Distributed Multipole Moments in Atomistic Force Fields: Implementation and Applications

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch–Naturwissenschaftlichen Fakultät

der Universität Basel

von

Nuria Selina Plattner

aus Frauenfeld, TG

Basel, 2009

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von:
Prof. Dr. Markus Meuwly
Prof. Dr. Michele Parrinello

Basel, den 23. Juni 2009

Prof. Dr. E. Parlow   (Dekan)

Acknowledgment

I would like to thank my supervisor Prof. Markus Meuwly for providing me with this project. I appreciated his enthusiasm and the inspiring discussions throughout my PhD study. I enjoyed working on this project.

I would also like to thank Prof. Michele Parrinello who kindly accepted to act as external examiner.

Furthermore, I would like to thank Prof. Sally Price and Prof. Anthony Stone for helpful informations about distributed multipole moments.

It has been a pleasure to work with the members of the Meuwly group. In particular, I appreciated the constructive discussions with Dr. Michael Devereux.

Abstract

The accuracy of atomistic force fields depends on the complexity of the interatomic potential function as well as on the parametrization of the potential. In conventional force fields, the electrostatic potential is represented by atom-centered point charges. Point charges can be understood as the first term of multipole expansions, which converge with increasing number of terms towards the accurate representation of the molecular potential given by the electron density distribution.

Here, the distributed multipole analysis (DMA) is used to obtain atomic multipole moments. The accuracy of distributed multipole potentials is tested for several molecules and compared to point charge potentials. The investigation is focused on convergence of the multipole expansion and conformational dependence. Energies and forces required for molecular dynamics (MD) simulations with atomic multipole potentials are implemented into the CHARMM program. Important points to consider for the implementation are the orientation of the multipole moments and the conformational dependence of multipole parameters.

The implementation is applied to different systems: The splitting of the infrared (IR) absorption band for photodissociated CO in Myoglobin is analyzed comparing different multipole models for CO. A relationship is established between the IR frequency and the CO orientation in the binding pocket. The experimental IR spectrum of CO in amorphous ice is reproduced using multipole potentials for CO and water. The relationship between infrared frequencies and ice structures is analyzed. Furthermore, atomic multipole moments are applied to methane and CO clathrate hydrates. Lattice modes are calculated and compared to experiment. The influence of different guest molecules on lattice modes and structure is characterized.

Contents

Abbreviations

AIM	Theory of atoms in molecules
CAMM	Cumulative atomic multipole moments
CCSD(T)	Coupled-Cluster with single, double and perturbative triple excitations
CHARMM	Chemistry at HARvard Molecular Mechanics
CO	Carbon Monoxyde
DFT	Density functional theory
DMA	Distributed multipole analysis
ff	force field
HF	Hartree-Fock
INS	Inelastic Neutron Scattering
IR	infrared
KKY	Kumagai, Kawamura and Yokokawa
LJ	Lennard-Jones
Mb	Myoglobin
MbCO	Carbonmonoxy Myoglobin
MD	Molecular Dynamics
MM	Molecular Mechanics
MP2	Møller-Plesset second order
NPT	Isothermal-isobaric ensemble
NVE	Microcanonical ensemble
NVT	Canonical ensemble
PMM	Perturbed Matrix Method
QM	Quantum Mechanics
Rank 0	Charge
Rank 1	Dipole
Rank 2	Quadrupole
Rank 3	Octopole
RRKR	rotational Rydberg-Klein-Rees
SCF	Self consistent field
UB	Urey-Bradley
VdW	Van-der-Waals

Units

Å	Ångström
pm	Picometer
nm	Nanometer
a_0	Bohr radius
e	Elementary charge
K	Kelvin
kcal	Kilocalorie
kJ	Kilojoule
fs	Femtosecond
ps	Picosecond
ns	Nanosecond
cm^{-1}	Wavenumber

x

1 Introduction

"True realism consists in revealing the surprising things which habit keeps covered and prevents us from seeing."

Jean Cocteau, *Le Mystère Laïc*, 1928

1.1 Atomistic force fields

Atomistic force fields[1–4] are an important tool for understanding the interactions and dynamical processes in complex molecular systems relevant to different areas of science. In biology, atomistic force fields are e.g. used to simulate the motion of proteins and to model protein-ligand and protein-protein interactions.[5,6] In physics, properties of condensed matter are analyzed using force field models, for example the structure of ices and crystals.[7]

The concept of atomistic force fields consists in an approximation of the interactions between atoms. A popular strategy is to divide the interaction into bonded and non-bonded terms, where the bonded terms include bonds, angles and dihedrals and the nonbonded terms include electrostatics described by point charges and a Lennard-Jones(LJ) potential to represents dispersion and Pauli repulsion. The force field potential according to this scheme is resumed in Equation 1.1.

$$V_{ff} = V_{bonds} + V_{angles} + V_{dihedrals} + V_{elec} + V_{LJ} \tag{1.1}$$

This approximation makes simulations of large systems like proteins possible due to its low computational cost.

With the development of specialized computer clusters and suitable algorithms for parallelization, it is now possible to perform atomistic MD simulations for systems of

relevance to biology of physics up to the microsecond time scale.[8] One of the limitation of these simulation is the inaccuracy of the available force fields.[9,10] Therefore it is important to evaluate possible improvements for the current force fields.

1.2 The CHARMM force field

The main instrument of the present work is the CHARMM program.[11,12] This program includes an empirical force field potential[1] conceptually corresponding to Equation 1.1. In detail, the potential function is composed of the following terms:

$$V_{ff}(\vec{r}) = \sum_{bonds} K_b(b - b_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 + \sum_{UB} K_{UB}(S - S_0)^2$$
$$+ \sum_{dihedrals} K_\chi(1 + cos(n\chi - \delta)) + \sum_{impropers} K_{imp}(\phi - \phi_0)^2$$
$$+ \sum_{nonbond} \epsilon((\frac{R_{min_{ij}}}{r_{ij}})^{12} - (\frac{R_{min_{ij}}}{r_{ij}})^6) + \sum_{nonbond} \frac{q_i q_j}{r_{ij}}, \qquad (1.2)$$

where K_b, K_θ, K_{UB}, K_χ and K_{imp} are the bond, angle, Urey-Bradley(UB), dihedral and improper force constant. The corresponding bond length, angle, UB distance, dihedral angle and improper torsion angle are represented by b, θ, S, χ and ϕ, with the subscript zero representing the equilibrium value. ϵ is the LJ well depth, R_{min} the distance at the LJ minimum and q_i the atomic partial charge. For the nonbond interactions, the sum runs in principle over all pairs of atoms which are more than two bonds apart or belonging to different molecules. However, the details of the nonbond list depend on the system and the method chosen for the calculation. For periodic systems, there are two possibilities of handling the nonbonded interactions. The faster but less accurate method is to truncate the interactions at a defined distance which needs to be smaller than half the simulation cell size in order to avoid interactions of an atom with its periodic image. The more accurate method is to use Ewald summation.[13]

Compared to the simplified terms in Equation 1.1, atom interactions two bonds apart are described by two terms, namely the angle term and the UB term. Atom interac-

tions three bonds apart are described by improper and dihedral terms. This shows that the functional form used for this force field is already quite elaborate, however compared to the full quantum mechanical description of the interatomic potential, this is a rough approximation.

The accuracy of this approximation depends in the first place on the values chosen for the different parameters, i.e. force constants and partial charges. Force field parameters can be obtained from *ab initio* calculations or by fitting of calculated system properties to experimental data. The evaluation and improvement of parameters is an important task in the force field development, but even with an optimal set of parameters, the accuracy of the force field remains limited by the functional forms which describe the potential.

1.3 The quantum mechanical description

Describing interatomic interactions corresponds quantum mechanically to the problem of solving the time independent Schrödinger equation[14] for a system of M nuclei and N electrons, which in Dirac notation[15] is given by

$$\hat{H} \mid \Psi \rangle = E \mid \Psi \rangle, \tag{1.3}$$

where $\mid \Psi \rangle$ and E are the wavefunction and the energy of the system and $\hat{H}$ is given by

$$\hat{H} = -\sum_{i=1}^{N} \frac{1}{2} \nabla_i^2 - \sum_{A=1}^{M} \frac{1}{2M_A} \nabla_A^2 - \sum_{i=1}^{N} \sum_{A=1}^{M} \frac{Z_A}{r_{iA}} + \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{1}{r_{ij}} + \sum_{A=1}^{M} \sum_{B>A}^{M} \frac{Z_A Z_B}{R_{AB}}, \tag{1.4}$$

with r_{iA}, r_{ij} and r_{AB} being the distance between an electron and a nucleus, the distance between two electrons and the distance between two nuclei, respectively. Z_A is the atomic number of nucleus A, M_A is the ratio of the mass of nucleus A to an electron, ∇_i^2 and ∇_A^2 are Laplacian operators for the differentiation of the positions.[16]

The equation is usually solved by separating nuclear and electronic wavefunctions based on the Born-Oppenheimer approximation.[17] This approximation is also the ba-

sis of *ab initio* MD simulations.[18,19] For these simulations, the electronic Schrödinger equation is solved at each step and the nuclei are then propagated classically on the potential energy surface of a given electron density distribution. Obviously these simulations are more accurate than the empirical force field approach in Equation 1.1, but they are computationally more demanding since they typically scale with the third power of the number of electrons.

The computational effort can be reduced if the *ab initio* accuracy is only needed for one part of the system and an empirical potential is sufficient for the rest of the system. The system can in this case be divided into a quantum mechanical and a molecular mechanical part, which interact through specific linking potential terms. This proceeding is called a quantum mechanics/molecular mechanics (QM/MM) approach.[20,21]

1.4 Improvement of force field accuracy

From Equations 1.1 and 1.2, it is obvious that some types of intermolecular forces are missing. Compared to a systematic classification of long-range intermolecular interactions into electrostatics, induction and dispersion,[22,23] the electrostatics in Equation 1.2 are represented by a simple coulomb potential and the dispersion is approximated by the attractive part of the LJ potential, whereas the induction is completely missing.

Based on this observation, two types of force field improvement can be distinguished:

A) Improvement of the functional forms of the potential functions for existing types of interactions. For the bonded terms, this can be done by replacing the simple harmonic potential by more realistic anharmonic functional forms. For the nonbonded terms, the point charges used for the electrostatic interaction can be replaced by a series of atomic multipole moments.

B) Addition of new terms to describe missing types of interatomic interactions. Missing types of interatomic interactions are e.g. the change of electrostatics either with molecular geometry or induced multipole moments, i.e. inter- and intramolecular polarization.

Adding more terms to the force field potential increases the number of parameters which have to be fitted to experimental data or derived from higher level calcula-

tions. Fitting procedures become generally more involved with increasing parameter space and require a larger amount of experimental reference data in order to be unambiguous. Therefore it is an advantage to have a defined procedure for deriving parameters for new potential functions from higher level calculations. Furthermore it is helpful if a new potential function can be associated separately to an observable. This allows to evaluate the quality of parameters for this potential function independent of other terms in the force field potential.

Many efforts for force field improvement in the last decade have been dealing with intermolecular polarization.[24] This strategy is justified by the fact that polarizability is completely missing among the force field terms, therefore it is certainly an improvement to include it. On the other hand, the energy contribution due to polarizability is for many systems significantly smaller than the electrostatics and decays faster with distance. This is particularly true for biological systems, which are mainly composed of the first row elements carbon, oxygen and nitrogen as well as hydrogen, which all have relatively small polarizabilities.

The importance of intermolecular polarization in force fields is related to the question how accurate the current representation of force field electrostatics is. If its already accurate, the addition of polarizabilities is obviously the next improvement step. If its very inaccurate, the addition of polarizabilities will be problematic, especially if the parameters are fitted to experimental data and therefore also attempt to compensate errors in the electrostatics. For this reason, accurate electrostatics are required for a meaningful description of polarizability. In the first place this means that the accuracy of the electrostatics has to be analyzed for a given problem and compared to the gain in accuracy by addition of polarizability.

The accuracy of the electrostatics depends on how appropriate the current functional forms are and how important the change of electrostatics with molecular conformation is. Evaluations concerning the accuracy of point charge models and of the change of electrostatics with molecular conformation can be found in Section 2.

If both, improved electrostatics and polarizability are found to be important for a given problem, it is possible to combine the improvement of electrostatics by distributed multipole moments with polarization terms. The theory of distributed po-

larizabilities[25] could be used for a rigorous treatment of polarizabilities in such a combined approach, but also other concepts may be possible.[26,27]

1.5 Atomic multipole moments

1.5.1 Population analysis methods

Electronic structure calculations[16,28] provide accurate wave functions for considerably large molecules. Integration over the probability distribution function for finding any electron in a given volume $d\vec{r}$ gives the electron density $\rho(\vec{r})$ for that volume element. The separation of the molecular electron density into atomic electron populations has been developed as a tool to analyze the molecular wave function.[29] First population analysis methods only considered the charge population of spatial volume elements or molecular orbitals. As later shown by Stone this procedure is incomplete. A quantitative reproduction of molecular properties like molecular multipole moments requires the inclusion of higher multipole moments into the population analysis.[30]

Point charges can be understood as a first term in a multipole expansion.[30,31] The potential which corresponds to the multipole expansion converges towards the potential derived from $\rho(\vec{r})$ on which the analysis is based. Several algorithms have been developed for obtaining atom-centered multipole expansions. The most widely used ones are the Distributed Multipole Analysis (DMA)[32], the Cumulative Atomic Multipole Moments (CAMM)[31] and the Theory of Atoms in Molecules (AIM)[33]. All methods include a partitioning of the electron density in real or molecular orbital space and an integration of $\rho(\vec{r})$ over a suitably defined space for all desired multipole ranks. The partitioning of $\rho(\vec{r})$, the allocation of atomic moments to particular atoms and therefore the resulting multipole expansion, its convergence, basis set dependence, conformational dependence and transferability differ between the three methods.

