = \tilde{\boldsymbol{A}}^{-1} = (\tilde{\boldsymbol{\alpha}}^{-1} + \tilde{\boldsymbol{T}})^{-1} . \tag{2.14}$$

The molecular polarizability is obtained by contracting the tensor $\tilde{\boldsymbol{B}}$ to a 3×3

tensor $\boldsymbol{\alpha}_{mol}$:

$$\boldsymbol{\mu}_{\mathbf{mol}} = \left[\sum_{p}^{N} \sum_{q}^{N} \boldsymbol{B}_{pq} \right] \boldsymbol{E} = \boldsymbol{\alpha}_{\mathbf{mol}} \boldsymbol{E} \,. \tag{2.15}$$

The three eigenvalues of $\boldsymbol{\alpha}_{\mathbf{mol}}$ then depict the xx, yy, and zz components of polarizability.

Thole contributed to the point dipole model by modifying the dipole interaction tensor as follows

$$\begin{aligned} (\boldsymbol{T}_{pq})_{ij} &= \delta_{ij} r^{-3} - 3 x_i x_j r^{-5} = (\alpha_p \alpha_q)^{-1/2} (\delta_{ij} u^{-3} - 3 u_i u_j u^{-5}) \\ &= (\alpha_p \alpha_q)^{-1/2} t_{ij}(\boldsymbol{u}) , \end{aligned} \tag{2.16}$$

where $\boldsymbol{u} = x/(\alpha_p \alpha_q)^{1/6}$ and δ_{ij} is the Kronecker delta. The most important part in this equation is the shape function of the interaction, t. This shape function does not depend on the atoms p and q, and it is based on some well behaved model of charge (electron) distribution around the cores of atoms. At this stage, Thole also replaced the polarizability tensors with an isotropic polarizability parameter (α_p) for each element.

Thole originally investigated many different forms for the charge distributions. Two of these, the linear and the exponential distributions, have been considered the most appropriate. The linear form of charge distribution is

$$\rho(u) = \begin{cases} \frac{3}{\pi} \frac{(a-u)}{a^4} & u < a \\ 0 & u \geq a \end{cases} . \tag{2.17}$$

For the exponential, the expression is

$$\rho(u) = (a^3/8\pi) e^{-au} \,. \tag{2.18}$$

The corresponding forms of the shape function are

$$\begin{aligned} t_{ij} &= (4a^3 - 3u^4)\delta_{ij}/a^4 - 3u^4 u_i u_j/(a^4 u) && u < a \\ t_{ij} &= \delta_{ij}/u^3 - u_i u_j/u^5 && u \geq a \end{aligned} \tag{2.19}$$

for the linear, and

$$\begin{aligned}\rho(u) &= \delta_{ij}/u^3[1-(a^2u^2/2+au+1)e^{-au}]- \\ &\quad 3u_i r_j/r^5[1-(a^3r^3)/6+a^2r^2/2+ar+1)e^{-au}]\end{aligned} \tag{2.20}$$

for the exponential charge distribution.

After modifying the dipole interaction tensor, Thole found the parameter a and element polarizabilities α_p by fitting the model into a set of 15 experimental molecular polarizabilities. The results were then tested by calculating polarizabilities for molecules not included in the learning set by using the optimized parameters and comparing the calculated values to experimental ones. Based on his fitting results Thole claimed that one needs only one polarizability per element. That is, Thole's polarizability for each atom in the molecule is independent of the chemical environment and hence, these polarizabilities should be well transferable. Later van Duijnen *et al.* [21] improved the fit by fitting the parameters in the model into a learning set of 52 molecules. This extended set also included molecules containing F, S, Cl, Br, and I so in addition to improving the old fit, Thole polarizabilities for these new elements were also determined.

Although both Thole and van Duijnen *et al.* solved the parameters by fitting into experimental polarizability data, this is not the only way of parametrizing the model. Often experimental data can be tricky to find and it may be hard to determine it's accuracy. Kaminski *et al.* [22] have demonstrated that one can assign both polarizabilities and charges in the molecular system by utilizing one-, two-, and three-body energies between molecules.

2.5 Electrostatic potential of a molecule

One on the fundamental problems in MD simulations is how to accurately and efficiently represent the charge distribution and electrostatic potential around a molecule. Several approaches have been suggested over the years. The most simple method is to calculate the electrostatic potential (ESP) on a grid around the molecule and reproduce this potential by fitting effective charges to chosen sites on a molecule. This idea is the core of Merz-Kollman (MK) [23], Charges from Electrostatic Potentials using a Grid (CHELPG) [24], and restrained electrostatic potential (RESP) [25] methods for charge fitting.

There are also methods based on estimating the partial charges form the elec-

tron density associated with an atom in a molecule. The most familiar of these is the Mulliken population analysis [26]. The weakness of the Mulliken population analysis is the basis set dependence and the fact that it is only meaningful if the basis set consists of basis functions that can be associated with an atomic site. With very complete basis sets, the Mulliken charges tend to become unphysically large. Natural population analysis [27] is a successor of Mullken method. It is based on the orthonormal natural atomic orbitals of the atoms in a molecule, and it resolves the many of the basis set related problem encountered in the Mulliken population analysis [28].

More advanced methods for reproducing electrostatics around a molecule include also higher multipole moments in addition to charges. The distributed multipole analysis (DMA) by Stone [29] utilizes the Gaussian form of computed wave functions of the molecule and calculates the multipole expansion directly form the charge distribution based on those wave functions. Recently Elking *et al.* have introduced their own multipole method called the Gaussian multipole model (GMM) [30] where a single Slater-type contracted Gaussian multipole charge density is assigned to each atom in a molecule and the Gaussian multipoles are fitted to the ESP around the molecule.

2.5.1 Merz-Kollman (MK)

The main difference between MK [23, 31] and CHELPG [24] methods is the way they generate the grid for the ESP approximation. In MK the grid of choice consists of several over-layered spheres (Connolly surfaces) around a molecule. These spheres are essentially scaled van der waals surfaces of the atoms in the molecule. The smallest scaling factor is 1.4 and conventionally three more surfaces with scaling factors 1.6, 1.8, and 2.0 are added (fig. 2.3). In the original paper by Sing and Kolmann [23] it was claimed that the resulting charges are actually quite insensitive to the choice of scaling factors and the smallest scaling was chosen to 1.4 to be sure that no artefacts from being too close to the atoms entered the fitting of the point charge.

The first version on the Merz-Kollman procedure [23] used non-linear fitting method which allowed the places of charges also outside atomic centres of the molecules to be optimized during the fitting. However, the advantages of adding off-center charges were not conclusive. In the second version of the algorithm [31] the fitting algorithm was updated to linear least squares procedure in which

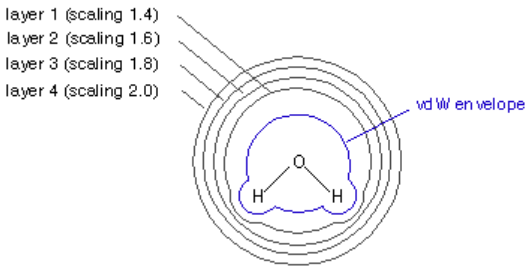


Figure 2.3: Illustration of point selection in the Merz-Kollman algorithm

the constraints are implemented via Lagrange multipliers. This update removed the need for iterative solution procedure and initial guess charges, but also made the optimization of off-center charge locations impossible.

In the least squares fitting the goal is to find the minimum of equation [31]

$$\gamma(q_1, q_2, \ldots, q_n) = \sum_{i=1}^{m} (V_{QM,i} - V_i)^2 \,, \tag{2.21}$$

where $V_{QM,i}$ is the quantum mechanically calculated ESP at point i and V_i is calculated from the charges q_j placed at the molecule

$$V_i = \sum_{j=1}^{n} \frac{q_j}{r_{ij}} \,. \tag{2.22}$$

Constraints (g) are placed in the fitting via Lagrange multipliers (λ). In order to keep the fitting linear, these constraints have to be functions of the charge values. The function to be minimized with w different constraints is

$$z = \gamma + \lambda_1 g_1 + \lambda_2 g_2 + \ldots + \lambda_w g_w \,. \tag{2.23}$$

The minimum of z can now be found by solving

$$\sum_{k=1}^{n} \frac{\partial z}{\partial q_k} = 0 \text{ and } \sum_{l=1}^{w} \frac{\partial z}{\partial \lambda_l} = 0 \,. \tag{2.24}$$

The most common constraint used is that the sum of the charges must equal the total charge of the molecule

$$g_1 = \sum_{j=1}^{n} q_j - q_{\text{tot}} = 0 \,. \tag{2.25}$$

For this constraint equations 2.23 and 2.24 give

$$\begin{aligned} z &= \sum_{i=1}^{m}\left(V_{QM,i}-\sum_{j=1}^{n}\frac{q_j}{r_{ij}}\right)^2+\lambda\left(\sum_{j=1}^{n}q_j-q_{\text{tot}}\right) \\ \frac{\partial z}{\partial \lambda} &= \sum_{j=1}^{n}q_j-q_{\text{tot}}=0 \\ \frac{\partial z}{\partial q_k} &= \sum_{i=1}^{m}\frac{2}{r_{ij}}\left(V_{QM,i}-\sum_{j=1}^{n}\frac{q_i}{r_{ij}}\right)+\lambda=0\,. \end{aligned}$$

This simplifies to

$$\begin{aligned} \sum_{j=1}^{n}q_j &= q_{\text{tot}} \\ \sum_{j=1}^{n}\sum_{i=1}^{m}\frac{1}{r_{ij}r_{ik}} &= \sum_{i=1}^{m}\frac{V_{QM,i}}{r_{ik}}+\lambda\,, \end{aligned}$$

where λ, being and arbitrary constant, has been left as $+\lambda$. Now defining

$$A_{jk} = \sum_{i=1}^{m}\frac{1}{r_{ij}r_{ik}} \tag{2.26}$$

$$B_k = \sum_{i=1}^{m}\frac{V_{QM,i}}{r_{ik}} \tag{2.27}$$

the problem can be presented as a matrix equation

$$\begin{bmatrix} A_{11} & A_{12} & \dots & A_{1n} & 1 \\ A_{21} & A_{22} & \dots & A_{2n} & 1 \\ \vdots & \vdots & \ddots & \vdots & 1 \\ A_{n1} & A_{n2} & \dots & A_{nn} & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix}\begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \\ \lambda \end{bmatrix}=\begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_n \\ q_{\text{tot}} \end{bmatrix} \tag{2.28}$$

that is

$$\boldsymbol{Aq}=\boldsymbol{B} \tag{2.29}$$

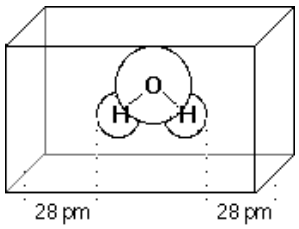


Figure 2.4: Illustration of point selection in the CHELPG algorithm

which can be solved for q

$$\boldsymbol{q} = \boldsymbol{A}^{-1}\boldsymbol{B}\,. \tag{2.30}$$

2.5.2 CHELPG

The CHELPG method follows much the same principles as the MK procedure. In order to make the fitted charges more rotationally invariant, a larger number of fitting points is used in the CHELPG method [24]. Also, the fitting grid is made independent from the molecular coordinate system by choosing a cube of points spaced 0.3-0.8 Å apart containing the molecule and including 2.8 Å of head space on all sides (fig. 2.4). Again, all points that fall inside predefined vdW radius of a nuclei will be discarded from the analysis to avoid artefacts. The least squares fitting described above is also used in the CHELPG scheme.

2.5.3 RESP

ESP charges fitted with MK and CHELPG methods have some weaknesses which were the inspiration behind the development on the RESP method [25]. The developers of RESP wanted to make the ESP fitted charges less conformationally dependent and more transferable between functional groups. Particularly, they wanted to eliminate the large charges which frequently occur in the deeply buried, and hence statistically poorly determined fitting centres in molecules.

In the RESP method [25], a penalty function is included in the least-squares fitting procedure. The objective of this is to hold the charges to a lower magnitude without compromising the quality of the fit to the ESP. With the penalty function, the object function for minimization is now defined by

$$\gamma = \sum_{i=1}^{m}(V_{QM,i} - V_i)^2 + \gamma_{\text{rstr}} = \gamma_{\text{esp}} + \gamma_{\text{rstr}} \tag{2.31}$$

the minimum of least squares fitting being

$$\frac{\partial \gamma}{\partial q_j} = \frac{\partial \gamma_{\text{esp}}}{\partial q_j} + \frac{\partial \gamma_{\text{rstr}}}{\partial q_j} = 0\,. \tag{2.32}$$

The initial choice for penalty function was a simple harmonic

$$\gamma_{\text{rstr}} = a \sum_{j}^{n} (q_0 - q_j)^2\,, \tag{2.33}$$

where a is the parameter for determining the strength of the restraint and q_0 is the target charge. A slightly better-working choice for the penalty function was found to be the hyperbolic

$$\gamma_{\text{rstr}} = a \sum_{j}^{n} \left((q_j^2 + b^2)^{1/2} - b \right)\,, \tag{2.34}$$

where a defines the asymptotic limits of the strength of the restraint and b determines how narrow the hyperbola is around the minimum. The target charge of zero was found to be a good choice.

Again, the problem can be presented as a matrix equation in the form of eq. (2.29). The diagonal elements of $\boldsymbol{A}$ are now given by

$$\boldsymbol{A}_{jj} = \sum_{i}^{m} \frac{1}{r_{ij}^2} + \frac{\partial \gamma_{\text{rstr}}}{\partial q_j}\,, \tag{2.35}$$

and the elements of $\boldsymbol{B}$ are

$$\boldsymbol{B_j} = \sum_{i}^{m} \frac{V_{QM,i}}{r_{ij}} + q_0 \frac{\partial \gamma_{\text{rstr}}}{\partial q_j}\,. \tag{2.36}$$

It is conventional to leave the hydrogens in the molecule unrestrained since they generally are well solvent-exposed and hence well determined in terms of the ESP fit. Usually, the Connolly surface based grid equal to the grid used in MK is the grid of choice to estimate the ESP in this method.

It is sometimes necessary to have identical charges in the nuclei which are equivalent in terms of conformational rotations, *e.g.*, all hydrogens in a methyl group must have the same charge, or otherwise rotations around the bond attaching the methyl group to the rest of the molecule will give rise to three differ-

ent conformational energies instead of resulting in three degenerate rotamers. The forced symmetry can be added to the least squares fit by modifying the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ in eq. (2.29). Preliminary $\boldsymbol{A}$ and $\boldsymbol{B}$ are generated as there were no equivalent charges. Then, the rows and columns of $\boldsymbol{A}$ (and the rows of $\boldsymbol{B}$) for centres to be fitted to the same charge are combined together to form a single row and column of $\boldsymbol{A}$ (and a single row of $\boldsymbol{B}$) giving rise to new, smaller versions of $\boldsymbol{A}$ and $\boldsymbol{B}$, which are solved as usual [25].

2.5.4 Distributed multipole analysis (DMA)

The idea of distributed multipole analysis [29, 32] (DMA) is to represent the charge distribution of a molecule by charges, dipoles, quadrupoles, and so on located at number of distributed points in a molecule. These multipoles are derived from the *ab initio* wave function of the molecule and assigned to sites chosen by the user.

The starting point of DMA is the expansion of the electron density in terms of primitive Gaussian basis functions in the form of

$$\chi_i(\boldsymbol{r}) = N_i x_i^{m_i} y_i^{n_i} z_i^{o_i} \exp[-\xi_i(\boldsymbol{r}_i)^2], \tag{2.37}$$

where $\boldsymbol{r_i} = \boldsymbol{r} - \boldsymbol{A_i}$ is the electron position relative to the position $\boldsymbol{A_i}$ of the primitive Gaussian, ξ_i is the exponent, N_i is the normalizing factor and n_i, m_i, and o_i are integer exponents of cartesian components x_i, y_i, and z_i of $\boldsymbol{r_i}$. The electron density is then

$$\rho(\boldsymbol{r}) = \sum_{ij} D_{ij}\chi_i(\boldsymbol{r})\chi_j(\boldsymbol{r}), \tag{2.38}$$

where coefficients D_{ij} are elements of the density matrix.

It has been shown by Boys [33] that a product of two Gaussians, $\chi_i\chi_j$, can be expressed as a Gaussian function centered at $P_{ij} = (\xi_i\boldsymbol{A_i} + \xi_j\boldsymbol{A_j})/(\xi_j + \xi_i)$. Thus each product of primitive functions gives an overlap charge density centered at its own point $\boldsymbol{P}$. The q component of the rank k multipole moment of this overlap density at site $\boldsymbol{P}$ is then defined in DMA as

$$Q_{kq}(\boldsymbol{P}) = -\int R_{kq}(\boldsymbol{r} - \boldsymbol{P}) D_{ij}\chi_i\chi_j, \tag{2.39}$$

where R_{kq} is a regular solid harmonics.