In the following, the DMA algorithm will be used to obtain atomic multipole expansions. Details of this method and the properties of the multipole potentials are explained in Section 2.

1.5.2 Force field electrostatics based on atomic multipole moments

As shown in Equation 1.2, the electrostatics in atomistic force fields are usually represented by point charges. These charges are typically obtained from a population analysis, or by fitting atomic charges which optimally reproduce the electrostatic potential or interaction energies.[34,35]

The potential function for the electrostatic energy based on atomic multipole moments is more complicated and computationally demanding than the simple coulomb interaction between point charges. As an example, the electrostatic potential of a multipole expansions up to quadrupole on two atoms a and b in spherical tensor notation can be written as

$$E_{elec} = E(q_a q_b) + \sum_{k=1}^{3} E(q_a \mu_b^k) + \sum_{l=1}^{3} E(\mu_a^l q_b)$$

$$+ \sum_{k=1}^{3} \sum_{l=1}^{3} E(\mu_a^l \mu_b^k) + \sum_{m=1}^{5} E(q_a \Theta_b^m) + \sum_{n=1}^{5} E(\Theta_a^n q_b) +$$

$$\sum_{k=1}^{3} \sum_{m=1}^{5} E(\mu_a^k \Theta_b^m) + \sum_{n=1}^{5} \sum_{l=1}^{3} E(\Theta_a^n \mu_b^k) + \sum_{m=1}^{5} \sum_{n=1}^{5} E(\Theta_a^n \Theta_b^m), \qquad (1.5)$$

where q_a i s the charge of atom a, μ_b^k is the direction k of the dipole moment vector on atom b, Θ_a^n is the element n of the quadrupole tensor on atom a.[22]

Since the accuracy of the multipole expansion can be gradually improved by adding higher multipole ranks, atomic multipole moments provide a possibility to systematically improve the accuracy of force field electrostatics. Evaluations concerning the rank of the multipole expansion required to considerably improve the accuracy of force field electrostatics are shown in Section 2.

2 Distributed multipole potentials

"Isolated material particles are abstractions, their properties being definable and observable only through their interaction with other systems."

Niels Bohr, *Atomic Physics and the Description of Nature*, 1934

In this chapter, the details for obtaining atomic multipole moments based on the distributed multipole algorithm are explained. Evaluations are carried out to illustrate the properties of distributed multipole potentials. They are focused on the convergence of the multipole expansion, the accuracy compared to *ab initio* and to point charges and the conformational dependence of multipole-derived potentials. Important points to consider for the use of atomic multipole moments as force field parameters are summarized and discussed in Section 2.5.

Parts of this chapter have been published in

"Higher order multipole moments for molecular dynamics simulations"
N. Plattner and M. Meuwly, **J. Mol. Model.** (2009) 15:687 − 694.

2.1 Distributed multipole analysis

For a closed shell system in the Hartree-Fock approximation, the electron density $\rho(r)$ can be expressed as

$$\rho(r) = \sum_{i=1}^{K} \sum_{j=1}^{K} P_{ij} \phi_i(r) \phi_j^*(r), \tag{2.1}$$

with K being the number of basis functions. For a given set of basis functions ϕ_i, the density matrix elements P_{ij} specify completely the charge density.[16] P_{ij} is given by

$$P_{ij} = 2 \sum_{m=1}^{N/2} C_{im} C_{jm}^*, \tag{2.2}$$

with C_{xm} being the coefficients of the basis functions and N being the number of electrons. In most modern electronic structure calculation programs, the basis set is composed of Gaussian functions

$$\phi_i^A = f_i(r - A) e^{-\alpha(r-A)^2}, \tag{2.3}$$

where $f_i(r - A)$ is a homogeneous polynomial, α determines the width of the function and A is the center of the Gaussian function, which is usually a nuclear position. The product of two Gaussian functions is itself a Gaussian function. The new function is centered on a line joining the initial centers and surrounded by the overlap charge density. The electron density is therefore expanded about as many centers as there are overlapping Gaussian functions. Based on the electron density, the multipole population can be evaluated in terms of multipole moments

$$Q_{kq}(A) = - \int R_{kq}(r - A) P_{ij} \phi_i^A \phi_j^B dv, \tag{2.4}$$

where q is a given multipole component(direction), k is the multipole rank and $R_{kq}(r)$ is a regular solid harmonic.

The resulting multipole expansions are distributed over all centers of overlap of the initial basis functions. The number of expansions can be reduced by transferring the multipole moments to fewer sites. A multipole expansion about any point A can be transformed into a multipole expansion about any other point B using

$$Q_{lm}(B) = \sum_{k=0}^{l} \sum_{q=-k}^{k} [\binom{l+m}{k+q}\binom{l-m}{k-q}]^{1/2} Q_{kq}(A) R_{l-k,m-q}(B-A), \qquad (2.5)$$

where k and q are ranks and components of the initial multipole expansion, l and m are the corresponding ranks and components of the final multipole expansion. It follows that a new multipole component of rank l on site B depends on the multipole moments of lower and equal rank on the initial site A.[22,30,32]

The new points for the multipole expansion can in principle be chosen arbitrarily. In the special case where all moments are transferred to the center of the molecule, the resulting multipole expansion corresponds to the molecular multipole moments. For atomistic force fields, the most convenient choice is to use the nuclear positions as centers for the expansions. In order to reduce the computational effort for the force field calculations, the number of expansion centers has to be reduced as much as possible. However, the convergence of the multipole expansion is improved with increasing number of multipole sites. This is due to the fact that for an accurate reproduction of the potential of the initial electron density distribution, it is important that the distance over which the multipole moments have been transferred is considerably smaller than the distance between the multipole sites and interacting charges or multipoles. This leads in practice to a tradeoff between computational cost and accuracy. Evaluations concerning this tradeoff can be found in Section 2.2.

The allocation of the electron density and therefore of the multipole expansions to the atoms is important for the properties of the resulting expansion. For the first version of the DMA algorithm,[30] it was proposed that the overlap densities should always be allocated to the site closest to the initial overlap center of the Gaussian functions, so in practice to the closest nucleus. For the CAMM algorithm,[31,36] half of the multipole population is allocated to each of the sites from which the Gaussians originate. The AIM algorithm[33,37] uses the gradient vector field of the charge density to partition a molecule into atoms, for which the corresponding atomic multipole moments are then evaluated by integrating over the volume space corresponding to this atom.

As larger basis sets became computationally affordable, it was found that for the basis functions with large exponents, a partitioning of the electron density in real space and not in molecular orbital space may be preferable. This partitioning was implemented

into the DMA algorithm using Lebedev spherical grid integration.[38,39] The new DMA algorithm now combines allocation of multipoles to the nearest site to allocation of electron density in volume space. The former method is applied to Gaussian functions with low exponents, the latter to Gaussian functions with high exponents. The partitioning of the electron density through space was found to improve the convergence of the multipole expansions for larger basis sets.[39] One of the consequences of this new algorithm is that the final multipole expansion depends on the choice of the relative integration radius for each atom. For converged multipole series, the choice of integration radii does not considerably affect the overall potential of the multipole expansion. For truncated multipole expansions in contrast, the choice of integration radii is important, as explained in Section 2.2.
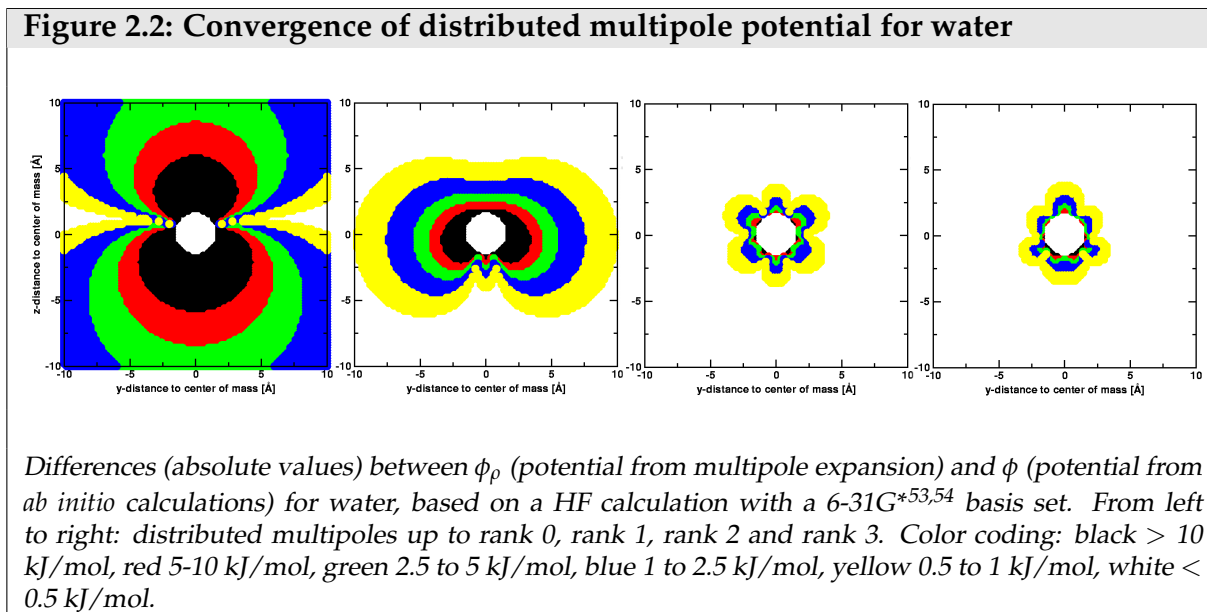
The distributed multipole analysis can be applied to electron densities obtained from different electronic structure calculation methods. The only limitation is that the basis functions need to be Gaussians. In the following, distributed multipole moments will be calculated based on densities obtained from restricted Hartree-Fock(HF),[40–44] Møller-Plesset second order(MP2)[45] and density functional theory(DFT)[46,47] calculations.

2.2 Accuracy and convergence

Important criteria for the quality of a multipole expansion are the accuracy compared to the potential of the underlying electronic structure calculation and the convergence of the multipole expansion with increasing multipole rank. Both of these properties can be evaluated based on interaction potentials between two molecules, as found in the literature,[48–51] or by placing a probe charge on grid points around the molecule and comparing the resulting potential at each point to the *ab initio* potential. The latter approach is chosen here. The region inside the Van-der-Waals radii is excluded for the following evaluations, since for the force field calculations the interaction there is dominated by the repulsive part of the LJ potential. The LJ half radius parameters of the CHARMM force field[1] are used to define the region excluded from the evaluations around each atom. The choice of LJ half radii instead of full radii is due to the fact that in CHARMM, full radii are calculated as the sum of the half radii of two interacting atom types, therefore their values vary significantly between different

interacting atom pairs. The half radii represent the region which is excluded for all interacting atom types. All of the following evaluations are carried out with the GDMA program,[39] based on electron densities from *ab initio* calculations with GAUSSIAN.[52]

As a first example water is chosen. Figure 2.2 shows a graphical representation of the difference between potential ϕ_ρ (derived from the multipole expansion) and the *ab initio* potential ϕ for water with different multipole ranks.

Figure 2.2: Convergence of distributed multipole potential for water

Differences (absolute values) between ϕ_ρ (potential from multipole expansion) and ϕ (potential from ab initio calculations) for water, based on a HF calculation with a 6-31G*[53,54] basis set. From left to right: distributed multipoles up to rank 0, rank 1, rank 2 and rank 3. Color coding: black > 10 kJ/mol, red 5-10 kJ/mol, green 2.5 to 5 kJ/mol, blue 1 to 2.5 kJ/mol, yellow 0.5 to 1 kJ/mol, white < 0.5 kJ/mol.

The analysis shows a good agreement to *ab initio* at the level of about quadrupole. This is the case for most of the tested systems.

For the use of distributed multipoles as force field parameters, it is important to know where the multipole series has to be truncated in order to get an optimal tradeoff between accuracy and computational cost. In this context, not only the accuracy for expansions up to a given multipole rank, but also the consequences of transferring parts of the population to fewer sites and the influence of the integration radii(see Section 2.1) should be tested. The consequences of the choice of integration radii on the water multipole moments are shown in Table 2.1. The consequences of integration radii and population shifts on the accuracy are shown in Table 2.2. The comparison of accuracies to *ab initio* in Table 2.2 is carried out in a cubic field of $10 \times 10 \times 10$ Å with 101 grid points in each direction .

13

Table 2.1: **Distributed multipole moments*****for different integration radii based on MP2 aug-cc-PVQZ**[55] **density for water**

atom name	relative integration radius			
charge[e]	q00			
dipole [ea_0]	q10 q11s			
quadrupole [ea_0^2]	q20 q21s q22c			
octopole [ea_0^3]	q30 q31s q32c q33s			

	Oxygen 0.8				**Hydrogen** † 0.2			
q	-1.153898				0.576952			
μ	0.334912	0.0			0.107979	-0.114351		
Θ	0.023976	0.0	-0.682931		-0.017299	0.019942	-0.012388	
Ω	0.408773	0.0	1.233701	0.0	0.041747	0.034621	0.103666	-0.076388
	Oxygen 0.7				**Hydrogen** 0.3			
q	-0.610304				0.30515			
μ	-0.213259	0.0			0.079759	-0.061051		
Θ	0.136491	0.0	-1.149071		-0.001955	-0.106929	-0.128946	
Ω	0.503113	0.0	1.262078	0.0	0.049096	0.077211	0.151472	-0.099292
	Oxygen 0.6				**Hydrogen** 0.4			
q	0.065351				-0.032676			
μ	-0.710046	0.0			-0.047583	0.085952		
Θ	0.101827	0.0	-1.341222		0.014648	-0.190347	-0.267331	
Ω	0.451977	0.0	0.573517	0.0	-0.008054	-0.027306	-0.036769	0.048199

The comparison of the multipole moments for the different integration radii shows that the multipole values at each level vary a lot for the different radii, e.g. the charges on oxygen change from 0.065351 to -1.153898 e. However, the differences are compensated by higher multipole ranks, as the dipole moment on oxygen changes direction and magnitude from 0.334912 to -0.710046 ea_0. The comparison of accuracies for different integration radii (Table 2.2) shows that this compensation is almost complete at rank 3, where all expansions reproduce the *ab initio* potential with deviations of only a few percent.

---

*Not all spherical tensor elements are populated in water due to symmetry reasons
†For the second hydrogens, signs for q11s, q21s and q31s are opposite.

Table 2.2: **Accuracy of truncated distributed multipole expansions with different integration radii based on MP2 aug-cc-PVQZ density for water**

atomic radii[‡]		Rank 1	Rank 0 / 2 [§]	Rank 1 / 2 [¶]	Rank 2	Rank 3
0.8 : 0.2	$\Delta\overline{E}$ [kJ/mol]	1.5974	0.3921	0.5674	0.5901	0.3779
	$\Delta\overline{E}$ [%]	46.38	7.11	8.93	9.26	3.89
0.7 : 0.3	$\Delta\overline{E}$ [kJ/mol]	3.1943	0.6039	1.0681	0.6252	0.3603
	$\Delta\overline{E}$ [%]	93.11	9.19	17.93	9.39	3.50
0.65 : 0.35	$\Delta\overline{E}$ [kJ/mol]	3.7628	1.0832	1.1547	0.5188	0.3525
	$\Delta\overline{E}$ [%]	111.34	17.85	19.45	7.18	3.94
0.6 : 0.4	$\Delta\overline{E}$ [kJ/mol]	4.1735	1.6365	1.1172	0.4185	0.3436
	$\Delta\overline{E}$ [%]	126.52	27.98	18.54	5.46	4.19
0.5 : 0.5	$\Delta\overline{E}$ [kJ/mol]	4.5163	2.8382	0.6966	0.6879	0.3248
	$\Delta\overline{E}$ [%]	146.98	49.81	9.44	10.45	4.33

While the results are very similar for all atomic radii at rank 3, the differences in accuracy are significant for multipole expansion truncated at rank 2 or rank 1. This is true for differences in integration radii as well as for shifting the multipole populations to fewer sites. If the expansion is truncated at the quadrupole level, for some integration radii it is equally accurate to use the quadrupole moment only on the oxygen molecule as to use the full expansion. As a consequence, the choice of the integration radii is important for truncated multipole expansions, but with the increase of the highest rank, the differences between potentials based on different radii vanish.

[‡]Relative radii for integration of Gaussian functions with exponents larger than 4. Oxygen:Hydrogen
[§]All moments higher than charges are transferred to the oxygen
[¶]All moments higher than dipole are transferred to the oxygen

2.3 Comparison to point charge potentials

From the evaluation in the last section, it is obvious that the accuracy of distributed multipoles up to quadrupole is significantly higher than the accuracy of charges obtained by the DMA algorithm. This accuracy gain should not be interpreted as the gain in accuracy compared to point charge potentials in general. For the parametrization of point charges in force fields, other algorithms are used, either the Mulliken population analysis[29] or fitting algorithms[35] which directly fit atomic charges in order to optimally reproduce the electrostatic potential of a molecule.

A comparison of the DMA accuracy compared to these more appropriate methods for obtaining point charges is shown in Table 2.3 for water and alanine. The evaluation is carried out with a probe charge of 1 e, on a cubic grid of $10 \times 10 \times 10$ Å for water and $15 \times 15 \times 15$ Å for alanine, with 101 grid points in each direction.

Table 2.3: **Point charge and multipole accuracies for water and alanine compared to *ab initio***

	HF 6-31G*		B3LYP[56,57] aug-cc-pVTZ	
	Water	*Alanine*	*Water*	*Alanine*
DMA, rank 0				
$\Delta\overline{E}$ [kJ/mol]	3.27	2.87	2.70	2.52
$\Delta\overline{E}$ [%]	64	2383	68	1653
DMA, rank 1				
$\Delta\overline{E}$ [kJ/mol]	0.63	0.333	0.71	0.36
$\Delta\overline{E}$ [%]	41	178	61	73
DMA, rank 2				
$\Delta\overline{E}$ [kJ/mol]	0.05	0.05	0.17	0.06
$\Delta\overline{E}$ [%]	1	6	4	8
Mulliken Charges				
$\Delta\overline{E}$ [kJ/mol]	0.52	0.84	2.18	6.05
$\Delta\overline{E}$ [%]	23	687	60	2993
Merz-Kollman				
$\Delta\overline{E}$ [kJ/mol]	0.28	0.08	0.44	0.08
$\Delta\overline{E}$ [%]	18	35	37	18

For both molecules and electronic structure calculation methods, the distributed multipole expansion converges at rank two. ϕ_ρ and ϕ differ by a few percent on average at

rank two, which makes the multipole expansion significantly more accurate than potentials from point charges for the same molecule, using Mulliken charges or charges fitted according to the Merz-Kollman scheme.[35]

2.4 Conformational dependence

To show the conformational dependence of multipole-derived potentials, two conformations (c1 and c2) of alanine and glycine, are considered. For glycine, distributed multipole moments for each of the two conformations are derived and compared to the *ab initio* potential of both conformations. The results are shown in Figure 2.4. Coordinates and multipole moments corresponding to this evaluation can be found elsewhere.[58]

The example in Figure 2.4 illustrates that the differences between ϕ and ϕ_ρ are considerably larger if the multipoles are transferred to another conformation. A quantitative analysis of the differences between ϕ_ρ and ϕ for glycine and alanine is provided in Table 2.4, where differences between the electrostatic potentials from *ab initio* calculations and different charge models are reported.

In order to calculate the data for another conformation, the multipoles are rotated according to the geometry of the atoms in the new conformation. This procedure corresponds to the assignment of local reference axis systems explained in Section 3. Additionally, the tables show data for the conformational dependence of point charges, which are calculated according to the Merz-Kollmann scheme[35]. Also for point charges, the deviations in another conformation are considerably larger. The data show that the conformational dependence of ϕ_ρ has to be taken into account. Otherwise the accuracy gained by using multipole expansions instead of point charges is likely to be compromised.
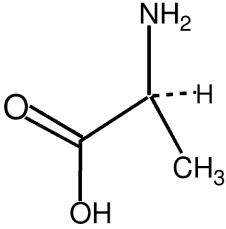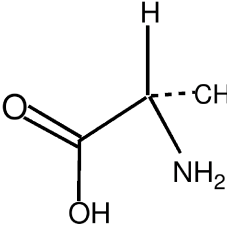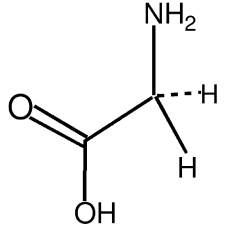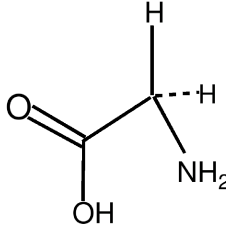
A setup with fewer multipole sites and integration radii chosen such that the conformational dependence is minimized is also shown in Table 2.4. The accuracy is improved by this approach, but the errors are still larger than for the moments calculated from the same conformation.

Figure 2.4: Conformational dependence of distributed multipole potentials
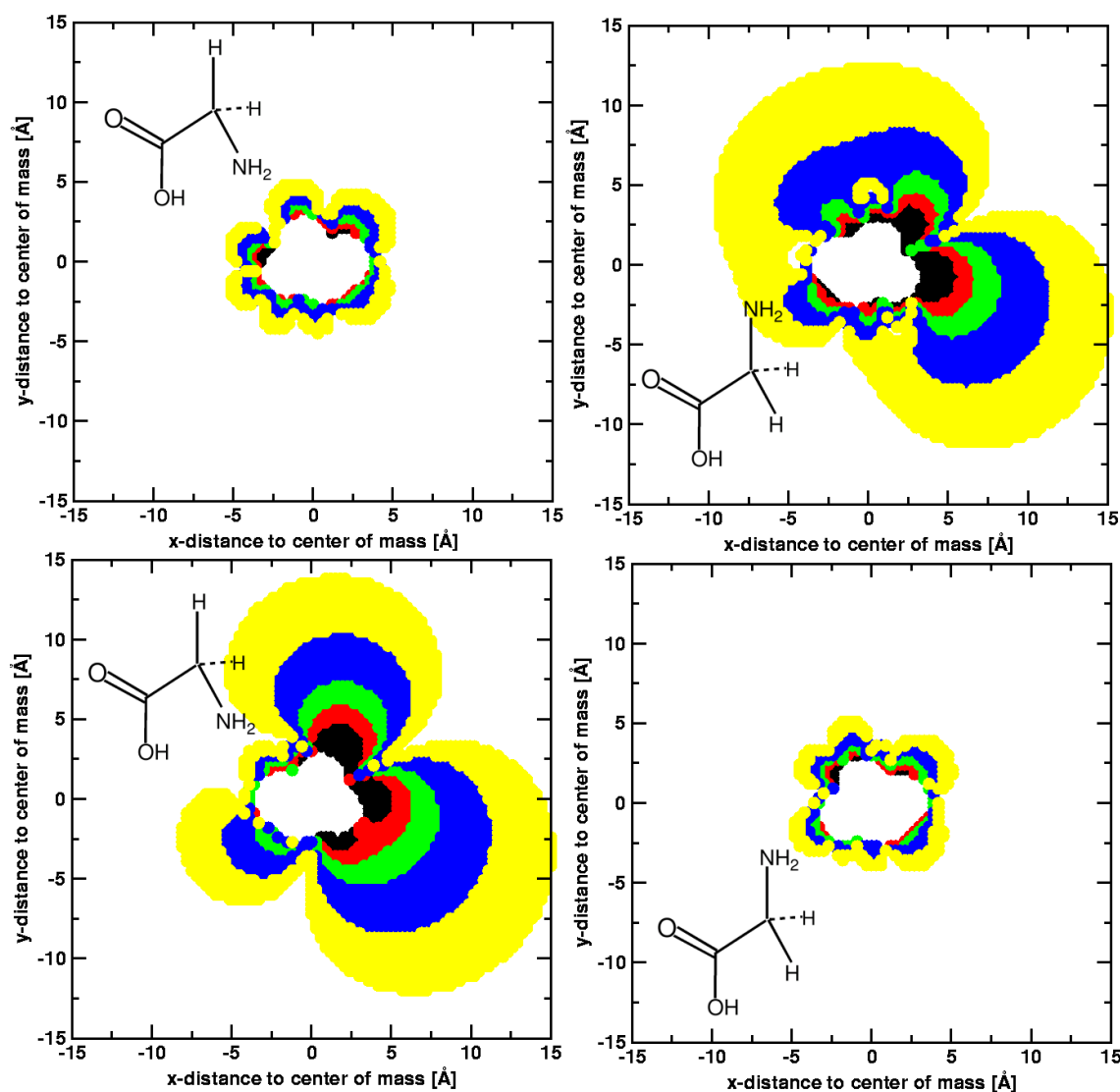
Differences (absolute values) between multipole expansion and ab initio potential for glycine in two different conformations based on HF 6-31G* calculations. Upper row: multipole moments calculated from conformation 1 of glycine. Lower row: multipole moments calculated from conformation 2 of glycine. Left column: comparison to ab initio potential of conformation 1. Right column: comparison to ab initio potential of conformation 2. Color coding: black > 10 kJ/mol, red 5-10 kJ/mol, green 2.5 to 5 kJ/mol, blue 1 to 2.5 kJ/mol, yellow 0.5 to 1 kJ/mol, white < 0.5 kJ/mol. The x−axis of the second conformation is inverted in this representation in order to allow better comparison of the two conformations.