A multipole expansion about the point $\boldsymbol{P}$ can be presented about any other point $\boldsymbol{S}$ by

$$Q_{lm}(\boldsymbol{S}) = \sum_{k=0}^{l} \sum_{q=-k}^{k} \left[\begin{pmatrix} l+m \\ k+q \end{pmatrix} \begin{pmatrix} l-m \\ k-q \end{pmatrix} \right] \times Q_{kq}(\boldsymbol{P}) R_{l-k,m-q}(\boldsymbol{S}-\boldsymbol{P}). \quad (2.40)$$

In DMA the multipoles at overlap centres $\boldsymbol{P}$ are presented with the help of this formula at one of a smaller number of sites $\boldsymbol{S}$, which are chosen beforehand the by user. This means that in a sense, the multipoles from sites $\boldsymbol{P}$ are moved to $\boldsymbol{S}$ and added together, but one must notice that the numerical value of the multipole changes in the moving process (eq. (2.40)). Usually, the final multipole sites $\boldsymbol{S}$ are chosen to be at the nuclei or at the centres of bonds in a molecule since this choice of sites is the most intuitive and gives a clear picture of the charge distribution. Also, it has been noticed that it gives good convergence of electrostatic potentials and interaction energies calculated from the multipoles [29].

The position of overlap centre $\boldsymbol{P}$ depends on the exponents ξ_i and ξ_j of the two overlapping primitive functions. The site $\boldsymbol{S}$, to which the multipole moments from $\boldsymbol{P}$ are relocated, in turn depends on the position of $\boldsymbol{P}$. In DMA, the multipole from $\boldsymbol{P}$ is always moved to the nearest possible site $\boldsymbol{S}$, or if two or more sites $\boldsymbol{S}$ are at equal distance from $\boldsymbol{P}$, the contribution is equally divided between them [29]. Since the multipole site $\boldsymbol{S}$, where multipole contribution is moved, is chosen based on the location $\boldsymbol{P}$, it follows that the distributed multipole analysis can be very sensitive to the values of the exponents, and hence to the basis set being used [32]. The sensitivity to basis is most severe when large basis sets with diffuse functions are used and the overlap densities extend over many atoms. Unfortunately, such a basis set is essential when one desires to obtain reliable results.

Version 2 of the DMA [32] attempts to make the multipoles less basis set dependent by using a different, grid-based integration method to calculate the multipole contributions from the overlap densities of diffuse functions. In turn the original DMA method is used for the more compact basis functions. The switch between the two methods is determined so that if the sum of exponents $\xi_i + \xi_j$ for a pair of primitive functions is less than the switch value Z the grid-based method is used, otherwise the original approach is utilized. The algorithm the program uses for compact primitives is exact and very efficient Gauss-Hermite quadrature. The grid-based method for the more diffuse cases is very much

slower and, in some cases, less accurate.

2.5.5 Gaussian multipole model (GMM)

Just like the distributed multipole analysis, the Gaussian multipole model [30] assigns multipole moments to atoms in a molecule, but instead of directly utilizing the *ab initio* calculated wave functions like in DMA, the GMM moments are solved by fitting to the ESP around a molecule. The idea in GMM is to use multipole moments which are weighted with Gaussian functions that resemble a wave function. With their recent work with Gaussian multipole model, Elking *et. al.* strived after more accurate dimer electrostatic energies since it has been noted that atomic point multipole models tend to underestimate the electrostatic interactions at dimer distances [30].

The model is composed of a nucleus and a single Slater-type contracted Gaussian multipole charge density on each atom. In the model a Slater type function $\exp(-r)$, which well depicts the actual form of the orbitals, is represented with a linear combination of computationally more convenient Gaussian type functions $\exp(-r^2)$ [28]. The correct contraction coefficients (d_μ) and exponents (a_μ) for the linear expansion are found by fitting

$$\exp(-r) = \sum_{\mu=1}^{N_c} d_\mu \exp(-\alpha_\mu^2 r^2), \tag{2.41}$$

where N_c is the degree of contraction. By adding a scaling factor λ one gets

$$\frac{\lambda^3}{8\pi}\exp(-\lambda r) = \sum_{\mu=1}^{N_c} c_\mu \left(\frac{\alpha_\mu^2\lambda^2}{\pi}\right)^{3/2} \exp(-\alpha_\mu^2\lambda^2 r^2), \tag{2.42}$$

where $c_\mu = d_\mu/8\pi \times (\pi/\alpha_\mu^2)^{3/2}$.

A contracted Gaussian multipole charge density $\rho(\boldsymbol{r},\boldsymbol{R})$ with moments Q_{lm} and nuclear center $\boldsymbol{R}$ evaluated at point $\boldsymbol{r}$ is given by

$$\rho(\boldsymbol{r},\boldsymbol{R}) = \sum_{l=0}^{l_{\max}} \sum_{|m|\leq 0} \frac{Q_{lm} R_{lm}^*(\boldsymbol{r}-\boldsymbol{R})}{(2l-1)!!} \rho_l(|\,\boldsymbol{r}-\boldsymbol{R}\,|;\alpha_\mu), \tag{2.43}$$

where l_{max} is the maximum order of Gaussian multipoles, R_{lm}^* is the complex

conjugate of a regular solid harmonic function, and ρ_l reads

$$\rho(r;\alpha_\mu)=\left(-\frac{1}{r}\frac{d}{dr}\right)^l\sum_{\mu=1}^{Nc}c_\mu\left(\frac{\alpha_\mu^2}{\pi}\right)^{3/2}\exp(-\alpha_\mu^2 r^2)\,. \tag{2.44}$$

For monopoles (l=0) the density is normalised to unity, ($\sum_\mu c_\mu$=1). The multipole moments of the charge density $\rho(\boldsymbol{r},\boldsymbol{R})$ with respect to the center $\boldsymbol{R}$ are defined in the GMM model as

$$Q_{lm}=\int\rho(\boldsymbol{r},\boldsymbol{R})R_{lm}(\boldsymbol{r}-\boldsymbol{R})d^3r\,. \tag{2.45}$$

The electrostatic potential V arising from the contracted Gaussian multipole density in eq. (2.43) is given by

$$V(\boldsymbol{r})_l=\sum_{l=0}^{l_{\max}}\sum_{|m|\le 0}\frac{Q_{lm}R^*_{lm}(\boldsymbol{r}-\boldsymbol{R})}{(2l-1)!!}V_l(|\,\boldsymbol{r}-\boldsymbol{R}\,|;\alpha_\mu)\,, \tag{2.46}$$

where V_l is

$$V_l(r,\alpha_\mu)=\left(-\frac{1}{r}\frac{d}{dr}\right)^l\sum_{\mu=1}^{N_c}c_\mu\frac{\mathrm{erf}(\alpha_\mu r)}{r}\,, \tag{2.47}$$

and the error function erf(x) is

$$\mathrm{erf}(x)=\frac{2}{\pi}\int_0^x\exp(-u^2)\,\mathrm{d}u\,. \tag{2.48}$$

The complete model for molecular charge density ρ^{GM} in GMM consist of effective nuclear charges Z_{eff} in addition to the set of contracted Gaussian multipole moments Q_{lm} with a single Slater exponential scaling parameter λ centered at each atom in the molecule. The effective charges are chosen as 1.0 for H, 4.0 for C, 5.0 for N, 6.0 for O and 7.0 for F, and 7.0 for Cl. The total molecular charge density is presented as sum over atoms in the molecule

$$\rho^{\mathrm{QM}}=\sum_a\sum_{l=0}^{l_{\max}}\sum_{|m|\le 0}\frac{Q_{lm,a}R^*_{lm}(\boldsymbol{r}-\boldsymbol{R_a})}{(2l-1)!!}\rho_l(|\,\boldsymbol{r}-\boldsymbol{R_a}\,|;\lambda_a\alpha_\mu)\,, \tag{2.49}$$

where $\boldsymbol{R_a}$ is the nuclear center of atom a and ρ_l is defined by eq. (2.46). The electrostatic potential due to the effective nuclear charges and the Gaussian multi-

pole charge density is expressed as

$$V(\boldsymbol{r};Q_{lm,a},\lambda_a) = \sum_a \frac{Z_{\mathrm{eff},a}}{|\,\boldsymbol{r}-\boldsymbol{R_a}\,|} + \sum_{l=0}^{l_{\max}} \sum_{|m|\leq 0} \frac{Q_{lm,a} R^*_{lm}(\boldsymbol{r}-\boldsymbol{R_a})}{(2l-1)!!} V_l(|\,\boldsymbol{r}-\boldsymbol{R_a}\,|;\lambda_a\alpha_\mu)\,, \quad (2.50)$$

where $\boldsymbol{r}$ is the point where the field is being evaluated and V_l is defined by eq. (2.47). For each atom, Q_{lm} and λ_a are treated as optimizable parameters and fitted to the *ab initio* electrostatic potential around the molecule. The fitting function is given by

$$\chi^2(Q_{lm,\lambda}) = \int w(r)[V^{\mathrm{GM}}(Q_{lm,\lambda;\boldsymbol{r}}) - V^{\mathrm{QM}}(\boldsymbol{r})]^2 d^3r\,, \quad (2.51)$$

where V^{GM} and V^{QM} are the ESPs calculated from the Gaussian multipoles (eq. (2.50)) and by *ab initio* methods. The fitting function is estimated on a grid of points around the molecule much the same way as in the MK, CHELPG, and RESP approaches, but the optimization is done using a Levenberg-Marquard [34] non-linear least squares algorithm .

The weighting factor $w(\boldsymbol{r})$ in eq. (2.51) serves much the same purpose as the exclusion of points inside the vdW radius in the CHELPG method or the choice of scaling of 1.4 ×vdW for the first point surface in the MK method. The form of $w(\boldsymbol{r})$ is

$$w(\boldsymbol{r}) = \begin{cases} \exp\{-\sigma[\ln\rho^{\mathrm{QM}}(\boldsymbol{r}) - \ln K_0]^2\} & \rho^{\mathrm{QM}}(\boldsymbol{r}) \geq K_0 \\ 1 & \rho^{\mathrm{QM}}(\boldsymbol{r}) \leq K_0 \end{cases}, \quad (2.52)$$

where $\rho^{\mathrm{QM}}(\boldsymbol{r})$ is the *ab initio* electron density. The weighting function $w(\boldsymbol{r})$ is small for regions of high electron density, and gives weighting of $w(\boldsymbol{r})$ =1 for regions of low electron density. There are two adjustable parameters, σ and K_0, which were chosen to be 0.3 and -6 by Elking *et al.* [30].

2.6 Combining Thole's model to models describing the ESP

In conventional MD simulations, molecules cannot respond dynamically to fluctuations of the electric field due to molecular motion. In addition, non-polarizable force fields are not able to model the dependence of electrostatic properties on

the conformation of the molecule. That is, a single set of fixed charges or multipoles is generally not applicable to a variety of conformations.

A possible solution to these problems could be the addition of polarizability. The method of choice for this in this work is the Thole model. The problem is that the intramolecular self-polarization significantly affects the charge distribution and ESP of a molecule, and this has to be taken into account when monopoles or multipoles are assigned to a molecule based on these properties. In other words, the effect of the intramolecular self-polarization has to be separated from the truly permanent electrostatics of a molecule when parametrizing a force field. In this work, two strategies for doing this are utilized: the ΔESP method by Cieplak *et al.* [35] and the analytic method by Ren and Ponder [14].

2.6.1 The ΔESP method

The ΔESP method is an iterative method for separating permanent electrostatics from polarization contributions while parametrizing a force field. First, the ESP around a molecule is calculated by *ab initio* methods (ESP(QM)), then monopoles or multipoles are fitted to the ESP by a method of choice. These are then used to self-polarize the molecule: the electric field from the charges/multipoles is placed in eq. (2.9), and the induced dipoles are calculated either by iteration or by matrix inversion. Finally, the ESP from the induced dipoles (ESP(ind)) is calculated and reduced from the ESP(QM). A new iteration round is started by fitting a new set of charges/multipoles into ΔESP=ESP(QM)-ESP(ind). The iteration can be chosen to terminate either when the fitted charges/multipoles no longer change, or the set of induced dipoles stays constant [35].

2.6.2 The analytic method

The idea behind the analytic method is that the multipoles resulting from a fit to quantum mechanical data can be considered a sum of permanent and induced contributions [14]

$$\boldsymbol{M_i} = \boldsymbol{M_i^p} + \boldsymbol{M_i^{\mathrm{ind}}}, \tag{2.53}$$

where the $\boldsymbol{M}$ denotes a vector of multipoles assigned to atom i in the molecule

$$\boldsymbol{M_i} = [q_i, \mu_{i,x}, \mu_{i,y}, \mu_{i,z}, Q_{i,xx}, Q_{i,xy}, Q_{i,xz}, \ldots, Q_{i,zz}]^T. \tag{2.54}$$

Now eq. (2.9) can be modified to

$$\boldsymbol{M}_i^{\text{ind}} = \alpha_i \left(\sum_{\{j\}} \boldsymbol{T}^{ij} \boldsymbol{M}_j^p + \sum_{\{j'\}} \boldsymbol{T}^{ij'} \boldsymbol{M}_{j'}^{\text{ind}} \right) , \tag{2.55}$$

where sites $\{j\}$ can be chosen as the sites outside the molecule containing site i and $\{j'\}$ as all the sites other than i, for example. The interaction tensor $\boldsymbol{T}$ is now expanded to include interactions for all the ranks of multipoles included in $\boldsymbol{M}$. When eq. (2.53) is substituted to above expression one yields

$$\boldsymbol{M}_i^{\text{ind}} = \alpha_i \left(\sum_{\{j\}} \boldsymbol{T}^{ij} (\boldsymbol{M}_j - \boldsymbol{M}_j^{\text{ind}}) + \sum_{\{j'\}} \boldsymbol{T}^{ij'} \boldsymbol{M}_{j'}^{\text{ind}} \right) . \tag{2.56}$$

If $\{j\}$ and $\{j'\}$ are chosen to be identical group of multipole sites, the expression in eg. (2.56) simplifies to

$$\boldsymbol{M}_{i,\alpha}^{\text{ind}} = \alpha_i \sum_j \boldsymbol{T}^{ij} \boldsymbol{M}_j . \tag{2.57}$$

Equations (2.56) and (2.57) can again be solved for $\boldsymbol{M}^{\text{ind}}$ by iteration (eq. (2.56)) or by direct matrix multiplication (eq. (2.57)). By subtracting the induced moments from the moments resulting from the fit to QM data one is left with the truly permanent multipoles [14].

2.6.3 Intramolecular interactions

Charge-charge interactions between atoms separated by one or two covalent bonds (called 1-2 and 1-3 interactions, see fig. 2.5) in a molecule are neglected in most force fields. There are different approaches to 1-4 interactions, involving atoms separated by 3 bonds, as for example OPLS and AMBER scale the interaction down by a factor, but CHARMM leaves it unscaled [11].

The addition of polarizability contributes to the confusion on how non-bonded interactions should be treated in a force field. It has been noticed that the iterative process on eq. (2.9) is prone to diverge [36, 37] when all the induced dipole-dipole and charge-dipole interactions are included and undamped. Same kind of phenomenon has also been observed developing the ΔESP approach [3]. The divergence has been speculated to originate from the 1-2 and 1-3 interactions by both Cieplak *et al.* [3] and Xie *et al.* [37]. However, their approach to the problem

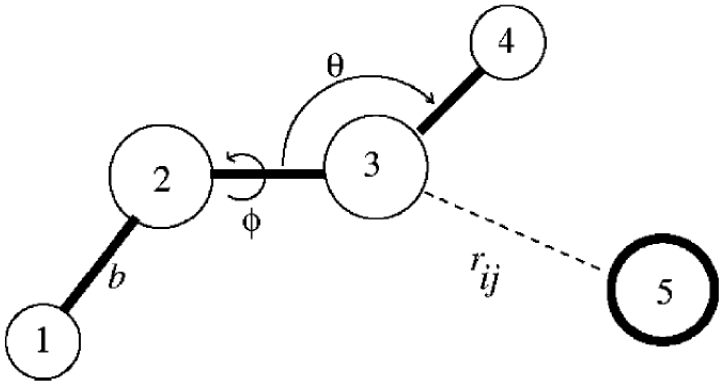


Figure 2.5: Illustration of interactions in a molecule [36].

is quite different. Cieplak *et al.* reparametrized the Thole model by calculating elemental polarizability parameters so that 1-2 and 1-3 interactions were neglected of and 1-4 interactions scaled down, whereas Xie *et al.* developed a more efficient method for solving eq. (2.9). Xie *et al.* also proved that it is essential to include all induced intramolecular interactions in the model in order to correctly describe the molecular polarizability tensor (at least when no re-parametrization of the Thole model is done).