Table 2.4: Conformational dependence of glycine and alanine

Alanine conformations

c1   c2

Glycine conformations

c1   c2

	Glycine		Alanine		Glycine		Alanine	
	c1	c2	c1	c2	c1	c2	c1	c2
	HF 6-31G**		HF 6-31G**		B3LYP cc-pVQZ		B3LYP cc-pVTZ	
Charges of c1								
$\Delta\overline{E}$ [kJ/mol]	0.09	0.53	0.07	1.03	0.11	0.44	0.08	1.05
$\Delta\overline{E}$ [%]	21	60	18	330	29	39	35	1317
Charges of c2								
$\Delta\overline{E}$ [kJ/mol]	0.17	0.15	0.75	0.08	0.17	0.15	0.47	0.10
$\Delta\overline{E}$ [%]	55	34	418	13	47	30	613	73
DMA of c1								
$\Delta\overline{E}$ [kJ/mol]	0.07	0.26	0.10	1.27	0.07	0.19	0.11	1.11
$\Delta\overline{E}$ [%]	4	30	11	295	5	25	15	829
DMA of c2								
$\Delta\overline{E}$ [kJ/mol]	0.16	0.07	0.63	0.10	0.14	0.11	0.54	0.11
$\Delta\overline{E}$ [%]	40	4	250	5	36	9	423	7
Setup with 3 Multipole Sites for Glycine, 4 Sites for Alanine								
DMA of c1								
$\Delta\overline{E}$ [kJ/mol]	0.10	0.12	0.10	0.32	0.10	0.13	0.10	0.46
$\Delta\overline{E}$ [%]	6	9	7	77	8	8	11	206

19

2.5 Atomic multipole parameters for atomistic force fields

In contrast to point charges, the values of the atomic multipole expansions do not correspond to chemical intuition. Charges may even be completely counterintuitive, because they are compensated by dipole moments of opposite sign. A quantitative evaluation of the whole expansion either compared to *ab initio* or to experimental values of molecular multipole moments is therefore required. In addition, the evaluation of multipole parameters should consider a number of points:

- The tradeoff between accuracy and computational cost

- Dependence of the multipole moments on the *ab initio* method and basis set

- Conformational dependence of the multipole parameters

- Transferability of parameters for equal fragments in different molecules

The tradeoff between accuracy and computational cost can be evaluated by comparing the accuracy of different parametrization, as shown in Section 2.2. The method and basis set dependence should be taken into account in the sense that appropriate *ab initio* methods and large basis sets should be used for the parametrization. The conformational dependence is the most demanding problem, although it is not important for all systems. For molecules which are too small or too rigid to undergo larger conformational changes such as H_2, water, methane or benzene, the assignment of multipole parameters is straightforward. On the other hand for molecules which undergo significant conformational changes like glycine and alanine, the dependence of the parameters on the molecular conformation has to be taken into account. The conformational dependence of atomic multipole potentials has already been investigated in earlier studies for the DMA, the AIM and the CAMM algorithm.[59–61] Different solutions have been proposed to this problem, not all of which are applicable in general.

For some molecules, the change of the atomic multipole moments with conformation can be described explicitly by an analytical function. This has been done successfully for CO.[62] (See Section 4.1) Except for this particularly simple case, more elaborate methods need to be used for molecules such as glycine and alanine. For the conformational dependence around torsion angles, the use of short Fourier series has been

shown to significantly improve the electrostatic potential.[48,63] A combined solution for atomic multipoles together with inter- and intramolecular polarization has been suggested[64], but no explicit dynamics has been carried out with this procedure and for larger molecules such as amino acids. Furthermore, it is unclear how well this approach is able to reproduce atomic multipole moments of another conformation, since only the combined electrostatic plus polarization energies have been evaluated and the individual contribution of each term to different conformations is not available. Systematic evaluations have been done for a neural network approach to model the change of AIM multipole moments with conformation.[27] For this approach, tests including explicit dynamics are not available yet. Another possible solution is to use conformationally averaged multipole moments to better reproduce the *ab initio* potential on average. The results for shifting the multipole moments to fewer sites and minimizing the conformational dependence shown in Section 2.4 indicates that this approach works for some cases. However, it is probably not a general solution and some of the accuracy of distributed multipole potentials is lost by this procedure. Finally, it may also be possible to use an iterative scheme in which the multipole moments are recalculated periodically. Such an approach has already been proposed and tested for energy minimizations of crystal structures.[65]

For a generalized use of distributed multipole parameters in atomistic force fields, the transferability of multipole moments between equal fragments in different molecules, e.g. for transferring amino acids between different peptides, needs to be considered. Evaluations concerning this problem can be found in the literature.[60,66] They reveal that this issue is less problematic than the conformational dependence of the multipole parameters.

3 Implementation of atomic multipoles into CHARMM

"Science is knowledge which we understand so well that we can teach it to a computer; and if we don't fully understand something, it is an art to deal with it."

Donald Knuth, *Computer Programming as an Art*, 1974

In this chapter, the details of implementing atomic multipole potentials into the CHARMM program[11,12] are explained. This is done in Section 3.1 on a theoretical basis, focusing on the additional difficulties arising from the multipole orientations. The implementation is based on the published terms for orientation-dependent multipole interaction energies.[22,67] Particular care is taken for the details of the first derivatives, which are not found in the literature and need to be formulated in a way suitable for an atomistic force field like CHARMM.

Section 3.2 gives an overview over the current implementation in CHARMM and is supposed to help understanding the multipole module in the context of the CHARMM code. A series of tests is shown in Section 3.3. These tests are supposed to illustrate the problems explained in Section 3.1 and allow comparison to other implementations. Finally in Section 3.4, open questions concerning applications and a more general implementation of atomic multipole moments are discussed.

3.1 Basic concepts

3.1.1 Atomic multipole parameters

As for other force field parameters, the data structure of atomic multipole moments can be organized according to types of equal molecules and residues. In addition, the geometry of the molecule needs to be specified.

The following specifications are required for each molecule:

- The number of multipole sites in the molecule; this specification is linked to the molecular geometry.

- The reference atom number for a prototype molecule (the first in a series of equal molecules)

- The highest multipole rank for each atom

- The values of the multipole parameters on each atom

- The range of atom numbers to which the same type of parameters apply

Based on these details, the multipole parameters are applied to all molecules of the same type in the system. The geometry specification is used to link the orientations of the atomic multipole moments to the molecular orientations. (See Sections 3.1.2 and 3.1.3 for details.) The final data structure combines atom-specific information like the values of the multipole moments to molecule-specific information like the indices of the atoms belonging to this molecule.

As explained in Section 2, the number of multipole moments per molecule can be optimized to reach an optimal tradeoff between accuracy and computational cost. Therefore it is important to specify the rank of the highest multipole moment on each atom and store this information to make the calculation of energies and gradients efficient.

3.1.2 Atomic multipole moments and molecular orientations

The basic problem for the implementation of atomic multipole moments into an atomistic force field is the dependence of the multipole moments on the orientation of the molecule. For a given internal geometry and orientation of a molecule, it is straightforward to calculate energies and forces for multipole moments based on the published terms.[22] The propagation of the atomic multipole moments in time in contrast causes a problem: the atomic multipole moments can not be propagated independently of the molecular geometry, since torques and moments of inertia for independent atoms are physically meaningless. Therefore the propagation of the atomic moments has always to remain linked to the molecular orientation.

In order to solve this problem, two main strategies are possible:

- Use an algorithm based on rigid rotation. For the rigid rotation of molecules, algorithms have already been developed[68] and tested for atomic multipole moments.[69] It may be possible to adapt these algorithms in order to allow for changes in the internal molecular geometry and therefore for vibrational degrees of freedom. However, it is not clear if an integration for pseudo-rigid molecules with continuously changing moment of inertia and torques due to vibrational degrees of freedom would be stable.

- Use atomistic forces and rotate the atomic moments according to the current molecular geometry. This strategy is linked to molecular reference axis systems, which define the orientation of the atomic multipole moments with respect to the molecular geometry.[64] This approach is easier to implement in an atomistic force field, where all the forces are calculated on an atomistic basis. The main problems of this approach arise for the calculation of the derivatives, which in consequence also need to contain terms for all the atoms which define the molecular reference axis system.

For the implementation of atomic multipole moments into the CHARMM program, the latter approach is chosen since it does not require major changes or additions in the structure of the code used for minimization or molecular dynamics. The additional code only needs to provide energies and forces for the atomic multipole interactions.

3.1.3 Molecular reference axis systems

Higher order multipole moments depend on the coordinate system in which they were calculated. For distributed multipoles, the moments obtained from GDMA[39] are defined with respect to the orientation of the molecule in the GAUSSIAN[52] calculation from which they are obtained. This initial orientation corresponds at the same time to an orientation of the corresponding molecule in space.

Figure 3.1.3: Molecular reference axis systems for water
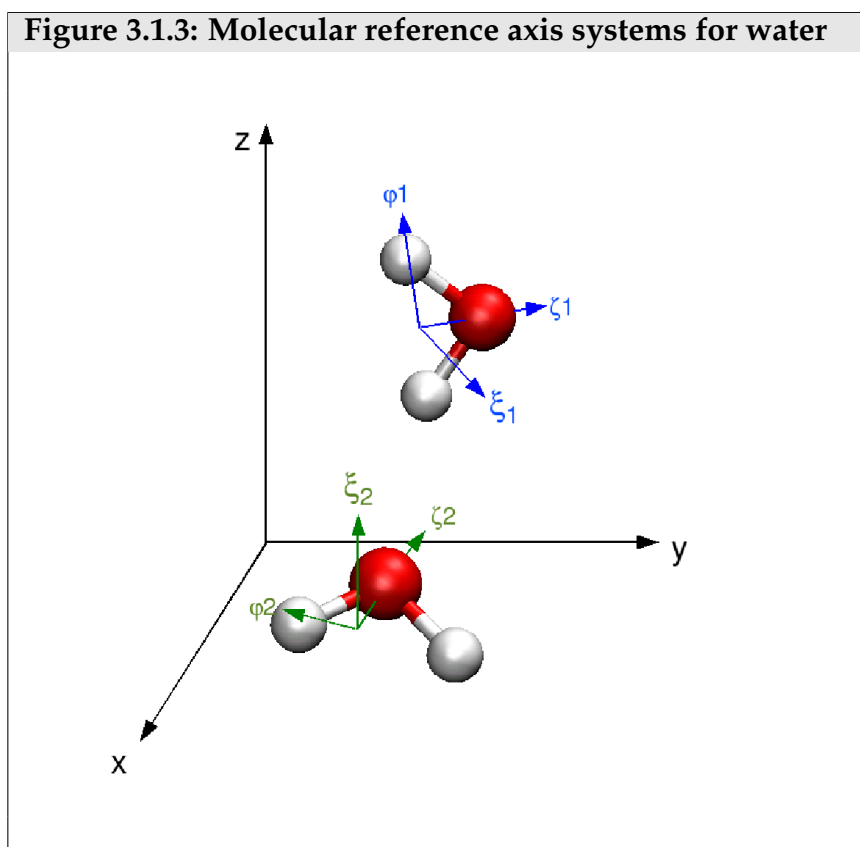
The relationship of molecular and multipole orientation can be maintained by the assignment of the multipole moments to reference axis systems, which are defined by the positions of the atoms. In the simplest case (linear molecule) one reference axis, defined by the two atoms, is sufficient. For water, all three axes have to be defined. For the following calculations, this is done by aligning the $z-$axis along the axis pointing from the oxygen to the center of the two hydrogens, the $y-$axis from one hydrogen to the other hydrogen, and the $x-$axis being then defined by orthogonality (see Figure 3.1.3).

For larger molecules such as glycine and alanine, several reference axis systems have to be defined, since the different parts of the molecule can rotate with respect to each other.

Based on the atom coordinates at each step, a rotation matrix can be calculated for each molecule which defines by how much the corresponding multipole moments or multipole interactions have to be rotated with respect to the global coordinate system. Formally, the rotation can be applied either to the multipole moments or to the orientation dependent part of the interaction. For the actual implementation, the latter approach is chosen. However, it has to be noted, that the consequences of this choice are only the actual organization of the code and probably numerical differences.

For molecules with more than one reference axis system, it is necessary to initially rotate all multipole moments into the corresponding axis system and use this new moments as parameters for the following calculation.

3.1.4 Energies and first derivatives

For the implementation of atomic multipole moments, additional interaction energies and their first derivatives have to be provided. The calculation of the energies is done based on the published terms for atomic multipole moments in spherical tensor notation.[22]

From the calculation of the molecular orientations with respect to the global coordinate system, a nonlinear molecule is described by three unit vectors $\vec{e}_\zeta$, $\vec{e}_\phi$, $\vec{e}_\chi$, corresponding to three orthogonal axes in the molecular reference axis system. If we consider a dipole-dipole interaction for dipoles placed on two atoms a and b belonging to different molecules, A and B, with the dipole orientation being defined as parallel to the ζ-axis of each molecule, the interaction energy is given by

$$E(\mu^a_{\zeta(A)}\mu^b_{\zeta(B)}) = \mu^a_{\zeta(A)}\mu^b_{\zeta(B)}\frac{3r^A_{\zeta(A)}r^B_{\zeta(B)} + c_{\zeta(A)\zeta(B)}}{R^3}, \tag{3.1}$$

where $\mu^a_{\zeta(A)}$ is the dipole moment in direction of $\vec{e}_\zeta$ on molecule A, $r^A_{\zeta(A)}$ is defined by the relative orientation of the two dipole moments and $c_{\zeta(A)\zeta(B)}$ is the scalar product of the two orientational vectors, $\vec{e}_{\zeta(A)} \cdot \vec{e}_{\zeta(B)}$. The magnitude of $r^A_{\zeta(A)}$ is obtained as the scalar product of $\vec{e}_{\zeta(A)}$ and $\vec{r}_{ab}$, with $\vec{r}_{ab}$ being a unit vector pointing from atom a to atom b. R is the interatomic distance. Based on this expression, the interaction energy is defined with respect to the molecular orientation described by the unit vectors of the corresponding molecular reference axes. The higher order terms are more complicated, but there are no new variables. All interactions can be calculated as a function of R, $r^A_{\zeta(A)}$ and $c_{\zeta(A)\zeta(B)}$ given for the atoms in both molecules and for all molecular reference axes.

From the expressions for the interaction energies, we can derive the expressions for the forces, which are the first derivative of the energy with respect to the global x-, y- and z-coordinate. The energy expression contains three types of variables which are dependent on the global coordinates, namely the interatomic distance R, the normalized interatomic separation vector $\vec{r}_{ab}$ and the molecular unit vectors $\vec{e}_\alpha$.

For a rigid molecule with only one central dipole moment and one central LJ potential, the derivative of the energy with respect to $\vec{e}_\alpha$ corresponds directly to the torques on the molecule, i.e. the rotation of the molecule is entirely defined by this derivative.

For atomic moments placed on fully flexible molecules in contrast, the rotation of the molecule and therefore the changes in $\vec{e}_\alpha$ depends on the differences in the translational forces between the atoms in the molecule, where the translational force on each atom is given by the sum of the derivatives for all energy terms as listed in Equation 1.2. The derivative of $\vec{e}_\alpha$ can be formed with respect to the changes of the atom positions which define the reference axis system. However, for the total rotation of the molecule which defines the change of $\vec{e}_\alpha$, the derivative of $\vec{e}_\alpha$ is only a small contribution. Based on these considerations, the derivatives for atomic multipole moments can be defined at two levels of accuracy:

1. Approximate derivatives: Treat R and $\vec{r}_{ab}$ as variables, $\vec{e}_\alpha$ as a constant. In this case, forces are only added to the atom where the multipole moment is located.

2. Full derivatives: include all the variables R, $\vec{r}_{ab}$ and $\vec{e}_\alpha$ into the derivative. The resulting expression is dependent on the definition of the reference axis systems and adds forces to all atoms which define the reference axis system, no matter if they carry multipole moments or not.

The first level of accuracy is an approximation which is valid if either the reorientation of the multipole moment and therefore the change of $\vec{e}_\alpha$ is small, or if the multipolar force contribution to the total rotation of the molecule is small. In practice, the first level of accuracy is sufficient to conserve the total energy during an MD simulation if only parts of the molecules in a system carry atomic multipole moments or if the rotation of the molecules is small, which is the case for simulation at low temperature or for systems where the rotation is energetically unfavorable. Detailed tests can be found in Section 3.3

The second level of accuracy is required to conserve the total energy for systems with larger rotations in each step. The implementation is more involved, since the actual expressions for the derivatives depend on the reference axis system, therefore in principle each type of reference axis systems needs its own set of derivatives for all multipole interaction terms. For linear molecules, the expression is straightforward, but it becomes complicated for systems with more reference axes, especially if one axis is defined by orthogonality as it is the case for water. Since the orthogonal axis depends on the unit vectors corresponding to the other two directions, the derivatives in consequence also include the elements of the other two vectors.

Another problematic issue for the complete derivative are the additional forces on atoms on which no multipole moments are located. These forces do not have a direct physical meaning. They are the consequence of accepting the use of molecular reference axis systems. At this point, the concept of atomistic force fields, which consist in the separation of intra- and intermolecular forces, comes to its limits, since the atomic multipole moments require a combination of molecular and atomistic forces.

3.2 Code overview

A schematic representation of the code is shown in Figure 3.2 The different parts of the code are explained here in more detail.

A Input section; on initialization of the multipole module, template reference atoms are read for each molecule type. They are then copied to the specified range corresponding to the same type of molecule.

B The orientations of all molecules with multipole moments are calculated based on the definition of the corresponding reference axis system. On this basis, the orientation unit vectors $\vec{e}_\alpha$ are calculated for each direction and each molecule.

C For each atom pair on the nonbond list, it is checked whether at least one atom of the pair has multipole parameters. If this is the case, the distance R and $\vec{r}_{ab}$ are calculated.

D The scalar products of $\vec{e}_\alpha$ and $\vec{r}_{ab}$ are formed.

E According to the specification of the highest multipole rank on each atom, the interaction energies and derivatives are calculated. Since there are many redundancies in the terms due to symmetry, the actual energy and force expressions are organized as separate subroutines and the main part of the code just defines which interaction pair (e.g. μ_{zeta} on A, θ_{10} on B) is being calculated.

F Energies and forces are summed up. For the second level of accuracy, (see previous section) two sums have to be formed, one over the terms which only apply to the atoms with multipoles, the second over all atoms in the reference axis system.

For periodic boundary conditions, the additional nonbond energy between primary and image atoms requires the summation of energies and forces over an additional loop. The calculation is done in the same way as for the primary nonbond interactions, but with the image nonbond list instead of the primary nonbond list.

For several applications, (see Section 4) additional features are added to the main code, namely anharmonic bond potentials and changes of the multipole moments with bond lengths. Both of these features can be combined with the calculation of the molecular orientation, since they require the bond length which is needed in this part anyways.

Figure 3.2: Flow chart of CHARMM multipole module

Blue color shows the standard CHARMM code, red color the additional multipole module MTP
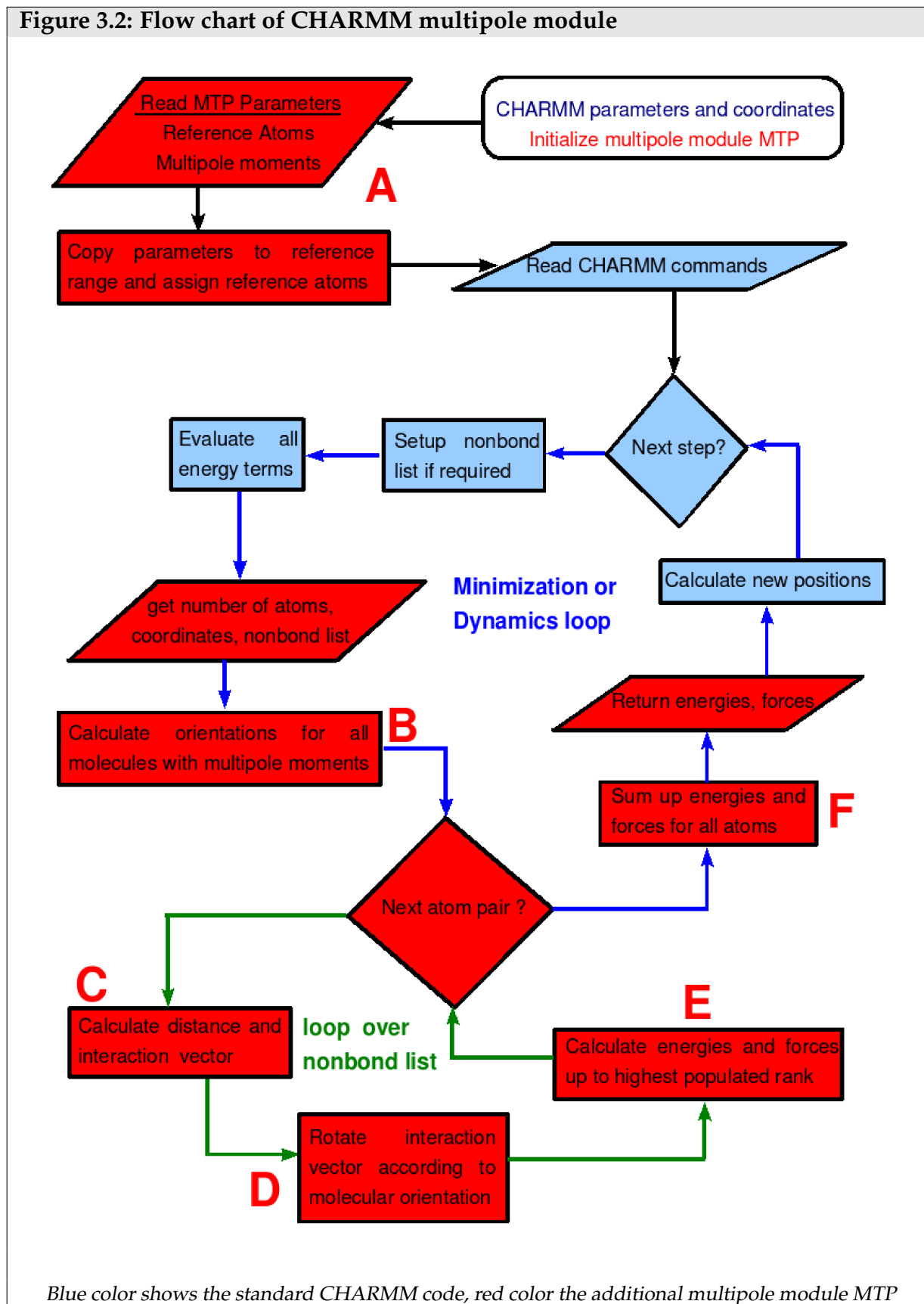
3.3 Validation

3.3.1 Interaction energies

The multipole interaction energies were compared between grid-based potentials in a model program to the program ORIENT.[70] For this purpose, a probe charge and a probe dipole moment were placed on a cubic grid around different probe molecules represented by multipole expansions. The energies were evaluated at each grid point and compared between the two programs. The energies were found to be identical for the two programs at all grid points within the numerical accuracy. The same setup of a probe charge on a cubic grid was used to compare the potential energy to the *ab initio* potential. Some of these tests are shown in Section 2. The comparison to ORIENT shows that the published interaction terms[22] are interpreted correctly while the comparison to *ab initio* provides information about the accuracy of a given multipole parametrization.

Based on these evaluations, the interaction energy terms were then implemented into the CHARMM program. The only change with respect to the test program was the change of units. All calculations of the multipole module are done in atomic units, input coordinates and output energies and forces are converted with respect to the units in CHARMM. (Å and kcal/mol)

The rotation of atomic multipole moments based on molecular reference axis systems was tested for several small molecular clusters. The interaction energy of the cluster was evaluated, then the entire cluster was rotated by randomly chosen rotation angles about all rotation axes. This procedure was repeated several times and the interaction energy of the cluster was evaluated and compared to the initial orientation at each point. The energies were found to be identical within the numerical accuracy of the CHARMM program at all rotation angles for all tested reference axis systems.

3.3.2 Derivatives

Two types of tests for the first derivatives are possible with the CHARMM program:

- Comparison of the analytical derivatives to numerical derivatives. The numerical derivatives are obtained by energy calculations for small displacements of the atoms.

- Evaluation of the fluctuations in the total energy during MD simulations in the NVE ensemble.

Numerical derivatives were calculated for both levels of accuracy of the derivatives explained in Section 3.1.4 using several small test systems. The accurate terms were found to produce differences between numerical and analytical derivatives in the same order of magnitude as other energy terms of the CHARMM force field. The use of approximate derivatives lead to differences of one or two orders of magnitude larger. The evaluations of fluctuations in the total energy will be explained in the following for MD simulations of several test systems.

Comparison of the different levels of accuracy for the forces

The fluctuations in total energy are compared between accurate and approximate derivatives for two test systems. As a first test system, two dioxygen molecules are chosen. The two molecules form a dimer at low temperatures and interact by LJ potentials and atomic multipole moments. The test runs include 100 minimization steps using the steepest descent algorithm, 20 ps of heating and equilibration to the test temperature and 100 ps of dynamics in the NVE ensemble with a timestep of 1 fs. The results are shown in Figure 3.3.2a.