That said, divergence is not the only point of view that has to be considered when choosing how to treat the intramolecular interactions. In their paper [14] describing how to combine the Thole model with DMA using the analytic method (elaborated in section 2.6.2) Ren *et al.* note that although the model works fine with all the intramolecular interactions included, they would prefer a slightly modified, group based 1-2 model. This is because of the large 1-2 direct induction that resulted from the original model. They argue that unrealistic interaction energies may arise if an unphysically large portion of the local electrostatics is considered to be the result of induction, and an ideal model would avoid such unphysical intramolecular polarization.

There is also the option to damp interactions in a way consistent to the Thole model. This approach is utilized, among others, by Ren *et al.* [38] in the paper where they describe the AMOEBA water model, and by Xie *et al.* Following the notation by A. J. Stone [39] the interaction tensor elements for different order

multipoles can be obtained by

$$\begin{aligned} T &= \tfrac{1}{R} \\ T_\alpha &= \nabla_\alpha T = -\tfrac{R_\alpha}{R^3} \\ T_{\alpha\beta} &= \nabla_\alpha T_\beta \\ T_{\alpha\beta\gamma} &= \nabla_\alpha T_{\beta\gamma} \\ &\dots \\ (\alpha,\beta,\gamma, &\dots = 1,2,3), \end{aligned}$$

where R is the distance between interaction sites. Using the different forms of charge distribution given in eqs. (2.17) and (2.18) one can derive the damped interaction tensor elements (T^D). The damped first order element in the direction α reads

$$T_\alpha^{\mathrm{D}} = T\nabla_\alpha\rho + \rho\nabla_\alpha T\,. \tag{2.58}$$

That is

$$T_\alpha^{\mathrm{D}} = -a^{-4}(4au^3 - 3u^4)\frac{R_\alpha}{R^3} \tag{2.59}$$

for the linear distribution ($u \le a$) and

$$T_\alpha^{\mathrm{D}} = -\frac{R_\alpha}{R^3} + \left(\frac{1}{2}a^2u^2 + au + 1\right)e^{-au}\frac{R_\alpha}{R^3} \tag{2.60}$$

for the exponential. The higher order terms are

$$\begin{aligned} T_{\alpha\beta}^{\mathrm{D}} &= \lambda_5\tfrac{3R_\alpha R_\beta}{R^5} - \lambda_3\tfrac{\delta_{\alpha\beta}}{R^3} \\ T_{\alpha\beta\gamma}^{\mathrm{D}} &= -\lambda_7\tfrac{15R_\alpha R_\beta R_\gamma}{R^7} + \lambda_5\tfrac{3(R_\alpha\delta_{\alpha\beta}+R_\beta\delta_{\alpha\gamma}+R_\gamma\delta_{\alpha\beta})}{R^5}, \end{aligned} \tag{2.61}$$

where for the linear case

$$\begin{aligned} \lambda_3 &= a^{-4}(4au^3 - 3u^4) \\ \lambda_5 &= \left(\tfrac{u}{a}\right)^4 \\ \lambda_7 &= \tfrac{1}{5}a^{-4}u^5 \end{aligned}$$

when $u \le a$ and $\lambda_3 = \lambda_5 = \lambda_7 = 1$ when $u > a$. Thole lambdas for the exponential

distribution are

$$\begin{aligned}\lambda_3 &= 1-(\tfrac{1}{2}a^2u^2+au+1)e^{-au}\\ \lambda_5 &= 1-(\tfrac{1}{6}a^3u^3+\tfrac{1}{2}a^2u^2+au+1)e^{-au}\\ \lambda_7 &= 1-(\tfrac{1}{30}+\tfrac{1}{6}a^3u^3+\tfrac{1}{2}a^2u^2+au+1+\tfrac{1}{6}a^3u^3)e^{-au}\,.\end{aligned}$$

Now the row of damped interaction tensor depicting the interaction in direction α reads

$$\boldsymbol{T_\alpha} = [-T^{\mathrm{D}}_{\alpha}, T^{\mathrm{D}}_{\alpha 1}, T^{\mathrm{D}}_{\alpha 2}, T^{\mathrm{D}}_{\alpha 3}, -T^{\mathrm{D}}_{\alpha 11}, -T^{\mathrm{D}}_{\alpha 12}, \dots]\,. \tag{2.62}$$

This notation is consistent with the one presented in eqs. (2.55), (2.56), and (2.57). Thole defined the dipole interaction tensor with different sign (equations (2.9) and (2.10)) but one can easily see that the tensor derivation here produces the same dipole-dipole interaction both in the undamped and damped case.

Chapter 3

Methods

3.1 *Ab initio* calculations with `Gaussian`

The molecular structures used for testing the polarizable and non-polarizable versions of the five different point charge/multipole methods were optimised by using `Gaussian09` [40] and CSC `SOMA2` [41, 42] interface. The optimization was done using six consecutive steps. First, preliminary optimization was done with AM1/STO-3G, B3-LYP/3- 31G*, and B3LYP/6-311++G(d,p) levels of theory. Next, eigenfrequency calculation was performed with B3LYP/6-311++G(d,p) to make sure that an energy minimum was reached. Finally, one more optimization with a more accurate MP2/aug-cc-pVTZ method was conducted to obtain the final optimized structure. The molecules chosen for testing were 1-butene, butane, dimethylethylamine, methyl ethyl ether, methyl formate, propanal, 1-propanol, and propionic acid (fig. 3.1). These particular molecules were chosen because they represent different functional groups commonly found in biomolecules (particularly phospholipids). The molecules also have a dihedral angle in their backbone, but they are still fairly simple molecules with few atoms.

For testing how the results from each method change under conformational variance, one dihedral angle was rotated in each test molecule (fig. 3.1). This was done by using the `Gaussian` "scan" option in MP2/aug-cc-pVTZ level of theory. Rotations where done in 18 steps of 10 degree increment resulting in a total scan of 180 degrees. The angles of the global minimum energy conformation and the second minimun conformation (according to the scan) are presented in table 3.1 accompanied with the *ab initio* dipole moments of the molecules.

For MK, CHELPG, and RESP methods the appropiate potential and grid points

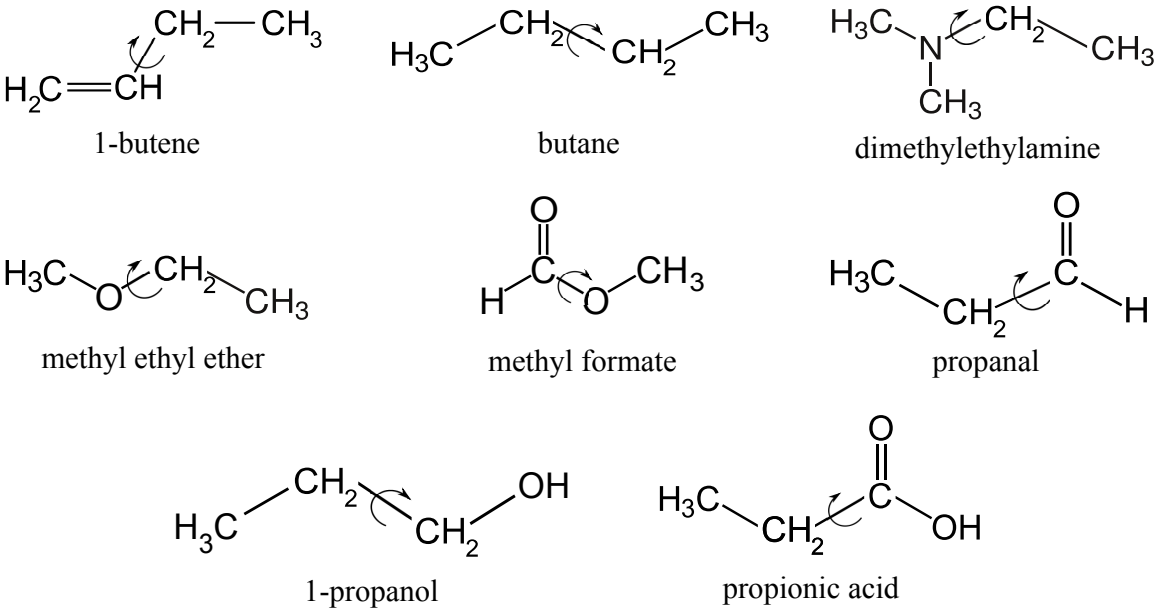


Figure 3.1: Molecules used for testing the performance of the methods. The dihedral angle to be rotated is illustrated by an arrow in each case.

	Global minimum	Second minimum	Dipole moment
1-butene	-118	2	0.17
butane	180	70	0.00
dimethylethylamine	66	166	0.29
methyl ethyl ether	180	-	0.55
methyl formate	0	-	0.79
propanal	0	120	1.24
1-propanol	-180	-60	0.61
propionic acid	0	110	0.73

Table 3.1: The dipole moments and minimum conformations of the test molecules according to the *ab initio* calculations. The dipole moments, calculated with MP2/aug-cc-pvtz, are presented here in atomic units. The global minimum energy conformation and second minimum conformation dihedral angles are in degrees.

were provided by conducting a MK/CHELPG calculation with `Gaussian` and using the undocumented `Gaussian` output option iop(6/33=2) to print out the potential and the locations of the grid points. The calculation of the potential was done in MP2=FC/aug-cc-pVTZ level of theory. The default MK grid was used both for MK and RESP with maximum point density on 6 point per unit area. For CHELPG the default grid was used with point density of 1 point per unit area.

For the formatted checkpoint file needed in DMA and GMM calculations a single point calculation on MP2/aug-cc-pVTZ level of theory was conducted.

3.2 Damping

In this work, damping was the method of choice for treating the intramolecular electrostatic interactions. This makes the interactions consistent with the ones utilised in the parametrization of the Thole model done earlier by the author [43] and used in this work. Hence, we follow the recommendations of Xie *et al.* [37] discussed earlier. In addition, the damping scheme makes the treatment of all multipole interactions consistent with the approach presented by Thole [18] for dipoles, and it is relatively easy to implement in different methods as no information about bonds is needed (as opposed to full exclusion of some interactions). No full exclusions of interactions were used in addition to damping because the parametrization was done without exclusions, and the damping alone was sufficient to prevent the convergence problems for four out of the five methods studied in this work.

3.3 The parametrization of the Thole model

The parameters of Thole model used in this work (polarizabilities α_C, α_O, α_H, α_N, and parameter a) where obtained from the previous work from the author [43]. The values for these parameters can be found in table 3.2. In this work, the exponential version of the Thole model was used (eqs. (2.20) and (2.62)).

The parameters were calculated by an optimization process utilizing experimental molecular polarizabilities together with the optimized molecular geometries for a set of molecules. The learning set used for the fitting was build with care and consisted of 37 molecules. Some common solvent molecules, such as water and cyclohexane, were included in the set to increase the overall performance of the parametrization. Even more importantly, molecules that represent functional groups found in phospholipids were included in the learning set to make sure that the parametrization will be useful for lipid simulation purposes. The fact that experimental polarizability for a molecule is a weighted average over polarizabilities of all the conformations of that molecule was also considered when building the learning set. For this reason molecules with high

	Thole		**van Duijnen**		**Our work**	
	linear	*exponential*	*linear*	*exponential*	*linear*	*exponential*
α_{H}	0.5140	0.4270	0.5189	0.4138	0.3044	0.3128
α_{C}	1.4050	1.2850	1.5079	1.2886	2.0111	1.7669
α_{N}	1.1050	0.9670	1.1269	0.9716	1.7276	1.5389
α_{O}	0.8620	0.7471	0.9475	0.8520	0.7609	0.7405
a	1.6620	2.0890	1.7278	2.1304	2.5416	1.5779

Table 3.2: The values for Thole model parameters α_{H}, α_{C}, α_{O}, α_{N}, and a used in this work. The same parameters optimized by Thole [18] and van Duijnen *et al.* [21] are also presented for comparison.

symmetry, very little rotation around C-C, C-N and C-O bonds, or unambiguous minimum energy conformations were chosen into the learning set.

The optimization of polarizabilities α_C, α_O, α_H, α_N, and parameter a was done with an evolutionary strategy using covariance matrix adaptation. The fitting was performed for all these parameters simultaneously. The performance of the parameters was ensured by building an additional test set of 18 molecules and seeing how well the experimental polarizabilities of these molecules and the molecules in the learning set are reproduced when using the new parameters for the Thole model.

In the previous work by the author [43], the new parametrization used in this work was concluded to be, for the most parts, an improved version compared to the previous sets of parameters presented in the literature [18],[21]. It was also observed that Thole model is usually not able to reproduce the experimental polarizabilities for alkenes as well as for other types of molecules. Although the new parametrization of the Thole model, presented in [43], was able to improve the poor fit of the previous parametrizations by Thole [18] and van Duijnen [21] also in the case alkenes, it was speculated that the parametrization could be further improved by adding a new carbon type for double bonded, sp2 hybridized carbon. Hence, one would part from the original idea by Thole that the isotropic polarizability assigned to each atom of a molecule in the Thole model is independent on the chemical environment of that atom and only depends on the element.

3.4 MK/CHELPG/RESP and ΔESP method

MK, CHELPG, and RESP methods were combined with the Thole model using the ΔESP approach by Cieplak *et al.*. The code for all the point charge methods were written in C by the author, and combined with the C codes for the ΔESP method and the Thole model also written by author. The solving of eq. (2.9) was done by matrix inversion instead of self-consistent iteration in order to minimize the possibility of divergence [37] and eliminate the need of initial guess for the iteration.

Originally, the ΔESP method was used together with the RESP method so that in the intermediate stages the charges were not calculated by RESP but by some other charge fitting method. RESP fitting was done only in the fully iterated potential. As the nature of RESP fitting itself is also iterative, we suspect that this was done to avoid convergence issues and to reduce the computation time. In this work, a full RESP fitting was done also in the intermediate stages of the ΔESP iteration. No convergence issues due to this choice were observed and the computational time increase was fully acceptable with the size of molecules (8-16 atoms) used here. Values of $a = 0.0005$ au and $b = 0.01$ e where used for the hyperbolic restraint (eq. 2.34).

3.5 DMA and the analytical method

Since distributed multipole analysis does not directly use the electrostatic potential in assigning the multipoles to atoms it was not possible to use the ΔESP method to combine it with the Thole model. Therefore the analytical method was chosen, more precisely the simplified approach of eq. (2.57) was used. That said, the formulas presented in section 2.6.2 as in the original work [14] imply that originally a single atomic Thole polarizability was used to induce a full set of multipoles from monopoles to quadrupoles. In this work, the Thole polarizabilities were only used to induce dipoles.

DMA multipoles were calculated by the original DMA code provided by A. J. Stone. Multipoles up to quadrupoles were used, except for hydrogens for which the expansion was limited to rank 1 (dipoles) as instructed in the DMA manual [44]. In DMA this means that higher moments on H atoms were transferred to the nearest atom with a higher limit.

To include polarizability, the DMA was combined to the analytical method by

a C code written by the author. It is worth noting that adding polarizability to DMA with this method will not alter the algorithm's ability to reproduce the ESP around a molecules as the original DMA dipole will essentially just be presented as a sum of permanent and induced dipoles on each atom (eqs. (2.53), (2.57)).

3.6 GMM and ΔESP method

The code for the Gaussian multipole model was provided by D. Elking. This code was then combined with the Thole model using the ΔESP method. Again, solving of eq. (2.9) was done by matrix inversion in order to avoid possible convergence issues. GMM multipoles up to quadrupoles were used. GMM requires an initial guess for λ and all the multipole moments for each atom in a molecule. In this work, initial guess of 0 for all the multipole moments and 4 for λ (eqs. (2.49) and 2.50) were used as recommended by Dr. Elking.

The code provided by D. Elking was slightly altered version of the original, as the original contained a modified part of a commercial software. This part is related to the grid selection used for estimating the fit to *ab initio* potential and the electron density. The version used in this work utilized the `Gaussian09 cubegen` tool to create potential and electron density grids for the GMM fitting. Medium point density with `cubegen` option -3 was used for the grids.

The interactions of Gaussian multipoles are different from the interaction tensor approach used with the other methods. The electric field to be inserted in eq. (2.9) can be calculated by taking the negative gradient of eq. (2.50), and from that one can see that the contracted Gaussian nature of the multipoles distorts the form of the interactions compared to the interaction tensors presented in 6.3. This makes the assignment of Thole damping factors (λ_3, λ_5, λ_7) more complicated. Here, the damping was assigned so that the damped terms corresponded to the same directional terms ($\delta_{\alpha\beta}$, R_α, $R_\alpha R_\beta$, $R_\alpha R_\beta R_\gamma$, $R_\alpha \delta_{\alpha\beta}$) as in the original damped tensor elements in eqs. (2.59), (2.60), and (2.61).