The tests show that at 5 K where the dimer structure changes only slightly during the dynamics, approximate and accurate derivatives conserve the total energy well. The difference between the average energy and the energy at each step is small. Compared to the fluctuations in potential and kinetic energy, the fluctuations in the total energy are smaller by two to three orders of magnitude. At 10 K, the change of the dimer structure is larger and faster. As a consequence the difference in the fluctuations of the total energy become larger between the two levels of accuracy. For the

Figure 3.3.2a: Energy fluctuations for the O$_2$ dimer
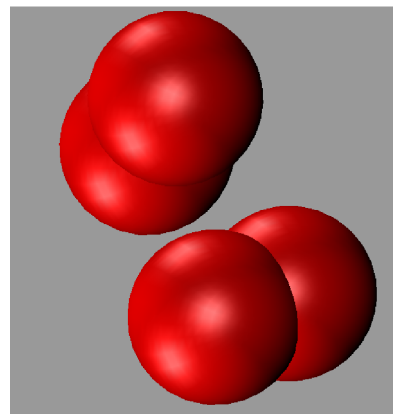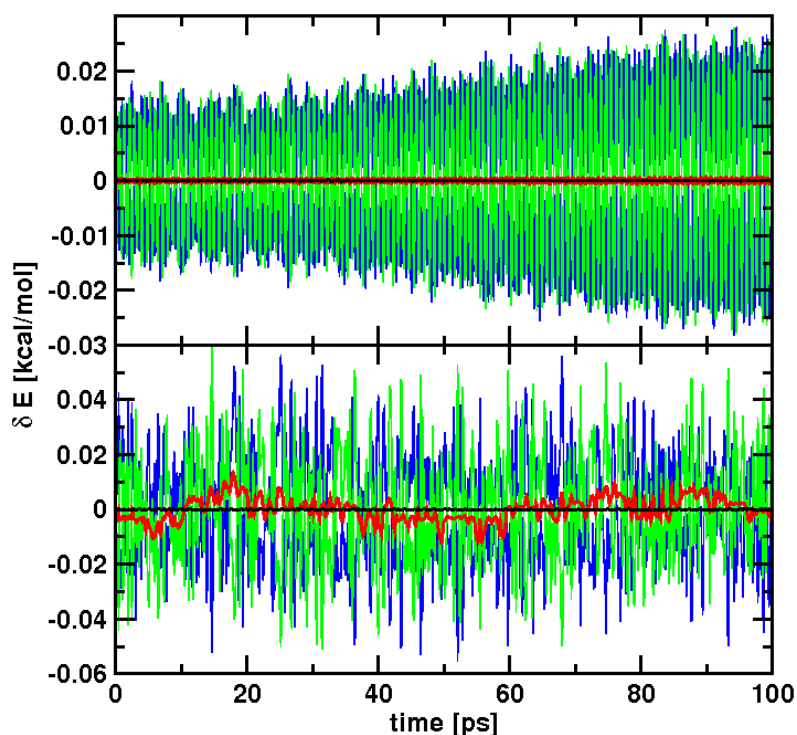
Oxygen Parameters

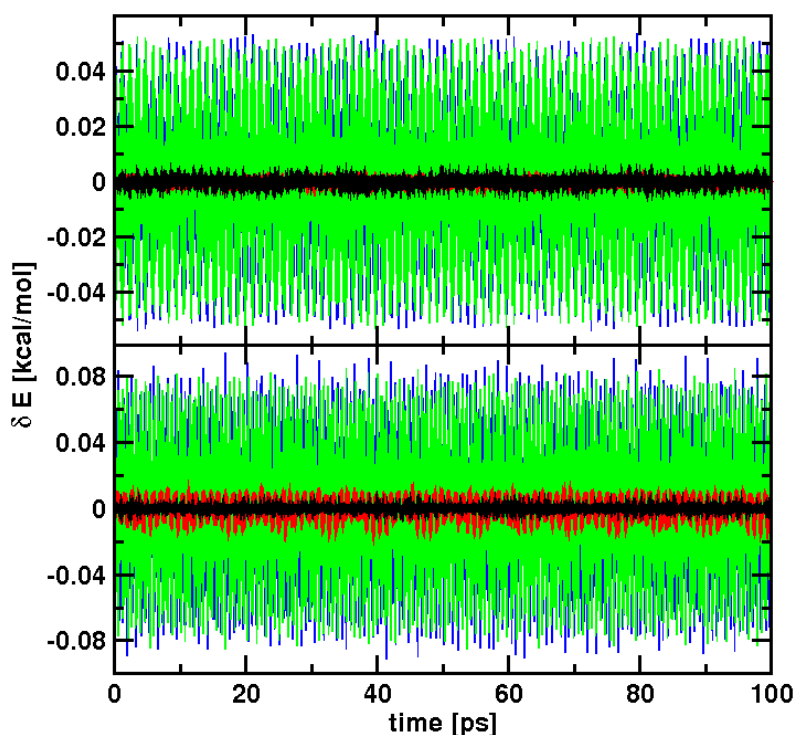|	
ϵ	-0.12 kcal/mol
$\sigma/2$	1.7 Å
μ	-0.288298 ea_0 *
θ	0.541545 ea_0^2

The difference of each energy term minus its average is plotted for trajectories of two dioxygen molecules. Red lines correspond to the first level of accuracy in the derivatives, black lines to the second level of accuracy, green lines show the kinetic energy and red lines the potential energy. Temperatures of the initial equilibration are 5 K (top) and 10 K (bottom).

approximate derivatives, the energy fluctuations are only about one order of magnitude smaller than the fluctuations in potential and kinetic energy. For the accurate derivatives, the difference in the energy fluctuations between the total energy and its contributing terms remains at two to three orders of magnitude. An additional analysis of the contribution of the VdW and electrostatic energy fluctuations to the fluctuations in the potential energy (not shown here) reveals that the fluctuations in both terms are approximately equal.

The same test is repeated for a molecular cluster with one water and two CO molecules.(Figure 3.3.2b) The setup is the same as before, except that a timestep of 0.4 fs is chosen to allow flexible O-H bonds on the water molecule with an anharmonic bond potential.[71] The parameters for this test can be found in Table 3.3.2.

---

*opposite signs on the two oxygens

Figure 3.3.2b: Energy fluctuations for the CO-H$_2$O cluster

The difference of each energy term minus its average is plotted for trajectories of one water and two CO molecules. Red lines correspond to the first level of accuracy in the derivatives, black lines to the second level of accuracy, green lines show the kinetic energy and red lines the potential energy. Temperatures of the initial equilibration are 5 K (top) and 10 K (bottom).

The findings are very similar to the first test system. At 5 K, the difference between the two levels of accuracy is not visible, but it becomes larger at 10 K, where the changes in the cluster structure are faster and larger.

Both test systems have relatively weak intermolecular interactions, therefore already a small increase of the temperature leads to faster rotation of the molecules. This faster rotation of the molecules is related to the larger fluctuations in the total energy for the approximate derivatives.

Table 3.3.2: H_2O and CO parameters used for testing						
Water Electrostatics				**Lennard-Jones**		
charge[e]	q00					
dipole [ea_0]	q10 q11s					
quadrupole [ea_0^2]	q20 q22c				ϵ [kcal/mol]	$\sigma/2$ [Å]
Oxygen		**Hydrogen**		O(H_2O)	-0.1521	1.7682
W1				H	-0.0460	0.2245
q	-0.705394	0.352697		O(CO)	-0.1200	1.7000
μ	-0.130332	0.091939	-0.081400*	C	-0.1100	2.1000
Θ	0.091999 -1.280201					
W2						
q	-0.276216	0.138108				
μ	-0.485045	0.031667	-0.003249*			
Θ	0.130550 -1.662565					
The W1 set of parameters is used for the molecular cluster with one water and two CO molecules. For CO electrostatics, see parameters for Model B used for CO in Section 4.1. The change of moments with bond lengths is not applied for this test.						

---

*opposite sign on the two hydrogens

Energy fluctuations for different temperatures and parametrizations

If the temperature of an MD simulation is increased, the fluctuations in the kinetic and the potential energy become larger since the trajectory samples regions of the potential energy surface which are farther away from the minimum. It is important to know whether the stability of the trajectory remains unaffected by the temperature change, i.e. if the fluctuations of the total energy remain proportional to the fluctuations in the contributing terms.

Another important question concerns the parametrization of multipole potentials: Does the distribution of the molecular multipole population among the different multipole ranks affects the stability of the trajectory and therefore the conservation of the total energy? As shown in Section 2.2, it does not affect the converged properties of the multipole potential.

Both questions are addressed here using the water pentamer as a test system (Figure 3.3.2c). The simulations are equilibrated at 10 and 200 K. The setup is the same as for the previous test, with a timestep of 0.4 fs. Two different sets of water parameters W1 and W2 are evaluated at each temperature. The parameter sets are shown in Table 3.3.2. Both sets are obtained from the same B3LYP aug-cc-pVTZ calculation for water, but with different radii for the Lebedev integration. For W1, the dipole moment of the water molecules is mostly described by the charges. For W2 in contrast, the atomic dipole moments contribute most to the molecular dipole moment.

The test results in Figure 3.3.2c show that the fluctuations in all energy terms are increased by about one order of magnitude as the temperature is increased from 10 to 200 K, which means that the conservation of the total energy remains proportional to the changes in the contributing terms. The differences between the multipole parametrizations W1 and W2 are small. The choice of parameters does not seem to affect the energy fluctuations.

**Figure 3.3.2c: Energy conservation tests for water pentamer**
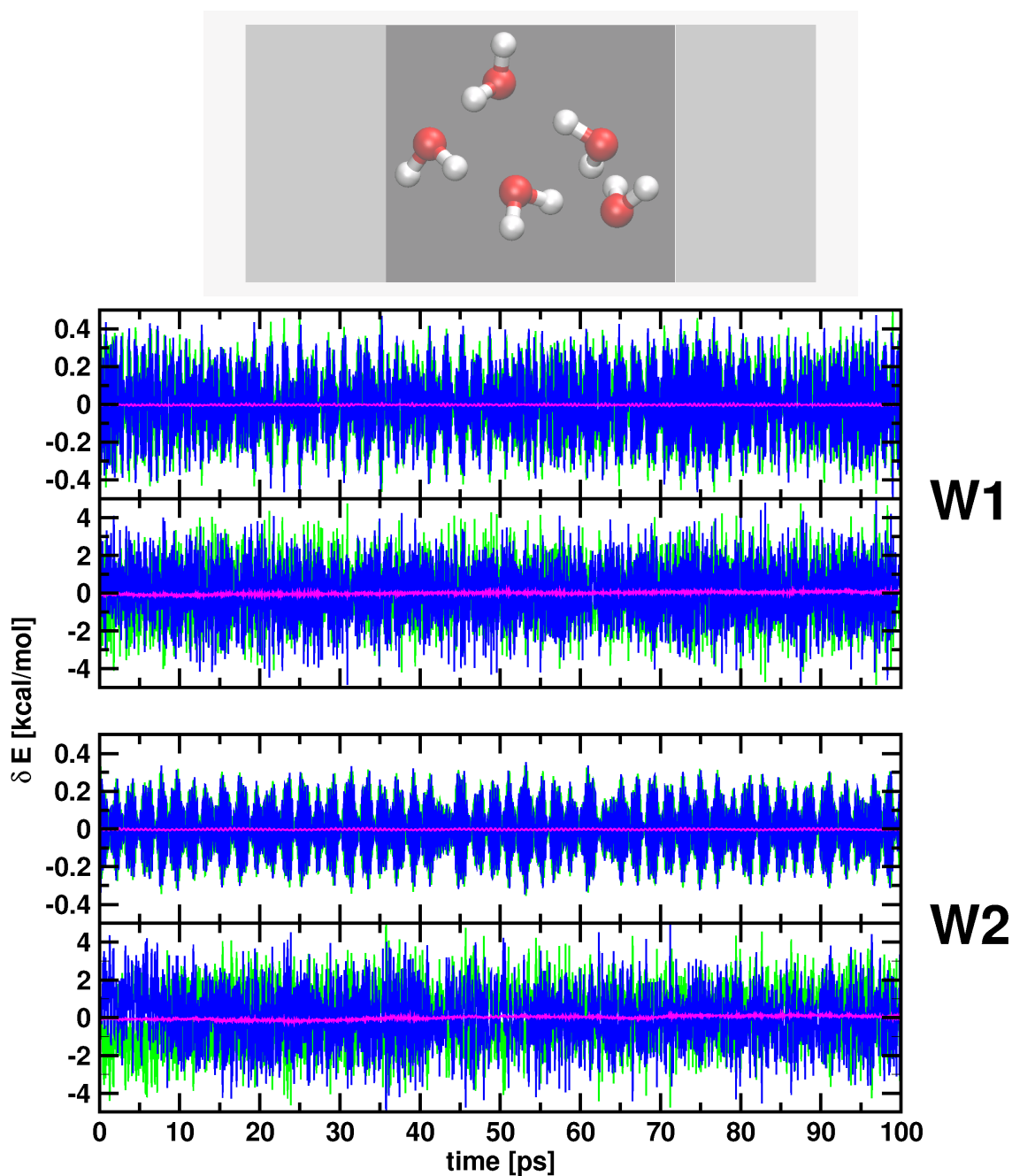
The difference of each energy term minus its average is plotted for trajectories of the water pentamer at different temperatures. Magenta lines correspond to the total energy, green lines show the kinetic energy and blue lines the potential energy. The two panels W1 and W2 correspond to the two different water parametrizations explained in Table 3.3.2. Each panel shows evaluations at 10 K (top) and 200 K (bottom).

Energy conservation using cutoffs and periodic boundaries

Additional problems in the conservation of the total energy arise if distance dependent cutoffs of the energies and forces are used. In general, cutoffs are required to reduce the computational cost of MD simulations. Furthermore, they are used for periodic systems, in order to avoid artifacts due to interactions of an atom with its periodic image. For these reasons it is important to test the effects of cutoffs.

The use of cutoffs means generally that the energy is not strictly conserved. Cutting the intermolecular interactions at a defined distance leads to a change in the potential energy without a corresponding compensation in the kinetic energy. However, these small changes in energy mostly cancel, since in total the atoms moving in and out of cutoff distances compensate for the energy changes. As a test system for this problem, CO clathrate hydrate is chosen. Further informations about this system can be found in Section 4.3. Parameters and setup used for these tests are the same as for the CO-water cluster shown in Table 3.3.2.

Figure 3.3.2d: CO clathrate hydrate
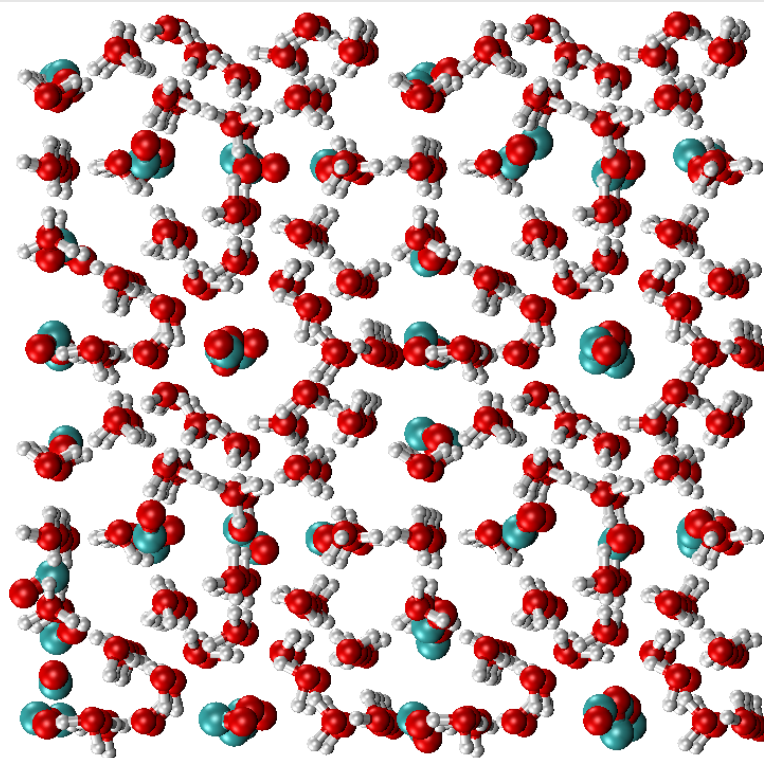
Optimized structure of clathrate hydrate simulation cell used for testing. Rendering:VMD[72]

Figure 3.3.2e: Energy conservation for CO clathrate hydrate
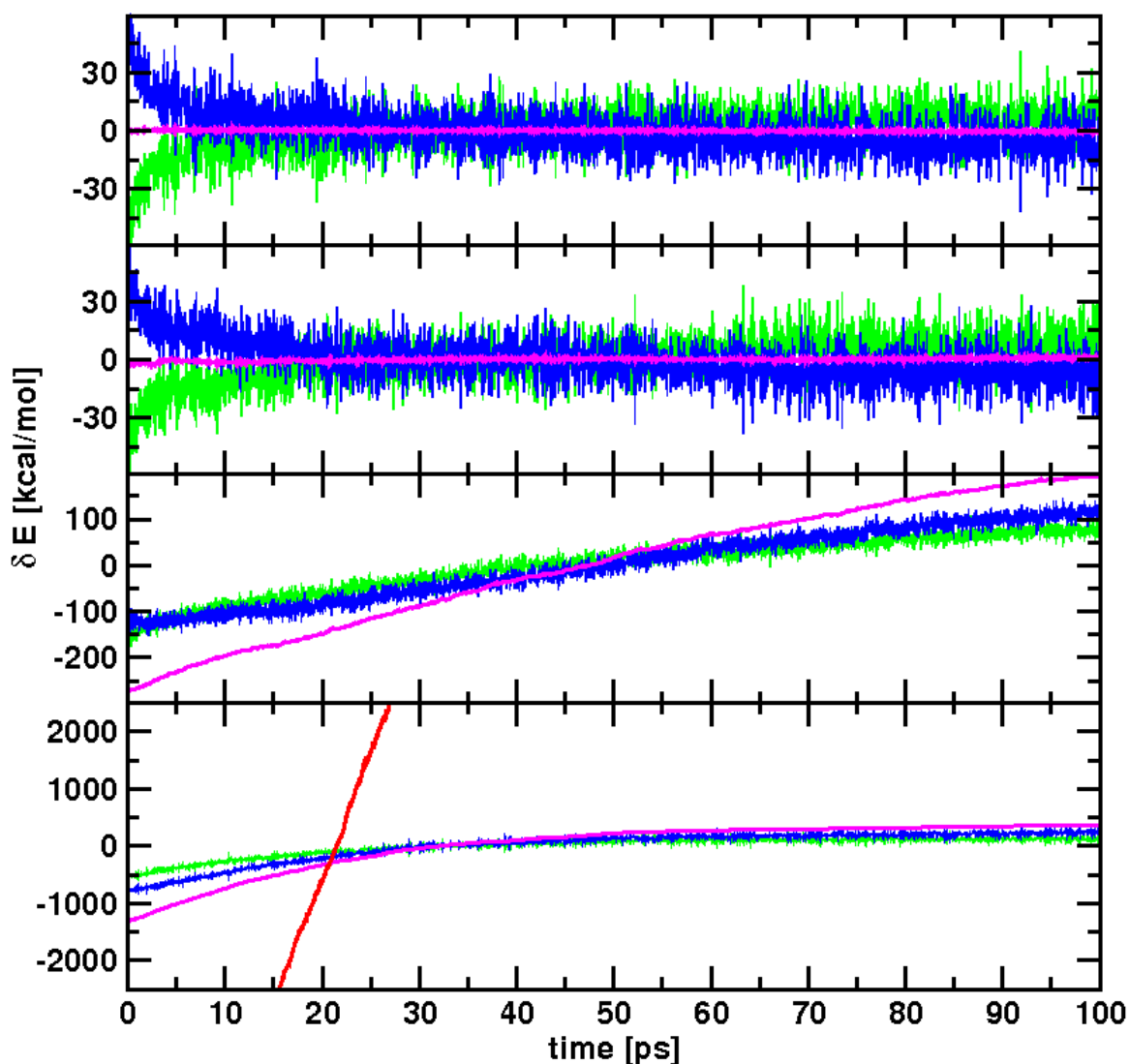
The energy conservation of the CO clathrate is evaluated with different cutoffs. From top to bottom: no cutoff, cutoff at 30 Å , cutoff at 20 Å , cutoff at 12 Å . The difference of each energy term minus its average is plotted. Magenta lines correspond to the total energy, green lines show the kinetic energy and blue lines the potential energy. The red line shows the drift in total energy for a trajectory with a cutoff of 12 Å and periodic boundaries.

For the clathrate unit cells without cutoff, the total energy is well conserved. Since no periodic boundary conditions are used, the system transforms from the symmetric clathrate unit cell shown in Figure 3.3.2d into a disordered sphere of water and CO. This structural rearrangement corresponds to the large fluctuations of the kinetic and potential energy at the beginning of the simulation.

If a cutoff of 30 Å is used, the total energy shows larger fluctuations, but it is still roughly conserved. Since the initial unit cell length is 24.06 Å , only very few interactions are outside the cutoff distance. With a cutoff of 20 Å, a large drift of about 100 kcal/mol is observed. For cutoffs of 12 Å the drifts becomes about five times larger. If the same cutoff is used with periodic boundary conditions, the drift increases again.

The evaluation was done using a switching function analogous to the standard CHARMM switching function.[11] This function scales the electrostatic interaction energy and force smoothly to zero with distance and is therefore supposed to improve the conservation of the total energy. However, for this evaluation the switching function did not reduce the total energy drift. The total energy curve was found to be smoother, but the drift remained roughly the same as for equivalent tests without switching function. Other functional forms for the energy scaling with distance have also been tested, but no improvement in energy conservation was observed.

From these results, it is obvious that the use of cutoffs is much more problematic for atomic multipole moments than for point charges. The errors due to cutting energies and forces with distance do not cancel. These findings are consistent with earlier implementations of atomic multipole moments into the DLMulti force field for rigid molecules and fragments.[69] The tests for the DLMulti force field show furthermore that even if Ewald summation is used, a drift in the total energy remains. Following the interpretation given for that force field, these difficulties are due to the fact that in contrast to point charges, higher multipole moments change their orientation if they are outside the cutoff distance. Therefore, the error in potential energy does not cancel. Another problem may be that the entire system can have a net dipole moment, which also leads to a drift in energy.

3.4 Discussion: Important points to consider for applications and further development

For a more general implementation of atomic multipole moments, the dependence on reference axis systems is the major difficulty. This is not only due to the fact that the reference axis systems have to be provided for all molecular geometries, but also because of the dependence of the derivatives on the reference axis system. The difficulties are reduced considerably if approximate derivatives are used, which do not include specific term for each reference axis system.

The use of approximate derivatives does not cause a direct problem in practice, since for most problems, MD simulations can be carried out in the NVT or NPT ensemble, where drifts in the total energy get absorbed. Nevertheless, the effects of errors in the derivatives should not be neglected, since it is important that the sampling of the free energy surface is not biased by systematic errors in the forces. This is not the case if the effect of the errors changes the results of the simulation in a random fashion. On the other hand, errors which e.g. systematically change the motion of one type of atoms should not occur.

For these reasons, control quantities like the conservation of the total energy evaluated in Section 3.3 should be used as a guideline to decide what system types the approximation may be used for. As far as the use of approximate derivatives is concerned, random errors are introduced if the rotation of the molecules is an oscillation about an equilibrium orientation, or if the rotation is slow in general. This is the case for solids and therefore for all systems in Section 4, namely CO in Myoglobin and mixed ices. In contrast, the approximate derivatives should not be used for decomposition studies of the same ice systems or for liquids. Therefore, the question which level of accuracy should be used for a more general implementation depends on the problem types to which the resulting atomic multipole force field is supposed to be applied to.

The use of approximate derivatives and discussions about derivative accuracies is not an issue present for standard force fields. In the field of *ab initio* MD in contrast, this issue is commonly discussed,[73,74] and criteria have been established to judge how useful an approximation or level of accuracy may be. Such criteria are the fluc-

tuations or drifts in total energy and the distribution of the energy to the different degrees of freedom according to the equipartition theorem. For other algorithms like path integral MD, it is accepted to systematically use separate thermostats for different degrees of freedom[75] or likewise to use Langevin dynamics. In this case, this is dictated by the non-ergodicity of path intergral molecular dynamics.[76]

As the empirical force fields become more sophisticated, the use of approximate derivatives and the availability of analytical derivatives also becomes an issue in this area. This is the case for the atomic multipole moments shown here or also for polarizable force fields, for which a new MD algorithm to solve the problem has already been proposed.[77] It seems to the author that not all developers pay sufficient attention to this problem, since e.g. for the implementation of distributed multipoles and polarizabilities in the TINKER force field,[78] no discussion of control quantities can be found in the literature. For the DLMulti force field in contrast, which includes atomic multipole moments on rigid molecules and fragments, a detailed discussion is provided.[69]

Given all these considerations, it is difficult to make a final conclusion about the optimal approach for the implementation of the derivatives in an atomic multipole force field. However, it is clear that this is an important issue which should be handled with care.

4 Applications

"For seeing life is but a motion of Limbs, the begining whereof is in some principal part within; why may we not say, that all Automata (Engines that move themselves...) have an artificial life?"

Thomas Hobbes, *Leviathan*, 1651

4.1 The Role of Higher CO-Multipole Moments in Understanding the Dynamics of Photodissociated Carbonmonoxide in Myoglobin

N. Plattner and M. Meuwly

The Role of Higher CO-Multipole Moments in Understanding the Dynamics of Photodissociated Carbonmonoxide in Myoglobin

Nuria Plattner and Markus Meuwly
Department of Chemistry, University of Basel, Basel, Switzerland

ABSTRACT   The influence of electrostatic multipole moments up to hexadecapole on the dynamics of photodissociated carbon monoxide (CO) in myoglobin is investigated. The CO electrostatic potential is expressed as an expansion into atomic multipole moments of increasing order up to octopole which are obtained from a distributed multipole analysis. Three models with increasingly accurate molecular multipoles (accurate quadrupole, octopole, and hexadecapole moments, respectively) are developed and used in molecular dynamics simulations. All models with a fluctuating quadrupole moment correctly describe the location of the B-state whereas the sign of the octopole moment differentiates between the Fe···CO and Fe···OC orientation. For the infrared spectrum of photodissociated CO, considerable differences between the three electrostatic models are found. The most detailed electrostatic model correctly reproduces the splitting, shift, and width of the CO spectrum in the B-state. From an analysis of the trajectories, the spectroscopic B_1 and B_2 states are assigned to the Fe···CO and Fe···OC substates, respectively.