3.7 The statistics

The relative root mean square deviation (RRMS) was used here as a key figure to depict how well the different methods reproduce the *ab initio* calculated poten-

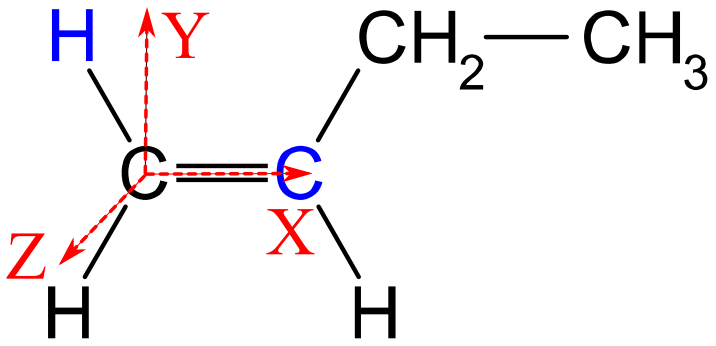


Figure 3.2: Local frame attached to a carbon atom in 1-butene. Atoms in blue are the neighbouring atoms used as a reference for assigning the x- and y-axis.

tials V_{QM}. RRMS reads

$$\mathrm{RRMS} = \sqrt{\frac{\sum_i^N (V_{\mathrm{QM},i} - V_i)^2}{\sum_i^N V_{\mathrm{QM},i}^2}}. \tag{3.1}$$

For GMM also weighted version was used to account for the electron density weighting used by the algorithm. The weighted RRMS is calculated as

$$\mathrm{RRMS_w} = \sqrt{\frac{\sum_i^N w_i (V_{\mathrm{QM},i} - V_i)^2}{\sum_i^N w_i V_{\mathrm{QM},i}^2}}, \tag{3.2}$$

where w_i is the weighting factor calculated by equation (2.52).

3.8 Conformational variance and the local frame

The variance of quality of fit and assigned charges/multipoles as a function of conformation were also studied. To do this, the charge/multipole fitting calculations were done separately for 4 conformations around the global energy minimum conformation, and 3 conformations around a second energy minimum conformation (table 3.1). Exceptions to this were methyl ethyl ether and methyl formate, as their second minimum in conformational energies were so high (8 kJ/mol and 22 kJ/mol above the global minimum conformation energy) that the conformations were determined to be irrelevant. This same set of conformations were also utilized for testing the performance of minimum energy conformation parameters. When studying the variance of assigned parameters, the charges/multipoles on the atoms in both ends of the 4 atom chain forming the

dihedral angle were of our interest.

In order for the multipole results of different conformations to be comparable, a local frame was attached to each atom of a molecule and the moments assigned to the atom where converted to this frame. The local frames where chosen so that x-axis pointed to a neighbouring atom with the longest chain attached. The y-axis was chosen orthogonal to the x-axis from the vector plane forming between the x-axis and a vector pointing to another neighbour of the atom at hand. The z-axis was chosen from the cross product of x- and y-axes (fig. 3.2).

Chapter 4

Results and discussion

4.1 The charge fitting algorithms: MK, CHELPG and RESP

In table 4.1 one can see an example of what kind of data was extracted from the charge fitting calculations using MK, CHELPG, and RESP. Although the magnitude of charges is independent of the coordinate system, the local frame definition was needed also in the case of charge fitting algorithms in order to convert the induced dipoles into a common coordinate system for different conformations of the molecule. In the columns "x-ref" and "y-ref" one can see which neighbouring atom was chosen to serve as the reference for the local coordinate system attached to each atom of the molecule. The numbers in these columns correspond to the atom indexing in the first column left.

4.1.1 Accuracy with respect to the ESP

The performance of the three different charge fitting algorithms was quite uniform. For the global minimum conformations of the test molecules, the polarizable versions of the charge fitting algorithms provided a better fit to the potential for all the molecules except 1-butene and butane (table 4.2). When comparing the results at a shell of points at 1.7×vdW radius from the atoms, one can see that the CHELPG method has the worst RRMS fit to the electrostatic potential, but MK and RESP provide very similar results. In our particular set of molecules, RESP gives on average a slightly worse fit but the differences are negligible.

Coordinates and local frame

		x	y	z	x-ref	y-ref
0	C	-1.319448	-1.226715	0.000000	1	9
1	O	-0.001216	-0.718914	0.000000	5	9
2	H	-1.250339	-2.311070	0.000000	6	3
3	H	-1.868410	-0.899626	0.889206	6	2
4	H	-1.868410	-0.899626	-0.889206	6	3
5	C	0.000000	0.697790	0.000000	1	8
6	C	1.433967	1.171915	0.000000	5	2
7	H	-0.533397	1.066807	-0.885171	5	8
8	H	-0.533397	1.066807	0.885171	5	7
9	H	1.471603	2.260717	0.000000	0	10
10	H	1.952482	0.804683	0.883541	0	9
11	H	1.952482	0.804683	-0.883541	0	10

Non-polarizable and polarizable fitting results

	RESP	RESP+Thole			
	q	q	μ_x	μ_y	μ_z
C	-0.07193	0.05766	0.36381	0.12019	0.00000
O	-0.37664	-0.79539	0.08061	-0.08331	0.00000
H	0.09638	0.12191	0.02079	0.01696	0.00000
H	0.05588	0.08259	0.01780	-0.00366	-0.00699
H	0.05588	0.08259	0.01780	-0.00366	0.00699
C	0.32485	0.60589	0.02422	-0.30930	0.00000
C	-0.35561	-0.35541	-0.02212	-0.11742	0.00000
H	-0.00229	-0.04714	-0.00445	-0.00346	-0.01932
H	-0.00229	-0.04714	-0.00445	-0.00346	0.01932
H	0.07701	0.12574	0.00367	-0.02507	0.00000
H	0.09938	0.08435	-0.00732	-0.00576	-0.01282
H	0.09938	0.08435	-0.00732	-0.00576	0.01282
RMS	0.00228	0.00152			
RRMS	0.22117	0.14719			
Induction energy		-38.74			

Table 4.1: An example of a local frame definition and charge fitting data: RESP fitting data for the minimum energy conformation of methyl ethyl ether. Atom coordinates are in Å, charge and dipole moment data are presented in atomic units and the unit of induction energy is kJ/mol.

	propionic acid	propanal	1-propanol	methyl formate
MK	0.0811	0.0945	0.1204	0.1407
MK+Thole	0.0784	0.0769	0.1076	0.0726
CHELPG	0.0891	0.0987	0.1273	0.1535
CHELPG+Thole	0.0864	0.0819	0.1209	0.0853
RESP	0.0819	0.0947	0.1207	0.1406
RESP+Thole	0.0783	0.0772	0.1083	0.0730
DMA	0.1520	0.1344	0.2265	0.1229
GMM	0.0076	0.0070	0.0162	0.0099
GMM+Thole	-	0.0119	0.0270	-

	methyl ethyl ether	dimethylethylamine	1-butene	butane
MK	0.2011	0.2748	0.2681	0.6898
MK+Thole	0.1370	0.1997	0.3055	0.7258
CHELPG	0.2159	0.3352	0.3840	0.7416
CHELPG+Thole	0.1531	0.2603	0.4323	0.8056
RESP	0.2020	0.2793	0.2696	0.6985
RESP+Thole	0.1371	0.1997	0.3029	0.7248
DMA	0.3106	0.5606	0.5252	1.5378
GMM	0.0101	0.0175	0.0212	0.0328
GMM+Thole	0.0113	0.0169	-	0.0645

Table 4.2: The RRMS error calculated for all the methods at 1.7× vdW distance from the molecule.

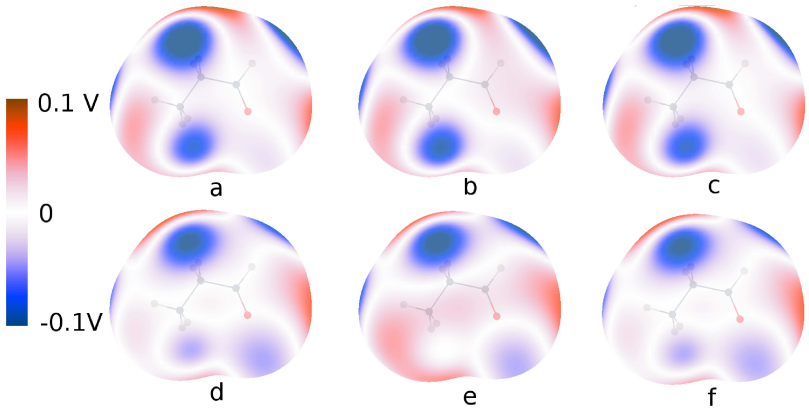


Figure 4.1: Difference between *ab initio* (MP2/aug-cc-pvtz) calculated electrostatic potentials and potentials from fitted charges and induced dipoles for a) MK b) CHELPG c) RESP d) MK+Thole e) CHELPG+Thole f) RESP+Thole.

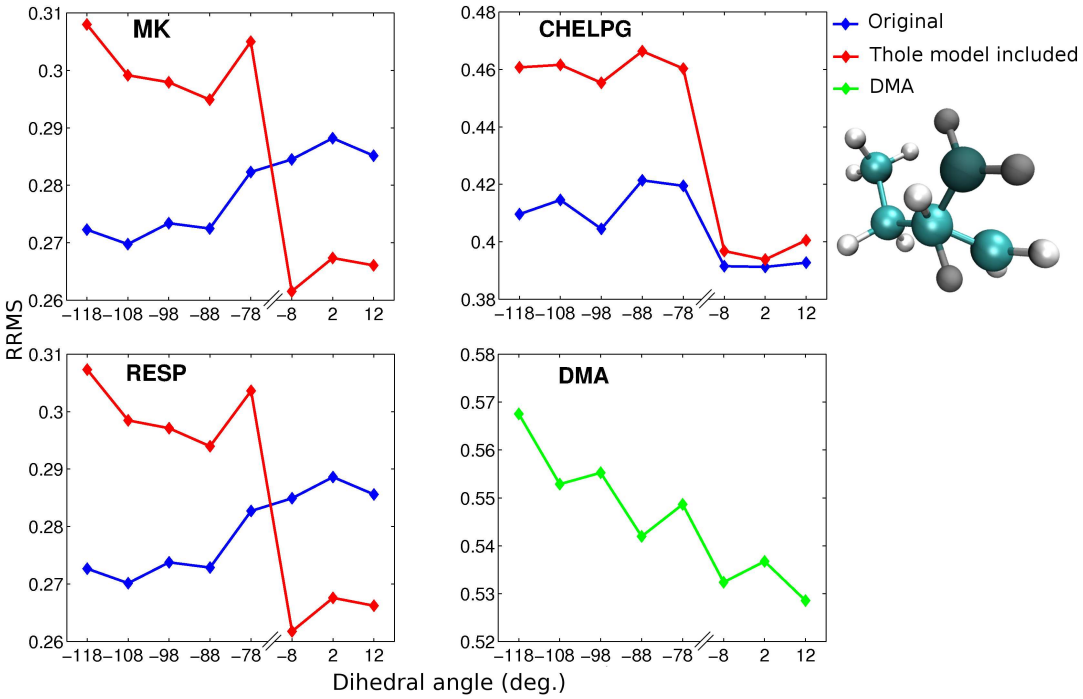


Figure 4.2: The RRMS errors of potential for different conformations of 1-butene. For MK, CHELPG, and RESP, the RRMSs were calculated in the grid used for the charge fitting, for DMA the MK grid was used. The different conformations are demonstrated by a ball-and-stick model depicting the global minimum conformation with a brighter shade and second minimum conformation with a darker shade (table 3.1).

These points are further illustrated in fig. 4.1 where the difference between *ab initio* potential and the potential from fitted charges and induced dipoles is visualized (by using `ORIENT` [45]) at 1.7×vdW for propanal. Here, one can clearly see how the addition of polarizability improves the fit.

In figs. 4.2-4.4 the conformational variance of the quality of fit for the original and polarizable versions of the three methods is presented. Here, the fitting of parameters was done separately for each conformation to demonstrate the best possible fit that can be obtained from the algorithm. The RRMS values are calculated in the grid used for the fitting in each method. Again, the RRMS errors of RESP and MK behave very similarly as a function of conformation.

As earlier, the non-polar molecules (table 3.1), 1-butene (fig. 4.2) and butane, seem to be problematic cases when polarizability is added. For 1-butene, the RRMS error is larger around the global minimum energy conformation for the polarizable case, but around the second minimum, one can see some improvement. The polarizable versions of MK and RESP do better around the second

minimum conformations than the non-polarizable ones. For polarizable CHELPG, the RRMS decreases around the second minimum but still remains slightly higher than for the non-polarizable version of CHELPG.

For all the other test molecules, polarizable versions of the algorithms provided better fits around the global minimum energy conformations. Example data of such case can bee seen in fig. 4.3 for 1-propanol. This was also the case for the second minimum conformations, with the exception of propanal (fig. 4.4) for which the original versions provided a better fit around the second minimum conformation for all the three charge fitting algorithms.

4.1.2 The performance of the minimum energy conformation parameters

The most simple way of assigning charges when parametrizing a force field is to use the parameters calculated for the minimum conformation. To see how well the ESP around different conformations could be reproduced by these parameters, the charges calculated for the minimum energy conformations were used to calculate the fit to the surrounding *ab initio* ESP for all the other conformations of the molecule. This was done both by using the original and polarizable versions of the MK, CHELPG, and RESP charge fitting procedures. In the non-polarizable case, the charges fitted to the ESP around the minimum energy conformation where straightforwardly applied to the other conformations. In the polarizable case, the minimum conformation charges were obtained by combining the charge fitting with ΔESP approach. The charges were then allowed to polarize the rest of the conformations of the molecule according to eq. (2.9). The fit to potential was calculated in MK-type grids for MK and RESP, and CHELPG-type grid for CHELPG. The results for 1-butene, butane, and 1-propanol can be seen in fig. 4.5. The data for the rest of the test molecules is presented in appendix A.

For most of our test molecules, the polarizable versions were able to reproduce the ESP around different conformations better than the non-polarizable ones when using the minimum conformation parameters. For dimethylethylamine, methyl formate, methyl ethyl ether, 1-propanol, and propionic acid, the performance of polarizable algorithm was better for all the conformations. The good performance of the polarizable charge fitting algorithms is demonstrated in fig. 4.5(c) for 1-propanol.

The only molecule for which the polarizable minimum parameters gave lar-

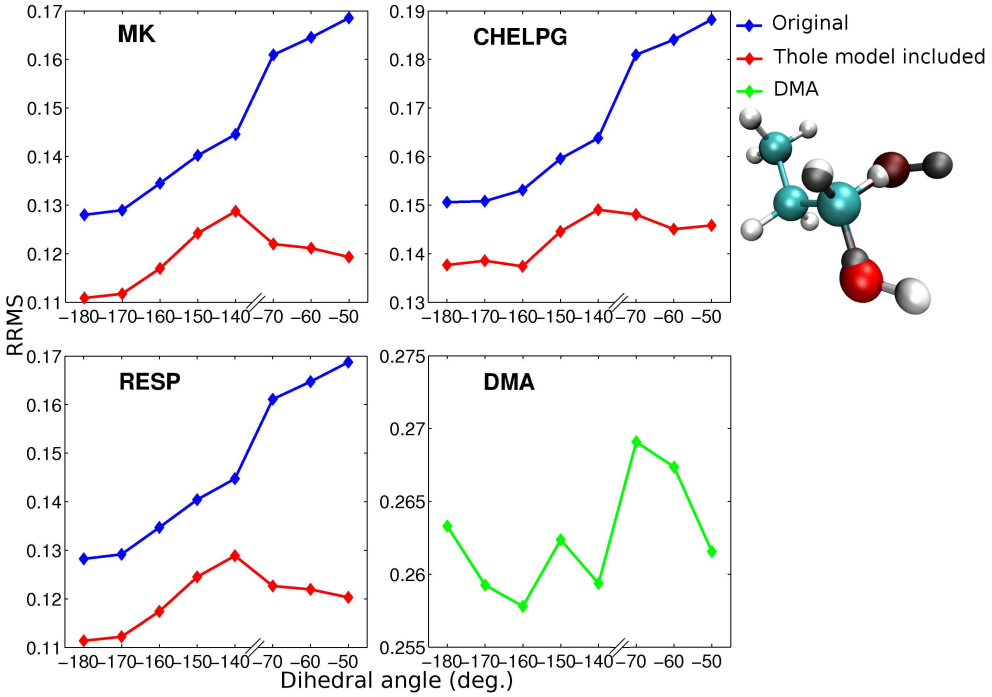


Figure 4.3: As in fig. 4.2 but for 1-propanol.

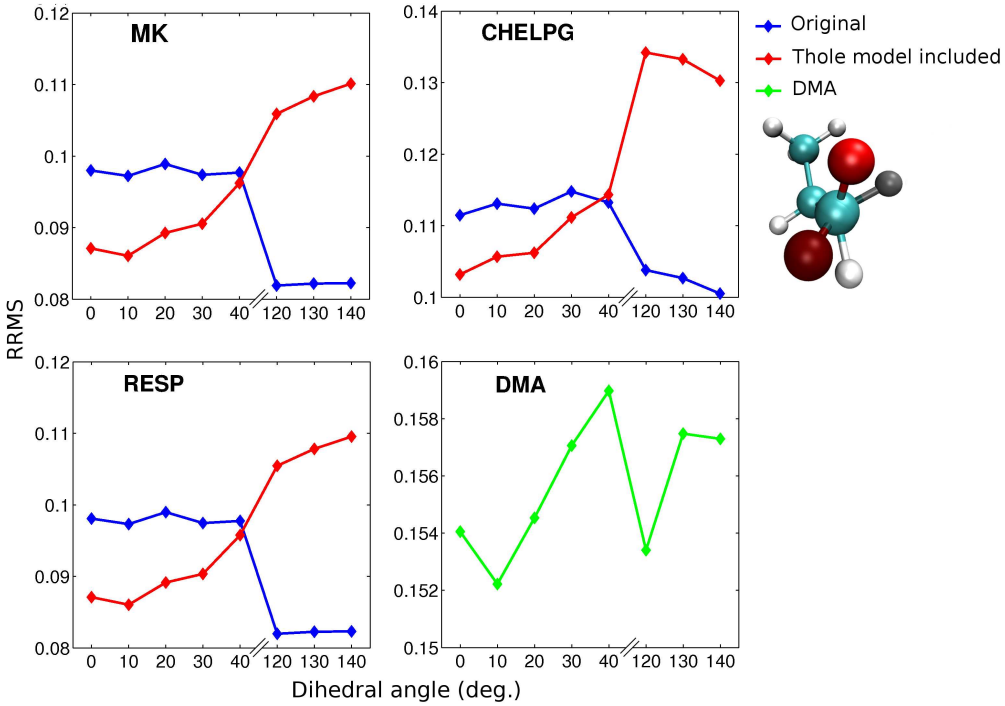
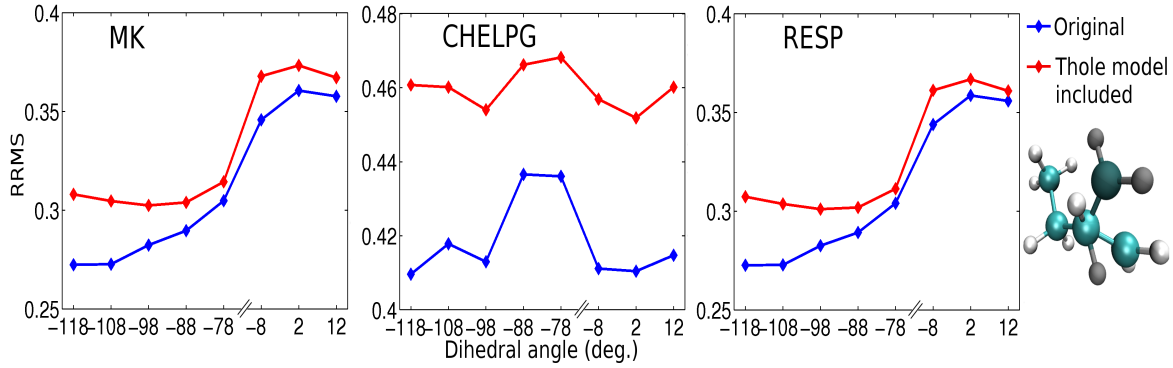
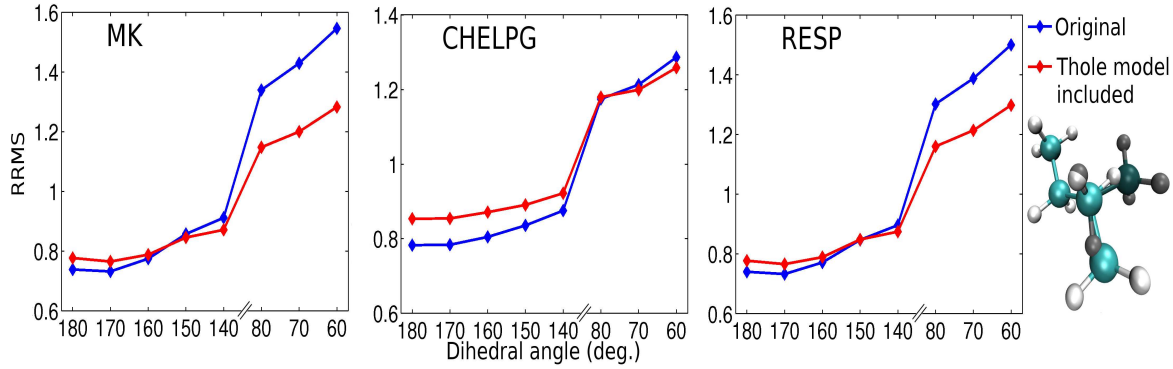


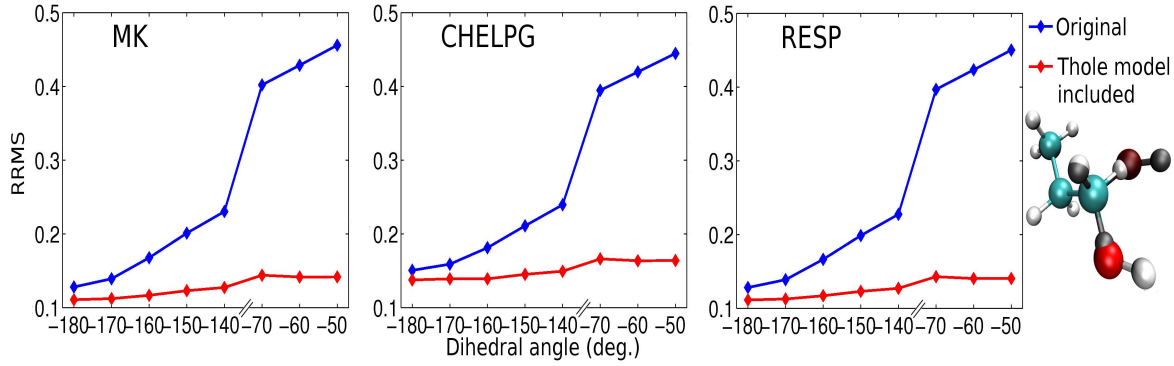
Figure 4.4: As in fig. 4.2 but for propanal.



(a) 1-Butene



(b) Butane



(c) 1-Propanol

Figure 4.5: The fit to surrounding ESP when using MK/CHELPG/RESP charges fitted for the minimum energy conformation of a) 1-butene b) butane c) propanol. The different conformations are demonstrated by the ball-and-stick models one the right depicting the global minimum conformation with a brighter shade and the second minimum conformation with a darker shade (table 3.1).

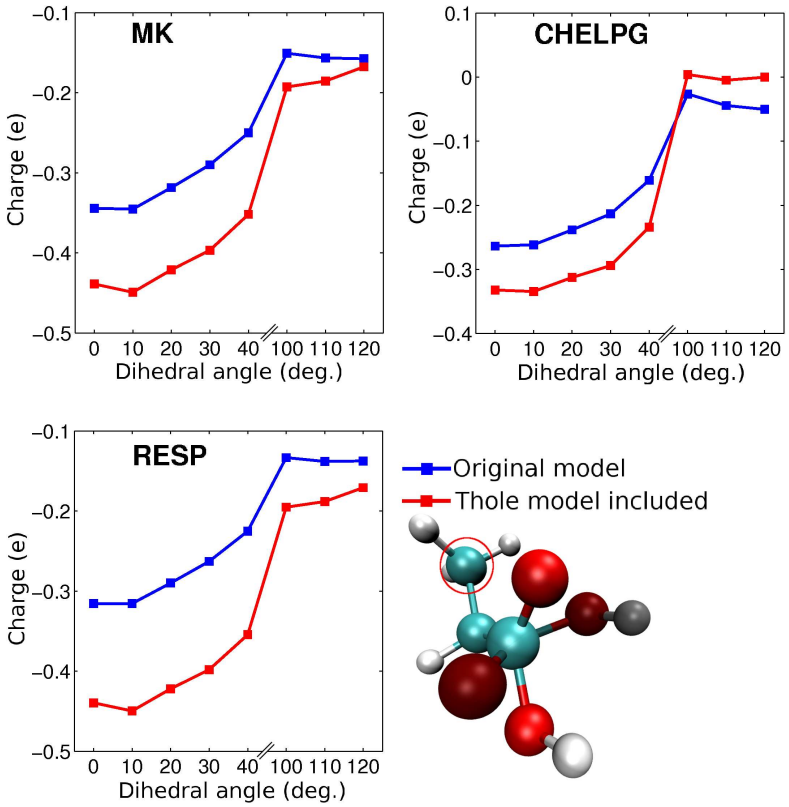


Figure 4.6: The variance of assigned charge as a function of conformation for propionic acid. The atom for which the charge is examined is circled and the second minimum conformation (table 3.1) is presented a with darker shade.

ger errors than the non-polarizable ones for nearly all the conformations was 1-butene (fig. 4.5(a)). For propanal and butane the situation differed between the first and second minimum conformations. For butane, the polarizable versions of the algorithms performed slightly worse around the global minimum energy conformation but around the second minimum, the RRMS errors of the original versions were larger (fig. 4.5(b)). From the polarizable versions of the three algorithms, CHELPG minimum energy parameters give slightly worse overall performance whereas MK and RESP results are almost identical.

4.1.3 The conformational variance of assigned parameters

Although using the parameters calculated for the minimum energy conformation is the simplest way of parametrising a force field, sometimes one can improve the performance of the force field by fitting the parameters for several conformations and using some kind of a weighted average. To see how the ad-

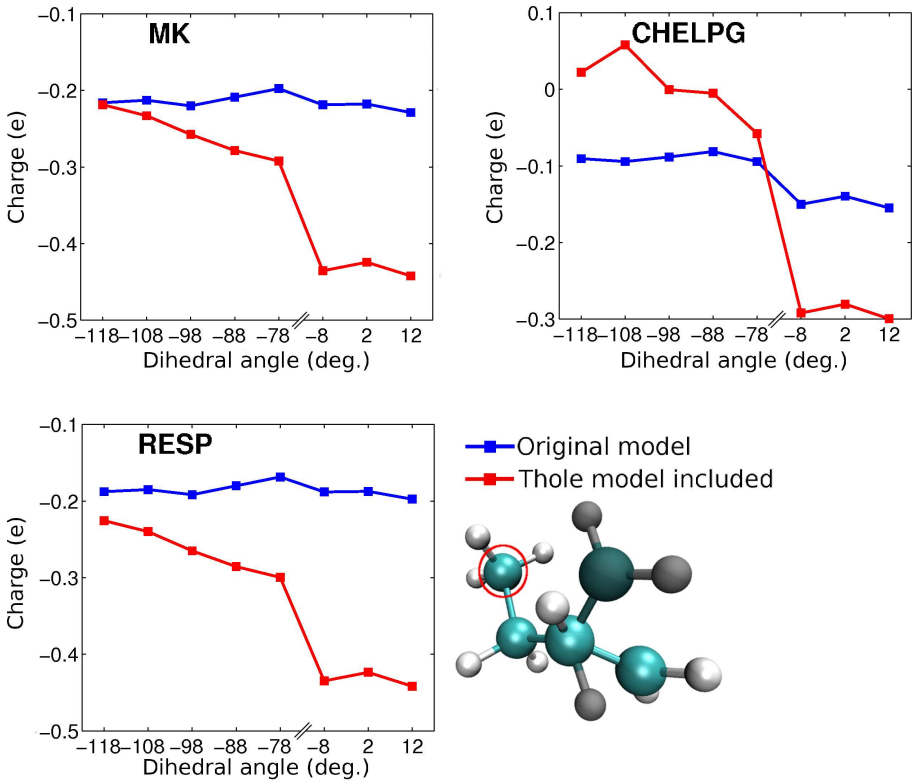


Figure 4.7: As in fig. 4.6 but for 1-butene.

dition of polarizability affects this, the fitting of parameters was done separately for each conformation and the conformational variance of assigned parameters was studied.

No common behaviour pattern could be determined when studying the charge fitting results from the non-polarizable and polarizable versions of the algorithms as a function of conformation. Most of the molecules studied here were borderline cases: one couldn't say definitely whether the original or the polarizable algorithm would be better in terms of the variance. MK and RESP once again gave very similar results and although CHELPG was was originally developed particularly to reduce the conformational variance of charges, no clear improvement compared to MK and RESP was detected.

Conformational variance of assigned parameters for propionic acid is depicted fig. 4.6 for the carbon at the end of the dihedral. The polarizable and non-polarizable algorithm results provide a similar variance for the first 4 conformations, but there is a large difference between the first and second minimum parameters for all three different algorithms. Very similar behaviour was also found

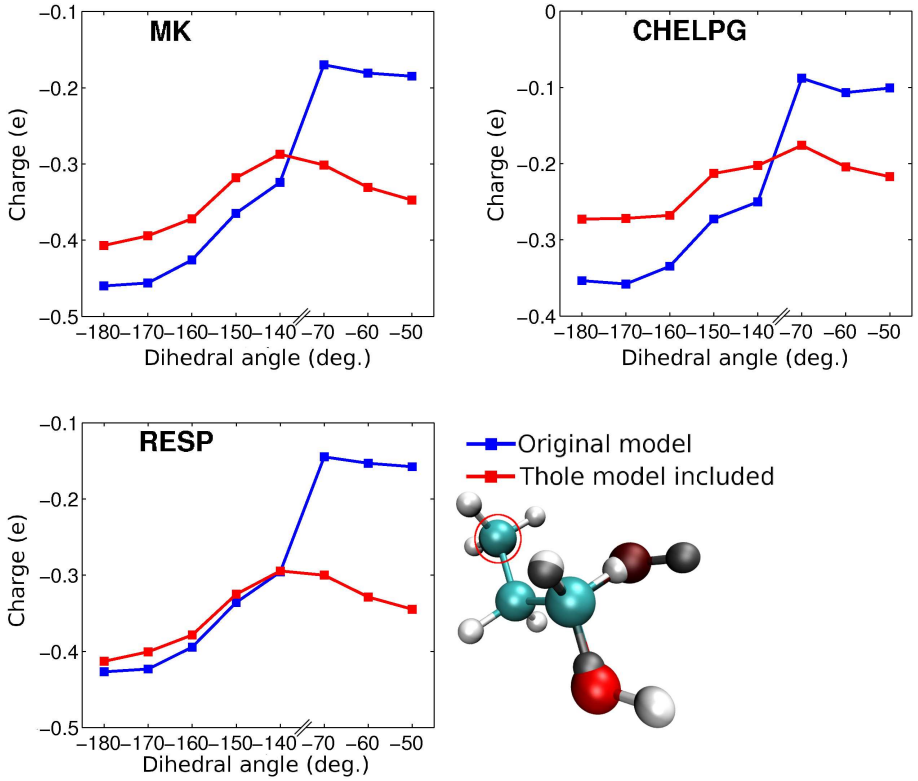


Figure 4.8: As in fig. 4.6 for 1-propanol.

for propanal in the case of the polarizable MK, CHELPG, and RESP.

The only test molecule for which the polarizable algorithm provided distinctly larger conformational variance was 1-butene (fig. 4.7). Here, the curves for assigned charge are considerably smoother for the original versions of the algorithm than the polarizable ones. The carbon at the end of the dihedral in 1-propanol serves as an example of a case were the polarizable versions perform better than the original versions of the algorithms (fig. 4.8). There is a large difference between the charges around the first and second minimum in the non-polarizable case whereas the curves for the polarizable case are present no such variance.

Altogether, the results in 4.6-4.8 indicate that while building a polarizable force field, one should proceed carefully if one plans to assign the parameters based on polarizable MK/CHELPG/RESP fitting data from multiple conformations. This is because the addition of polarizability can lead to larger conformational variance of fitted charges and using multiple conformations to fit the charges can actually make the results less accurate. It can also be noticed from figs. 4.6-4.8 that MK and RESP again provide very similar results and have almost exactly the same

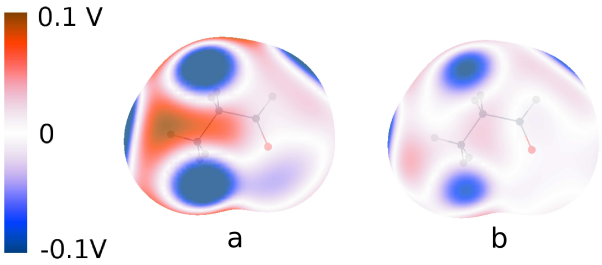


Figure 4.9: Difference between *ab initio* (MP2/aug-cc-pvtz) calculated electrostatic potentials and potentials from DMA multipoles for propanal. a) DMA multipoles up to quadrupole b) DMA multipoles up to octupoles.

variance of assigned parameters as a function of conformation.

4.2 DMA

4.2.1 Accuracy with respect to the ESP

It has been implied in the literature [46, 47] that in order to achieve good results with the DMA algorithm multipoles at least up to quadrupole have to be included in the calculations. Also, Ren *et al.* [14] use DMA multipoles up to quadrupoles in their efforts towards the polarizable AMOEBA force field. Our study indicates that one has to include at least octupoles to achieve as accurate results for the ESP around a molecule as given by the charge fitting algorithms MK, CHELPG, and RESP. This is seen by comparing the results in figs. 4.1 and 4.9. One can see that DMA up to quadrupoles gives clearly the largest difference to *ab initio* potential when compared to both the original and the polarizable versions of the charge fitting algorithms whereas DMA up to octupoles gives the smallest error.

In fact, methyl formate is the only test molecule for which DMA up to quadrupoles provided a better fit to the surrounding ESP at 1.7×vdW than the charge fitting methods (table 4.2). DMA would do better when comparing at larger distances. This is bacause DMA is not a fitting procedure based on the potential around a molecule, but the multipole moments are calculated from the quantum mechanical electron densities. The error demonstrated in fig. 4.9 compared to the *ab initio* calculated is almost purely due to the exclusion of higher moments from the electrostatic potential calculations, and these contributions from higher moments decay fast as a function of distance.

The conformational variance of the RRMS error between quantum mechanically calculated potential and the potential from the DMA multipoles (fitted for

each conformation separately) was added to figures 4.2-4.4 for reference. Since the grids used for calculating the RRMS were same for MK, RESP, and DMA, these values are directly comparable, and one can clearly see that MK and RESP provide a better fit also for all the other conformations for these test molecules.

4.2.2 The performance of the minimum energy parameters

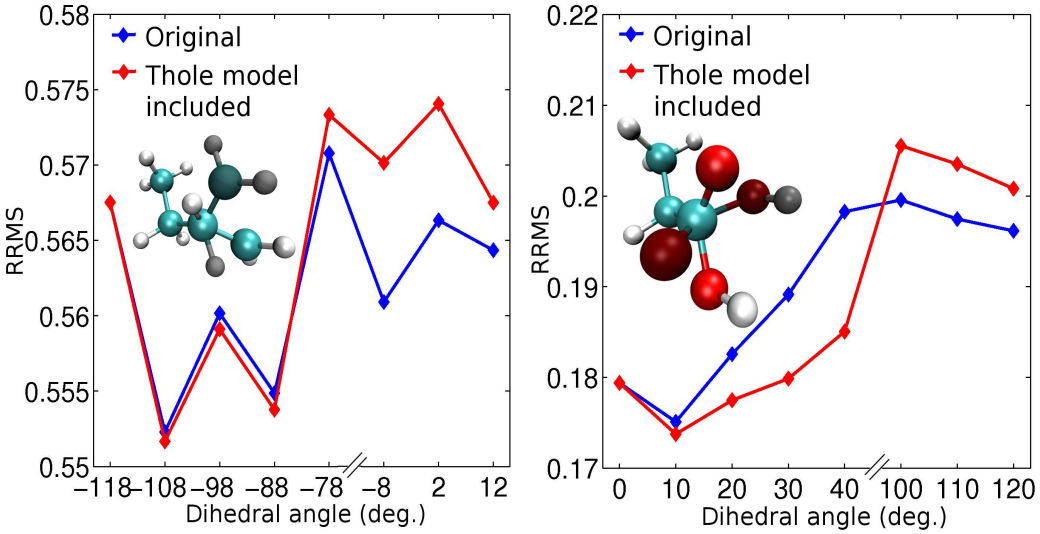
In fig. 4.10 one can see examples on how the polarizable and non-polarizable DMA parameters calculated for the minimum energy conformation perform for other conformations of the molecule. Here, the data for 1-butene, propionic acid, and 1-propanol is presented. The results for the rest of the test molecules can be seen in appendix B.

As mentioned earlier, the addition of polarizability in the case of DMA simply means dividing the DMA dipole into permanent and induced contributions. In the case of polarizable DMA the permanent multipoles are the parameters transferred from the minimum conformation. They are then allowed to induce dipoles for the other conformations. It follows that the ability of polarizable and non-polarizable DMA to reproduce the ESP will be exactly the same for the minimum energy conformation, but for the other conformations eq. 2.53 no longer holds, and the performances of polarizable and non-polarizable DMA starts to differ.

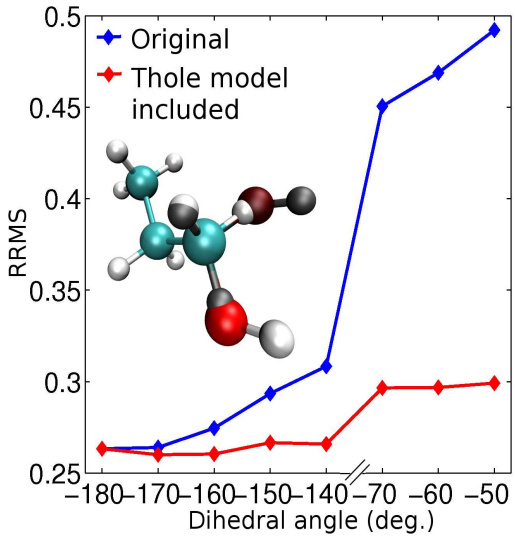
Overall, the polarizable version of DMA performs well when one starts to rotate the dihedral. Once again, 1-butene (fig. 4.10(a)) and butane are the most problematic cases for the polarizable version of the algorithm. For all the rest of the test molecules, the polarizable minimum energy conformation parameters provide a better fit for most of the conformations. For propionic acid, the polarizable DMA has slight problems around the second minimum conformation (fig. 4.10(b)). Same is true for the last studied conformation of methyl ethyl ether. That said, the most common behaviour of non-polarizable and polarizable DMA was the one that is demonstrated for 1-propanol in fig. 4.10(c) where the polarizable DMA provides a better fit for all the conformations. Here, we can see a clear improvement particularly around the second minimum.

4.2.3 The conformational variance of assigned parameters

Much of the same said about the conformational variance of assigned parameters in the case of polarizable and non-polarizable versions of the charge fitting

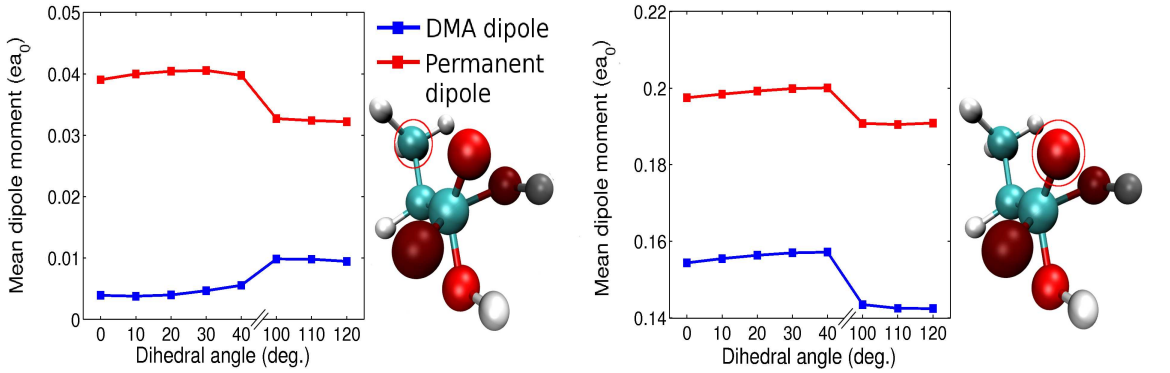


(a) 1-Butene (b) Propionic acid

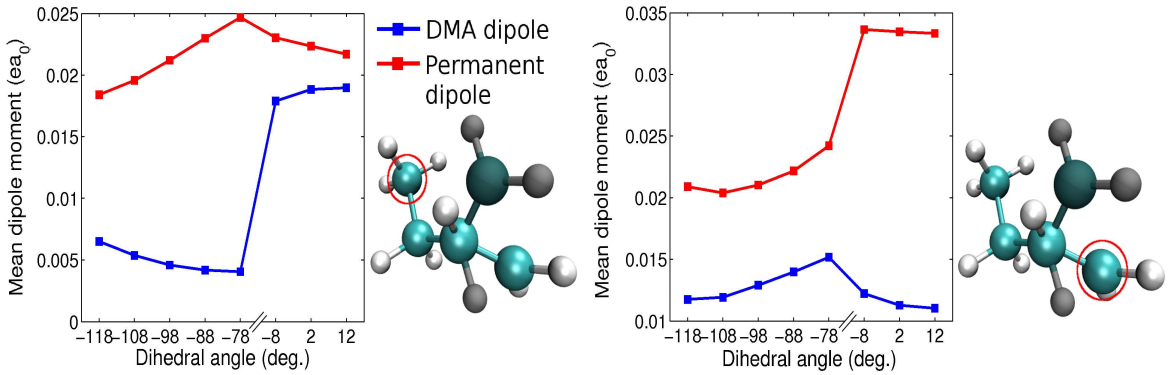


(c) 1-Propanol

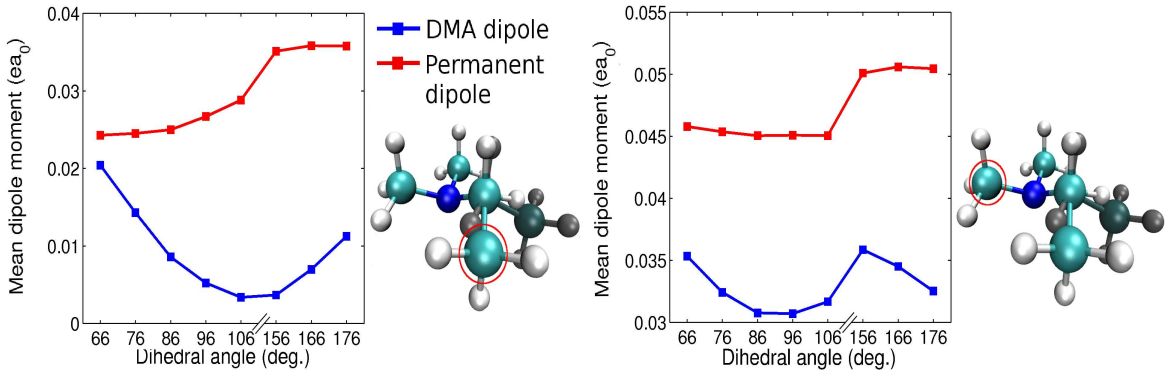
Figure 4.10: The fit to surrounding ESP when using both polarizable and non-polarizable DMA parameters fitted for the minimum energy conformation of the molecule. Results are presented for a) 1-butene b) propionic acid c) 1-propanol. The RRMS is calculated in the MK grid.



(a) Propionic acid



(b) 1-Butene



(c) Dimethylethylamine

Figure 4.11: The conformational variance of mean of dipole moment components μ_y, μ_y, and μ_z for 3 different molecules: a) propionic acid b) 1-butene c) dimethylethylamine. The DMA dipoles are presented in blue, and the dipole where polarizability has been extracted (permanent dipole) is presented in red. The atoms for which the moments are being examined are circled. The mean of dipole moments are in atomic units (elementary charge times the Bohr radius)

algorithms also applies to the polarizable and non-polarizable DMA (fig. 4.11). For most of the test molecules, the addition of polarization didn't make a clear difference to the conformational variance of assigned dipole moment. An example of this can be seen in fig. 4.11(a) where variance of the mean of dipole components is very similar for both the carbon and the oxygen atom. In the case of the carbon atom, the dipole moment from which polarizability has been extracted has slightly smaller difference between the global and the second minimum. For the oxygen atom, the situation is the opposite.

It is also possible that the behaviour is completely opposite when studying the moments assigned to different atoms in a molecule. From fig. 4.11(b) one can see that while the polarizable model performs overall well in the case presented on the left side of the figure, it provides a larger conformational variance compared to the original DMA when studying the carbon atom on the other end of the chain (right side of the figure). Of course, addition of polarizability can also reduce the conformational variance. This can be verified from fig. 4.11(c) where the improvement is clear in the case of dipole moment assigned to ethyl group carbon (left), but it is less obvious in case of the methyl group carbon (right).

All in all, one can make the same conclusion as in the case of the charge fitting algorithms: while building a polarizable force field, assigning the parameters based on data from multiple conformations may be a bad idea, since the addition of polarizability can make the variance of parameters as a function of conformation larger.

4.3 GMM

4.3.1 Computational requirements and convergence issues

As a method, GMM was considerably more demanding on the computational resources than any other method studied in this work. Particularly the memory consumption for one GMM fitting to the surrounding potential was large. For example, the memory requirement for fitting GMM multipoles to dimethylethylamine, containing 16 atoms, was around 5-8 GB of memory depending on the optimization level used when compiling the code. The memory consumptiont is probably mainly caused by the point selection around the molecule. GMM uses considerably more extensive grid (up to 10^6 grid points) for estimating the ESP than the point charge methods used in this work. Since the GMM version

provided by D. Elking had a different grid selection scheme compared to the version used in the original work [30], it is hard to say if the original version of GMM would perform any better.

The high memory consumption and long computation times mean that GMM was not the best method to be combined with the Thole model by an iterative approach. Solving the GMM multipoles, when the model was combined with the Thole model, could take up to 3-4 weeks on a regular PC. Of course, the process could have been accelerated by increasing the optimization level of the executable, but this would result in a higher memory consumption.

For 3 out of 8 test molecules, the ΔESP iteration failed to converge. GMM requires an initial guess for the scaling parameters λ (eq. 2.42) and all the multipole moments for each atom in a molecule. In this work, initial guesses recommended by Dr. Elking were used (section 3.6). By changing these, when combining GMM with the Thole model, convergence could maybe have been achieved, but unfortunately the long iteration time made the examination of different initial guesses unfeasible.

That said, more likely the convergence issues stem from the fact that GMM was originally developed because atom point multipole models tend to underestimate electrostatic interactions at close (dimer) distances. Correcting this problem has probably also led to the enhancement of 1-2 and 1-3 interactions so that the damping applied to these interactions in this work (section 2.6.3) can no more prevent the overpolarization when solving eq. 2.9.

4.3.2 Accuracy with respect to the ESP

Although GMM was inefficient, it was able to produce the potential around a molecule very accurately. From table 4.2 one can see that GMM gives by far the smallest error for all the molecules both in the non-polarizable and polarizable case. One should note that the RRMS errors for GMM in this table are non-weighted ones for maximum comparability instead of the weighted errors used for the rest of the GMM results (eq. 3.2).

Unfortunately, even when the ΔESP iteration was successful, the addition of polarizability into GMM reduced the quality of fit at 1.7×vdW with the exception of dimethylethylamine. The fairly complex fitting process and the weighting of points that GMM uses might be a contributing factor to the poor result. Nevertheless, even the polarizable GMM does give a better fit to the ESP than the

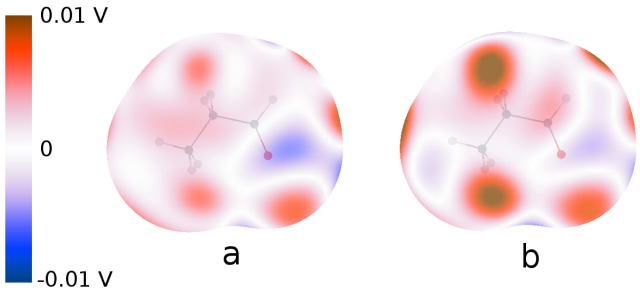
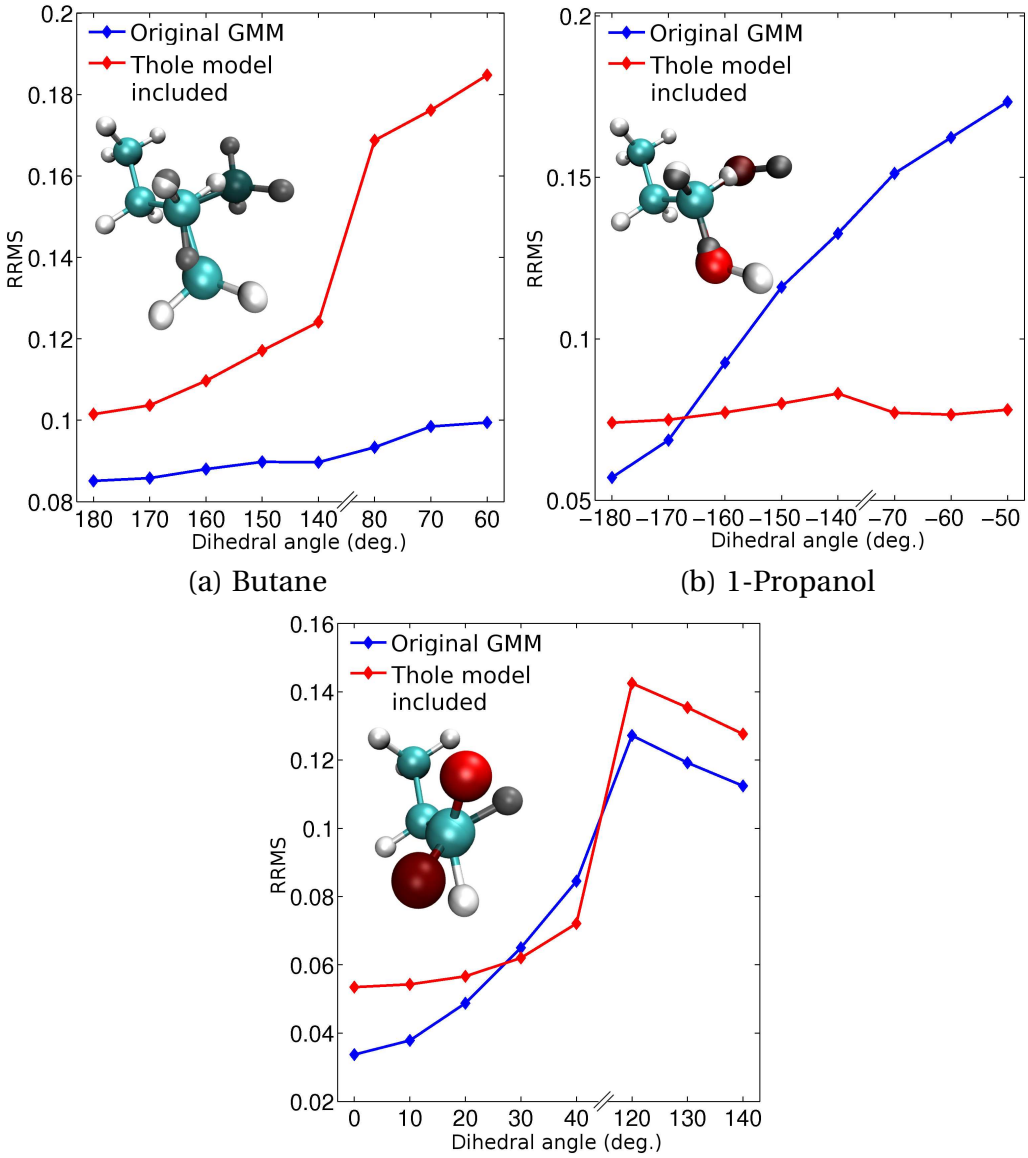


Figure 4.12: Difference between *ab initio* (MP2/aug-cc-pvtz) calculated electrostatic potentials and potentials from fitted GMM multipoles for propanal. a) GMM multipoles up to quadrupole b) GMM multipoles up to quadrupole combined with the Thole model.

polarizable and non-polarizable charge fitting methods, not to mention DMA. The overall excellent fit is also illustrated in fig. 4.12 where one can see that only by using a scale that is an order of magnitude smaller that the one used in figs. 4.1 and 4.9 for other methods one can see some difference between the potential from the GMM multipoles and the potential calculated *ab initio*.

4.3.3 The performance of the minimum energy conformation parameters

The slow iteration process made looking into conformational variance of optimal fit and assigned parameters unfeasible. That is, in the case of GMM the study was limited examining into how well the parameters calculated for the minimum energy conformation can reproduce the ESP around other conformations of that molecule. In the non-polarizable case, the GMM parameters fitted for the ESP around the minimum conformation were straightforwardly applied to the other conformations. In the polarizable case, the parameters where obtained from the GMM fitting performed together with the ΔESP iteration for the minimum conformation. These parameters where then applied to the other conformations of that molecule and allowed to induce dipoles for those conformations according to eq. (2.9). The results for this are presented in fig. 4.13 for 3 molecules. The data for the rest of the test molecules (for which the ΔESP iteration converged) can be seen in appendix C. Again, the non-polar butane (fig. 4.13(a)) seems to be the most problematic case for the polarizable version of GMM as it was the only molecule for which the weighted RRMS was constantly higher for the polarizable version of the algorithm. That said, the performance of the minimun energy conformation parameters for propanal (fig. 4.13(c)) and dimethylethyl-



(a) Butane (b) 1-Propanol

(c) Propanal

Figure 4.13: The fit to surrounding ESP when using both polarizable and non-polarizable GMM parameters fitted for the minimum energy conformation of a) 1-butene b) 1-propanol c) propanal.

	MK	CHELPG	RESP	DMA	GMM
1-butene	-24.54	-17.85	-24.32	-2.53	-
butane	-6.92	-5.81	-6.91	-2.39	-91.85
dimethylethylamine	-47.45	-35.38	-47.13	-8.57	-102.06
methyl ethyl ether	-39.30	-36.94	-38.74	-14.92	-72.61
methyl formate	-37.86	-37.18	-37.61	-13.35	-
propanal	-49.82	-50.44	-49.56	-20.12	0.00
1-propanol	-63.76	-62.06	-63.24	-12.26	-106.46
propionic acid	-47.55	-48.27	-47.39	-14.25	

Table 4.3: Induction energies (kJ/mol) for polarizable versions of all the methods used in this work.

amine was also quite poor. For these molecules, the polarizable version provided a better fit for 1-2 conformations.

Despite the poor performance at the minimun conformation, the polarizable minimum conformation parameters performed considerably better when one starts to rotate the dihedral for 2 molecules. For 1-propanol (fig. 4.13(b)) the polarizable version provided smaller error from the third conformation onwards. For methyl ethyl ether, the non-polarizable version provided a worse fit only for the minimum conformation. Same kind of trend can be detected for propanal and dimethylethylamineamine: the performance of the polarizable version improves around the first minimun conformation the further one gets from the minimum. These results might further imply that polarizability in itself was not the problem, but there is room for improvement in the strategy used for combining the GMM to the Thole model.

4.4 Induction energies

Induction energies were computed for each method using eq. (2.8), where $\boldsymbol{E_0}$ was calculated from the charges/multipoles produced by the method when it was combined with Thole's model. The induction energies are presented in kJ/mol in table 4.3 for the minimum energy conformations of the test molecules. One can see that the energies are the lowest for DMA, and again the differences between the charge fitting algorithms are small. CHELPG has slightly lower energies compared to MK and RESP for most of the test molecules. GMM produces clearly the highest induction energies.

Some reference for the order of magnitude of the induction energies can be obtained from the work by Söderhjelm and Ryde [48] where they compare induction and polarization energies of their own polarizable models versus the polarizable amber 2002 force field [35] for a system consisting of a protein binding site and a small ligand. Söderhjelm and Ryde obtained induction energies around -160 kJ/mol for their own model and around -100 kJ/mol for Amber ff02. Since the Amber force field is based on the ΔESP model and RESP charges, it should give a valid comparison point for the order of magnitude. One can say that the order of magnitude of the induction energies presented here for MK, CHELPG, RESP, and DMA is realistic. In the light of these results it also seems that the concern by Ren *et al.* [14] that large induction energies would arise from the polarizable DMA model, when 1-2 interactions are included, was exaggerated.

The induction energies for GMM are definitely above the higher limit of what is reasonable considering that systems studied here are significantly smaller than the system studied by Söderhjelm and Ryde [48]. This can be seen as another proof of the stronger 1-2 and 1-3 interactions as they would also lead to the immediate growth of induction energies.

Chapter 5

Conclusions

In this work, three charge fitting algorithms and two multipole methods were combined together with the induced point dipole model by Thole to see how the different methods would perform when polarizability contributions are self-consistently removed from the charge/multipole assigning process. In general, no universal behaviour pattern could be determined because the results varied a bit over the set of test molecules used in this work. This is probably partly due to the fact that the test molecules used here ranged from electrostatically neutral butane and 1-butene to considerably more polar propanal. One can see that the less polar 1-butene and butane were usually among the molecules that the polarizable versions of the algorithms had most problems with. The best performance was often obtained for 1-propanol which had mid-range polarity out of the molecules in the test set.

The lack of consistency might also relate to the parametrization of the Thole model. The original idea by Thole was that one would need only one isotropic polarizability per element without having to pay attention to the chemical environment of the atom in the molecule. However, when one takes a look at the data obtained for previous parametrizations of Thole model [18, 21] one can see that the model is not able to reproduce the experimental polarizabilities of alkenes as well as for other chemical groups. Although the parametrization used in this work performed better in that sense, there was still clearly room for improvement [43]. This suggests that the Thole model could benefit from the addition of different polarizability parameter for different carbon types which could in turn also increase the applicability of polarizability models on less polar molecules, such as alkanes and alkenes, from what we have seen in this work.

That said, some common trends could be established. For the charge fitting algorithms, the polarizable versions provided a better fit to the surrounding ESP in most cases. This was true both when the fit was done separately for each conformation a molecule and when the parameters fitted for the minimum energy conformations were used to reproduce the ESP around other conformations of that molecule. The differences between the three charge fitting methods were very small. Generally, CHELPG gave slightly worse fit to the potential.

MK and RESP provided very similar results in every analysis done in this work. This is understandable because RESP is an modification of MK developed to prevent the assignment of high charge values on the deeply buried atoms typical for MK and CHELPG. However, the test molecules used in this work are very simple and contain no such atoms.

It was surprising that DMA up to quadrupoles provided such a poor fit to the surrounding electrostatic potential. One source of the error is the exclusion of higher moments from the otherwise accurate DMA analysis and our results indicate that DMA up to octupoles is needed in order for DMA to beat the performance of the charge fitting algorithms at the close vicinity of a molecule. However, the inclusion of octupoles into a force field would be highly impractical, when already the inclusion of dipoles and quadrupoles leads to increasing complexity and computational cost. It was found encouraging that also in the case of DMA the polarizable minimum energy conformation parameters performed better than the non-polarizable parameters for most of our test molecules.

Convergence issues arose when combining GMM together with the Thole model. Based on the high induction energies GMM produced, we suspect the problems were due to strong 1-2 and 1-3 interactions, and the solution could be the total exclusion of these interactions instead of the damping used in this work. That said, GMM was by far the most accurate of all the methods studied here in reproducing the ESP around a molecule. Unfortunately, the polarizable version of GMM did not provide a better fit to the surrounding potential compared to the non-polarizable GMM even when convergence was achieved. Considering this, it is not a suprise that also the performance of the polarizable minimun energy conformation parameters, when applied to other conformations of the molecule, was somewhat poor. One should remember that GMM is a relatively new method compared to the other methods studied in this work. Hence, there is potential for further development both in GMM itself and in strategies used for combining it together with the Thole model. That said, the fairly complex functional forms for

the electric field and the potential resulting from the Gaussian multipoles makes the GMM less appealing for the force field purposes.

Altogether, the data presented in this work indicates that, at the moment, either MK or RESP would be the best choice for polarizable force field parametrizations out of the methods studied here. MK and RESP provide a good compromise between accuracy and computational efficiency not to mention the ease of force field implementation.

References

[1] H-D. Höltje, W. Sippl, D. Rognan, and G. Folkers. *Molecular modeling: basic principles and applications*. Wiley-VCH, 2008.

[2] M. F. Schlecht. *Molecular Modeling on the PC*. John Wiley & Sons, Inc, 1998.

[3] P. Cieplak, F-Y. Dupradeau, Y. Duan, and J. Wang. Polarization effects in molecular mechanical force fields. *J. Phys.:Condens. Matter*, 21, 2009.

[4] W. F. van Gunsteren, D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D. Geerke, A. Glättli, P. Hünenberger, M. Kastenholz, C. Oostenbrink, M. Schenk, D. Trzesniak, N. van der Vegt, and Haibo B. Yu. Biomolecular modelling: Goals, problems, perspectives. *Angew. Chem. Int. Ed.*, 45, 2006.

[5] S. W. Rick and S. J. Stuart. Potentials and algorithms for incorporating polarizability in computer simulations. In *Reviews in Computational Chemistry*, volume 18. Wiley, 2002.

[6] B. Brooks, R. Bruccoleri, B. Olafson, D. States, S. Swaminathan, and M. Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, 4, 1983.

[7] A. D. MacKerell, B. Brooks, C. L. Brooks, III, L. Nilsson, B. Roux, Y. Won, and M. Karplus. CHARMM: The energy function and its parameterization. In *Encyclopedia of Computational chemistry*, volume 1. John Wiley & Sons, 1998.

[8] W. Cornell, P. Cieplak, C. I. Bayly, Ian R. Gould, K. Merz, D. Ferguson, D.. Spellmeyer, T. Fox, J. W. Caldwell, and P. A Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117, 1995.

[9] GROMOS. Groninger molecular simulation program package. University of Groninger, Groninger, 1987.

[10] 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*, 110, 1988.

[11] A. D. MacKerell, Jr. Empirical forece fields for biological macromolecules: overview and issues. *J. Comput. Chem.*, 25, 2004.

[12] J. Klauda, R. Venable, A. MacKerell, Jr, and R. Pastor. *Considerations for Lipid Force Field Development*, volume 60 of *Current Topics in Membranes*. Elsevier, 2008.

[13] H. Yu and W. F. van Gunsteren. Accounting for polarization in molecular simulation. *Comput. Phys. Commun.*, 172, 2005.

[14] P. Ren and J. W. Ponder. Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations. *J. Comput. Chem.*, 23, 2002.

[15] T. D. Rasmussen, P. Ren, J. W. Ponder, and F. Jensen. Force field modelling of conformational energies: importance of multipole moments and intramolecular polarization. *Int. J. Quantum Chem.*, 107, 2007.

[16] J. E. Davis, O. Rahman, and S. Patel. Molecular dynamic simulations of a dpmc bilayer using non-additive interaction models. *Biophys. J.*, 96, 2009.

[17] J. Applequist, J. R. Carl, and K. Fung. An atom dipole interaction model for molecular polarizability. application to polyatomic molecules and determination of atom polarizabilities. *J. AM. Chem. Soc.*, 4, 1972.

[18] B. T. Thole. Molecular polarizabilities calculated with a modified dipole interaction. *Chem. Phys.*, 59, 1981.

[19] P. E. M. Lopes, B. Roux, and A. D. MacKerell, Jr. Molecular modelling and dynamical studies with explicit inclusion of electronic polarizability: theory and applications. *Theor. Chem. Acc.*, 124, 2009.

[20] L. Silberstein. Molecular refractivity and atomic interaction. *Philos. Mag.*, 33, 1917.

[21] P. Th. van Duijnen and M. Swart. Molecular and atomic polarizabilities: Thole's model revisited. *J. Phys. Chem. A.*, 102, 1998.

[22] G. A. Kaminski, R. A. Friesner, and R. Zhou. A computationally inexpensive modification of the point dipole electrostatic polarization model for molecular simulations. *J. Comput. Chemp*, 24, 2003.

[23] C. U. Singh and P. A. Kollman. An approach to computing electrostatic charges for molecules. *J. Comput. Chem*, 5, 1984.

[24] C. Breneman and K. Wiberg. Determing atom-centered monopoles from molecular electrostatic potentials. the need for high sampling density in formamide conformational analysis. *J. Comput. Chem.*, 11, 1990.

[25] C. Bayly, P. Cieplak, Cornell W.., and Kollman P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.*, 97, 1993.

[26] R. S. Mulliken. Electronic Population Analysis on LCAO[Single Bond]MO Molecular Wave Functions. I. *The Journal of Chemical Physics*, 23, 1955.

[27] A. Reed, R. Weinstock, and F. Weinhold. Natural population analysis. *The Journal of Chemical Physics*, 83, 1985.

[28] C. Cramer. *Essentials of computational chemistry: Theories and Models.* Wiley, 2004.

[29] A. J. Stone and Alderton M. Distributed multipole analysis: methods and applications. *Mol. Phys*, 100, 2002.

[30] D. Elking, A. Cisneros, J-P. Piquemal, Darden T., and L. Pedersen. Gaussian multipole model (GMM). *J. Chem. Theory Comput.*, 6, 2010.

[31] B. Besler, K. Merz, and Kollman P. A. Atomic charges derived from semiempirical methods. *J. Comput. Chem.*, 11, 1990.

[32] A. J. Stone. Distributed multipole analysis: Stability for large basis sets. *J. Chem. Theory Comput.*, 1, 2005.

[33] S. F. Boys. Electronic wave functions i. a general method of calculation for the stationary states of any molecular system. *Proc. Roy. Soc, A*, 200, 1950.

[34] W. H. Press, B. B. Flannery, S. A. Teokolsky, and W. T. Vetterling. *Numerical Recipes: The art of scientific computing*. Cambridge University Press, 1986.

[35] P. Cieplak, J. Caldwell, and P. A. Kollman. Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation: aqueous solution free energies of methanol and n-methyl acetamide, nucleic acid base, and amide hydrogen bonding and chloroform/water partition coefficients of the nucleic acid bases. *J. Comput. Chem.*, 22, 2001.

[36] J. Ponder and D. Case. Force fields for protein simulations. *Advances in protein chemistry*, 66, 2003.

[37] W. Xie, J. Pu, and J. Gao. A coupled polarization-matrix inversion and iteration approach for accelerating the dipole convergence in a polarizable potential functio. *The Journal of Physical Chemistry A*, 113, 2009.

[38] P. Ren and J. Ponder. Polarizable atomic multipole water model for molecular mechanics simulation. *The Journal of Physical Chemistry B*, 107, 2003.

[39] A. J. Stone. *The theory of intermolecular forces*. Oxford university press, 1996.

[40] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. Gaussian 09 Revision A.1. Gaussian Inc. Wallingford CT 2009.

[41] T. Kinnunen, T. Nyrönen, and P. Lehtovuori. Soma2 - open source framework for molecular modelling workflows. *Chem. Cent. J.*, 2, 2008.

[42] P. Lehtovuori and T. Nyrönen. Soma - workflow for small molecule property calculations on a multiplatform computing grid. *J. Chem. Inf. Model.*, 46, 2006.

[43] H. Antila. A reparametrization of thole's model for polarizable lipid force field applications. Special assingment, Aalto University, 2010.

[44] A. J. Stone. Distributed multipole analysis of gaussian wavefunctions, gdma version 2.2.02, 2011.

[45] A. J. Stone, A. Dullweber, O. Engkvist, E. Fraschini, M. P. Hodges, A. W. Meredith, D. R. Nutt, P. L. A. Popelier, and D. J. Wales. Orient: a program for studying interactions between molecules, version 4.5. University of Cambridge, 2002.

[46] G. G. Ferenczy, P. J. Winn, and C. A. Reynolds. Toward improved force fields. 2. effective distributed multipoles. *J. Phys. Chem. A*, 101, 1997.

[47] A. J. Stone. Intermolecular potentials. *Science*, 321, 2008.

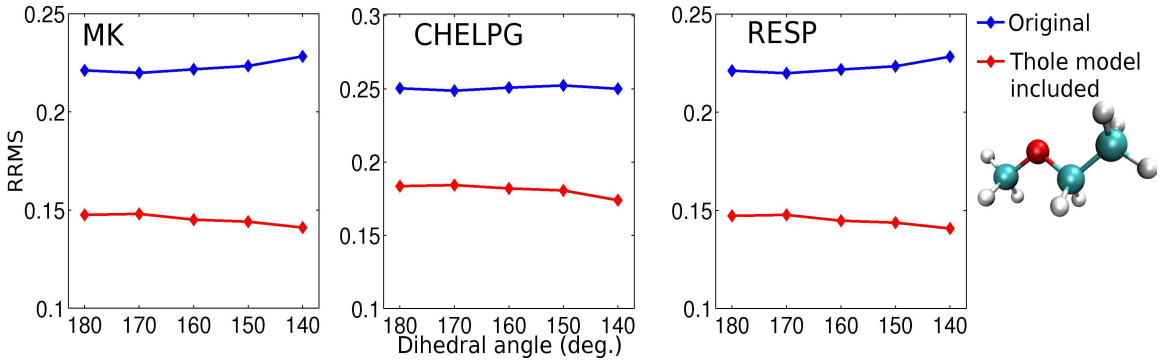
[48] P. Söderhjelm and U. Ryde. How accurate can a force field become? a polarizable multipole moldel combined with fragment-wise quantum-mechanical calculations. *J. Phys Chem. A.*, 113, 2009.

Appendices

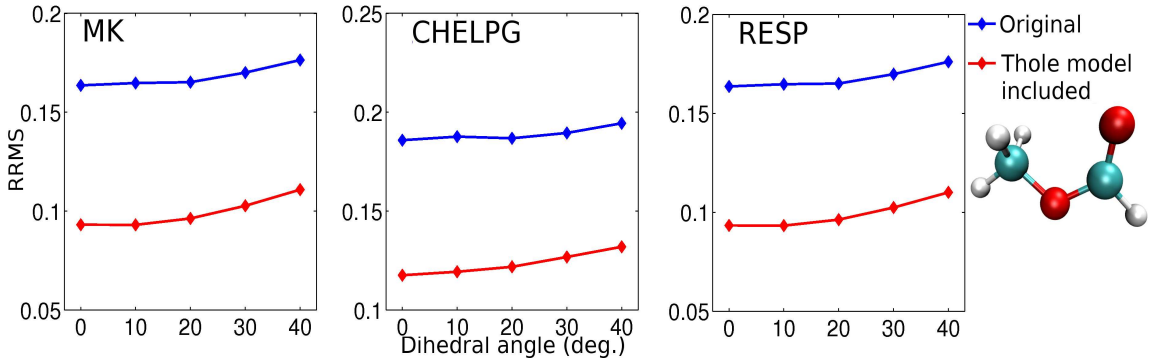
Appendix A

Additional data for MK, CHELPG and RESP

In this section data for the performance of the minimum energy conformation parameters in the case of methyl ethyl ether (fig. A.1(a)), methyl formate (fig. A.1(b)), dimethylethylamine (fig. A.2(a)), propanal (fig. A.2(b)), and propionic acid (fig. A.2(c)) is presented. The analysis covers both the non-polarizable and polarizable versions of MK, CHELPG, and RESP and is further elaborated in the section 4.1.2 of this work.

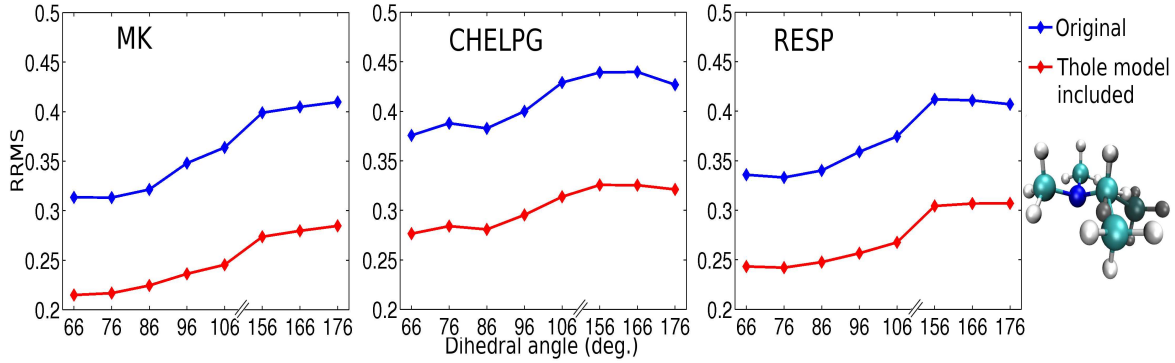


(a) Methyl ethyl ether

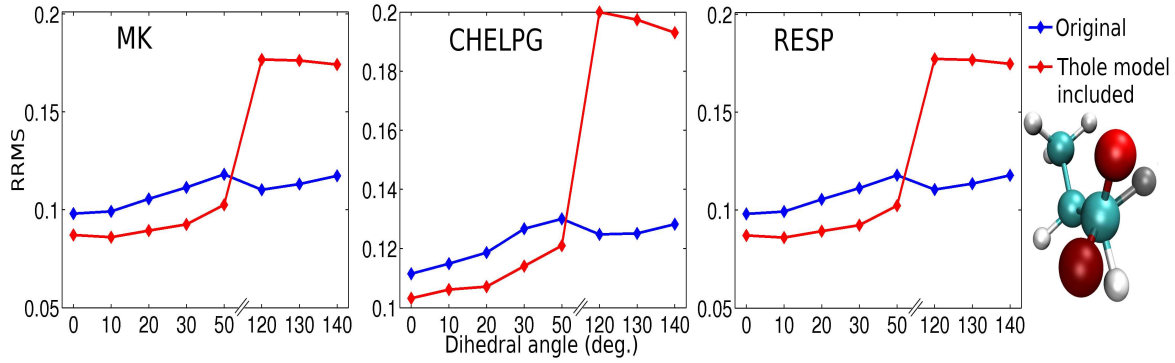


(b) Methyl formate

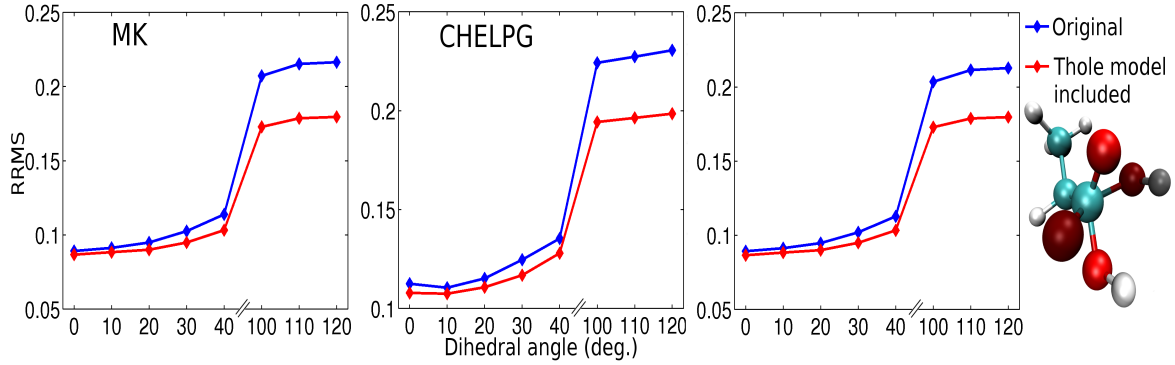
Figure A.1: The fit to the ESP when using MK/CHELPG/RESP charges fitted for the minimum energy conformation of a) methyl ethyl ether b) methyl formate. The global minimum conformation is depicted by the the ball-and-stick model one the right.



(a) Dimethylethylamine



(b) Propanal



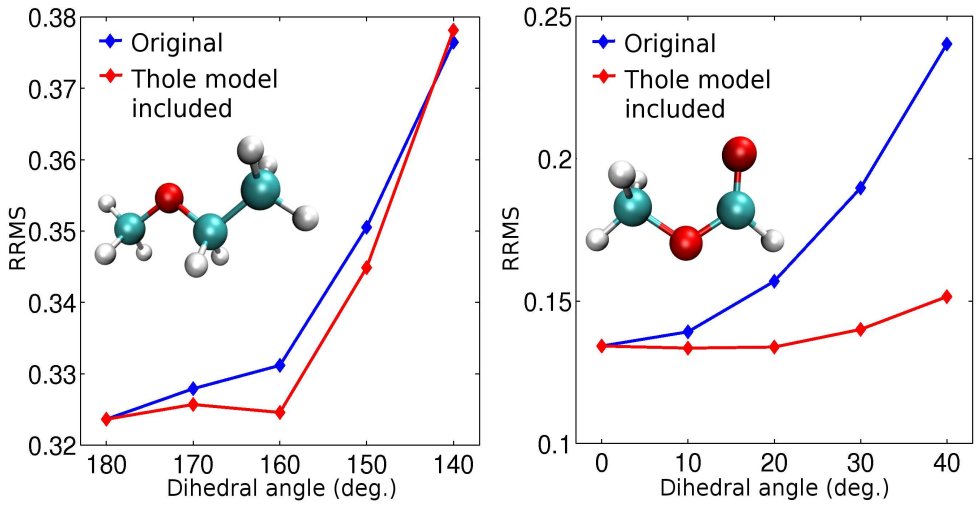
(c) Propionic acid

Figure A.2: The fit to surrounding ESP when using MK/CHELPG/RESP charges fitted for the minimum energy conformation of a) dimethylethylamine b) propanal c) propionicacid. The different conformations are demonstrated by the ball-and-stick models one the right depicting the global minimum conformation with a brighter shade and the second minimum conformation with a darker shade (table 3.1).

Appendix B

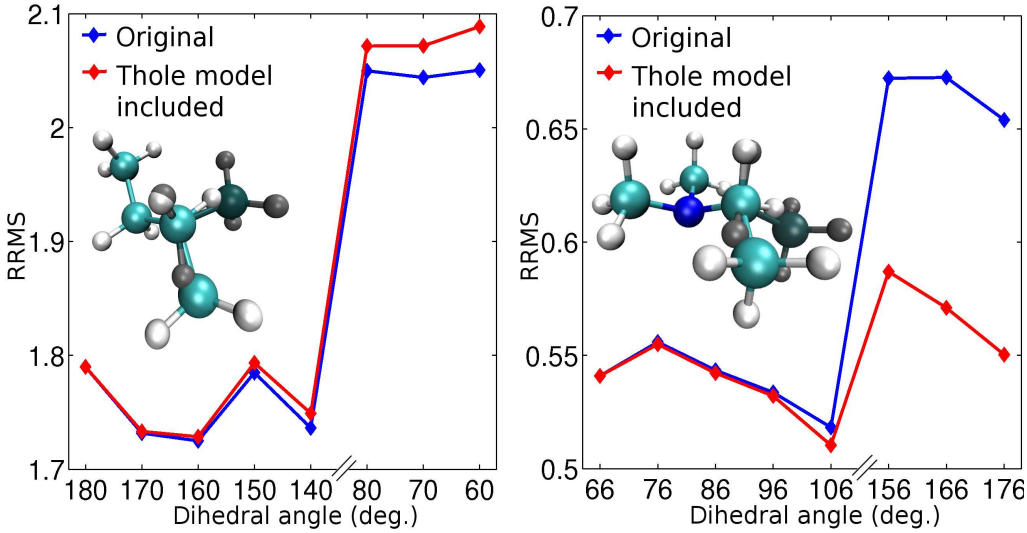
Additional data for the performance of DMA

In this section additional data for the performance of the DMA minimum energy conformation parameters is presented. The molecules covered here are methyl ethyl ether (fig. B.1(a)), methyl formate (fig. B.1(b)), butane (fig. B.2(a)), dimethylethylamine (fig. B.2(b)), and propanal (fig. B.2(c)). The analysis done to obtain this data is further elaborated in the section 4.2.2 of this work.

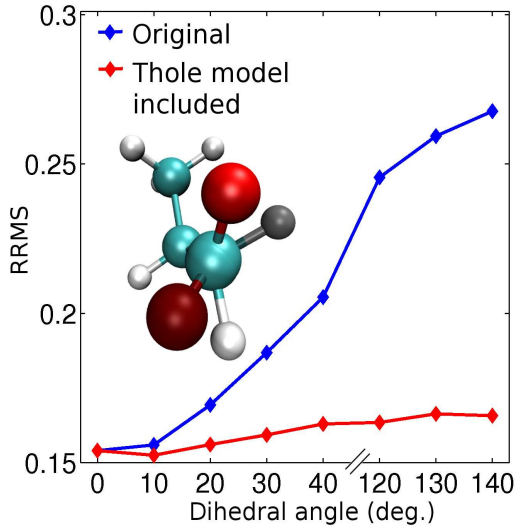


(a) Methyl ethyl ether (b) Methyl formate

Figure B.1: The fit to surrounding ESP when using both polarizable and non-polarizable DMA parameters fitted for the minimum energy conformation of the molecule. Results are presented for a) methyl ethyl ether b) methyl formate. The RRMS is calculated in the MK grid.



(a) Butane (b) Dimethylethylamine



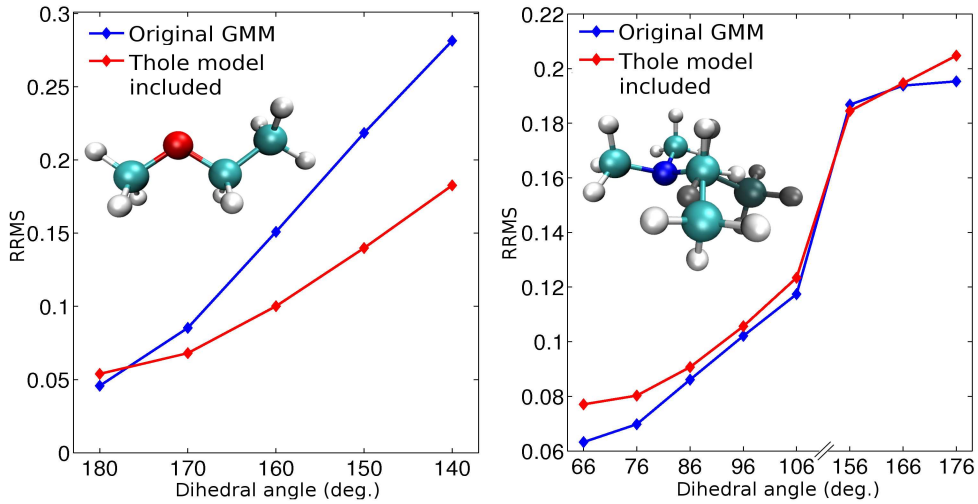
(c) Propanal

Figure B.2: The fit to surrounding ESP when using both polarizable and non-polarizable DMA parameters fitted for the minimum energy conformation of the molecule. Results are presented for a) butane b) dimethylethylamine c) propanale. The RRMS is calculated in the MK grid.

Appendix C

Additional data for the performance of GMM

Here the data for the performance of the GMM minimum energy conformation parameters is presented for the rest of the test molecules for which ΔESP iteration converged. The molecules covered here are methyl ethyl ether (fig. C.1(a)) and dimethylethylamine (fig. B.2(b)). The analysis done to obtain this data is further elaborated in the section 4.3.3 of this work.



(a) Methyl ethyl ether (b) Dimethylethylamine

Figure C.1: The fit to surrounding ESP when using both polarizable and non-polarizable GMM parameters fitted for the minimum energy conformation of a) methyl ethyl ether b) dimethylethylamine.