INTRODUCTION

The energetics, dynamics, and spectroscopic characterization of ligands in protein cavities provide fundamental information about intermolecular interactions and their role in controlling and influencing the relation between structure and function in proteins. Experimental and computational studies have revealed active sites within proteins which facilitate specificity for ligand binding and catalysis (1–4). Despite the remarkable progress in characterizing structural and dynamical aspects of ligands in protein active sites, the detailed (i.e., atomistic) and quantitative understanding of spectroscopic and kinetic data remains a central topic of biophysical research (5–7). One of the best studied examples, both experimentally and computationally, is the active site of myoglobin (Mb). Apart from its biological importance in storing and transporting dioxygen in the body, its involvement in vasodilation and in scavenging nitric oxide, Mb together with small diatomic ligands is a paradigm system for protein dynamics. Among these, photodissociated carbon monoxide (CO) from Mb is particularly well characterized. For this system, both structural and spectroscopic investigations for free CO in the heme pocket have been carried out (8–11).

Photodissociated CO in myoglobin is a system which has been investigated intensively by experiments (8–12) (low-temperature and time-resolved x-ray crystallography) and atomistic simulations (classical molecular dynamics (MD) and mixed quantum mechanical/molecular mechanics (QM/MM) MD simulations) (13–20). One general observation from all these efforts is that the details of the apparently simple processes of ligand dissociation and rebinding are actually quite complex. For example, to correctly locate the docking site (B-state) where the CO molecule settles after breaking the Fe-CO bond in agreement with the experimental findings (8,9) it was necessary to carry out atomistic simulations with charge models which capture the quadrupole moment of CO (13,16). Another feature which has repeatedly attracted the attention of experimentalists (21) and theoreticians (14,15,22,23) alike, is the split infrared spectrum of photodissociated CO in the docking site which was originally observed by Anfinrud and co-workers (9,24). The IR spectrum of Mb···CO shows two well-resolved bands (at 2130.5 cm^{-1} and 2119.0 cm^{-1}), split by 11.5 cm^{-1}. Compared to free CO (2143.3 cm^{-1}), the upper band is shifted by ~10 cm^{-1} to the red whereas the lower band by ~20 cm^{-1} (24). The splitting has been associated with two possible orientations (states B_1 and B_2) of the CO molecule in the docking site, either with the carbon closer to the iron center (Fe–CO) or with the oxygen closer to the iron center (Fe–OC) (24). Some effort has gone into identifying the spectroscopic B_1 and B_2 states with either Fe–CO and Fe–OC. Originally, the two peaks have been assigned to Fe–CO (B_2) and Fe–OC (B_1) from an analysis of the rate of appearance of the two bands in polarization-sensitive experiments (25). Further studies were based on classical MD simulations (14,15), a combination of experiment and density functional theory (DFT) calculations on model compounds (21), MD simulations together with DFT and quantum bound state calculations (22) and simulations using the perturbed matrix method (23). The first two studies did not explicitly correlate one state with a particular IR band (14,15), whereas the next two studies (21,22) find that B_1 correlates with Fe–CO and B_2 with Fe–OC. This contrasts with an opposite assignment from recent PMM calculations (23).

It has been recognized for some time that the CO infrared spectrum in the B-state depends on the accurate representation

of the electrostatic interactions between CO and its environment (13–15). But even with a refined fluctuating quadrupolar model for CO which correctly describes the location of the B-state, the splitting of the IR spectrum, and the energetics for CO rotation (16), it was not possible to directly correlate one spectral feature with a particular orientation of the CO molecule relative to the heme-iron atom from MD simulations (26). In this work, we address the question whether a systematically improved description of the electrostatics around the CO molecule together with atomistic simulations can help us to quantitatively understand various aspects of CO dynamics in the B-state. In a conventional force field (27,28), the electrostatic potential is represented by point charges, which are often obtained from a population analysis of the Hartree-Fock wave function from ab initio electronic structure calculations (29). Usually, atom-centered charges are obtained by fitting atomic charges which optimally reproduce the electrostatic potential (30,31) or by adjusting the overall potential of reference systems to experimental data. Using atomic point charges to represent the electrostatics around a molecule is largely a matter of convenience, but is also supported historically. It has been demonstrated in the past, in the theory of population analysis algorithms, that point charges can be understood as a first term in a multipole expansion (32,33). Each term represents the corresponding multipole population of the ab initio electron-density distribution. The expansion converges toward the electrostatic potentials of the electron-density distribution on which the analysis is based.

For carbon monoxide (CO), two point charges are not sufficient to describe the electrostatic potential, because only the dipole moment, which is very small, can be reproduced accurately. In the past, different models have been developed to represent the conformationally dependent or independent dipole and quadrupole moment of CO accurately with respect to gas phase data (13,16,34). Furthermore, the influence of CO multipole moments in MD simulations of vapor-liquid equilibria of pure CO has been investigated using point dipole and point quadrupole moments obtained from fitting to thermodynamic data (35). In an earlier study (16), the splitting of the CO absorption band could be reproduced using a three-point fluctuating charge model for the description of the CO electrostatics. This study aims at systematically investigating the influence of higher electrostatic moments than the molecular quadrupole on the ligand-protein interaction in MbCO. In particular it is of interest to determine whether using a convergent mathematical expression for the electrostatic field around a ligand (CO) is directly related to a stepwise improvement of the observables (IR spectrum, probability density functions) calculated from them.

This work is structured as follows. First, the computational methods used are described. Next, results from simulations using increasingly detailed electrostatic models are presented. Finally, the results are discussed in view of experimental and previous computational work.

COMPUTATIONAL METHODS

Molecular dynamics simulations

The computational setup for the molecular dynamics (MD) simulations closely followed that of previous studies of photodissociated MbCO (16). Briefly, hydrogen atoms were added to the x-ray structure (Protein Data Bank reference 1MBC) (36), and the heme pocket was solvated by a 16 Å sphere of equilibrated water molecules. The reaction region of 12 Å was centered on the heme-Fe atom (see Fig. 1). The dynamics inside this region was described using Newtonian dynamics, the dynamics in the buffer region (12–16 Å) by Langevin dynamics. Friction coefficients of 62 ps^{-1} and 250 ps^{-1} were applied to the oxygen atoms of water and the remaining nonhydrogen atoms. Nonbonded interactions were treated with a cutoff at 10 Å. The sensitivity of the results on the cutoff for the nonbonded interactions has been tested on a subset of three trajectories, for which a nonbonded cutoff of 10 Å and 12 Å was used, respectively. Comparison of the averages over the three trajectories showed only minor differences. This is in agreement with findings from Meller and Elber who used particle-mesh Ewald techniques for the electrostatics (15). Bound MbCO was equilibrated for 70–140 ps before photodissociation which was modeled by the sudden approximation (37): The Fe-C bond was deleted and the bound parameters replaced by a repulsive term during 0.1 ps of dynamics. After this, the different electrostatic models were used for CO together with the CHARMM CO Van der Waals parameters and the RRKR bond potential for the CO stretching coordinate (38). For comparison of the different electrostatic models (A, B, C, and C′; see below), eight trajectories of 1 ns each were run and analyzed. For model C, which was found to be the most accurate, eight additional trajectories were evaluated. Simulations were carried out for both possible protonation states, $His_\delta 64$ and $His_\epsilon 64$. The latter is the more likely (39–41), although the former was used in several previous simulations (13,16) which also provided benchmarks for this work.

The infrared spectrum for each trajectory is calculated from the time dependence of the dipole moment of the CO molecule. The time correlation function is accumulated over 2^n time origins, where n is an integer such that 2^n corresponds to between 1/3 and 1/2 of the trajectory, with the time origins separated by 1 fs. This function can then be transformed to yield $C(\omega)$ using a
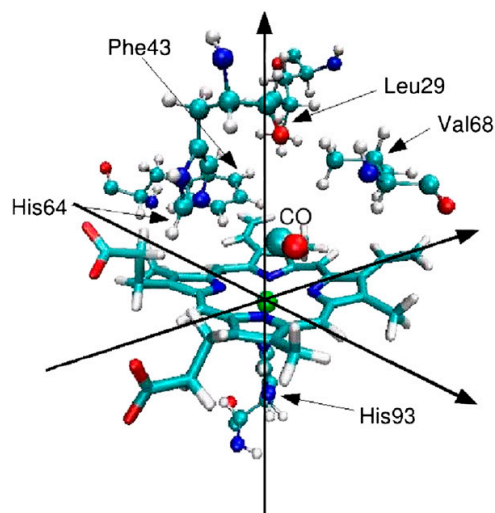

FIGURE 1 Myoglobin binding site together with the key residues and the CO molecule. The coordinate system is defined with respect to the iron center (*green sphere*) and the heme plane.

fast-Fourier transform with a Blackman filter to minimize noise (42). The final infrared adsorption spectrum is then calculated by evaluating

$$A(\omega) = \omega\{1 - \exp[-\omega/(kT)]\}C(\omega), \qquad (1)$$

where k is the Boltzmann constant and T is the temperature in Kelvin.

As a reference, the IR spectrum has also been calculated for free CO at 300 K using the RRKR potential. The spectrum shows one line at 2183 cm^{-1}. This corresponds to an offset of $\sim$40 cm^{-1} compared to the experimental value which is consistent with previous investigations (16). The offset has two different origins. First, it has been noted from simulations of CO in argon (43) that the integration of the equations of motions using the Velocity-Verlet algorithm with a time step of 1 fs shifts the frequency by 16 cm^{-1}. The remaining 24 cm^{-1} can be attributed to the classical treatment of the an-harmonic oscillator, which can be shown to have an energy-dependent fundamental frequency ω_E,

$$\omega_E = \alpha \sqrt{\frac{2(D_e - E_{tot})}{\mu}}, \qquad (2)$$

where E_{tot} is the total energy of the oscillator. All IR spectra discussed below are compared with the calculated spectrum for free CO (absorption at 2183 cm^{-1}).

Electrostatic models for CO

Multipole models

The distributed multipole algorithm (44) used here contains two different allocation methods for the multipole moments: the more localized basis functions are treated by a nearest-site algorithm. The less localized and diffuse basis functions are partitioned through space using the Lebedev integration method (45) over a spherical grid with 590 points. The atomic radii for this integration method are variable parameters. In this study, they have been adjusted such as to reproduce the molecular moments as accurately as possible for each model at the equilibrium C-O distance of 1.128 Å. Furthermore, the rank of multipoles is variable for each atom. Each molecular moment is composed of all atomic moments of lower and equal rank and was evaluated by using ORIENT (46).

Distributed multipoles are calculated using the program GDMA (44) based on Gaussian (47) formatted checkpoint files. Distributed multipoles for all models were evaluated for C-O distances $r \in [0.8..1.5]$ Å separated by 0.1 Å using the B3LYP functional and the aug-cc-pVQZ basis set (48,49). Different combinations of distributed multipoles on the two atoms have been evaluated, finally leading to three models (see Table 1). It was found that

TABLE 1 Molecular moments of CO calculated from the distributed multipoles on the C and O atom for models A–C

	Molecular moments				Highest atomic moment	
	$\mu[ea_0]$	$\Theta[ea_0^2]$	$\Omega[ea_0^3]$	$\Phi[ea_0^4]$	Carbon	Oxygen
Model A	0.04	−1.48	−2.68	−3.33	Dipole	Dipole
Model B	0.04	−1.51	3.42	−0.69	Dipole	Quadrupole
Model C	0.04	−1.51	2.86	−9.03	Quadrupole	Octopole
CCSD(T) (50, 51)	0.051	−1.47	3.46	−9.07		
Experiment*	0.048[†]	−1.58[‡]	4.15	−8.15		

The multipole moments (in atomic units) are evaluated for the equilibrium structure of CO.

*The values somewhat vary between experiments (for detailed information, see (50,51)).

[†]The experimental values have converged to this number.

[‡]This value varies between −1.4 and −1.68 ea_0^2 for different experiments.

distributed multipole analysis (DMA) faithfully reproduces the multipole moments of CO from high-level ab initio calculations (50,51) and experiment (see Fig. 2). Comparisons with other DFT methods and smaller basis sets showed that the molecular moments converge toward the values from B3LYP/aug-cc-pVQZ. The electrostatic potentials for the three models are represented as contour plots in Fig. 3. They show that the most important changes with increasing multipole order occur around the oxygen atom. In particular, whereas for the simplest model A only attractive interactions appear around the oxygen atom along the molecular axis, the more detailed model C has both attraction and repulsion. The changes of all atomic moments $q(r)$ as a function of the C-O bond length are to a good approximation linear. Therefore, they were described by linear functions $q(r) = \alpha + \beta(r - r_e)$ (see Fig. 4). Each atomic moment is defined by two parameters: its value α at the equilibrium bond length $r_e = 1.128$ Å and the slope β for the change with bond length. The coefficients of this linear expansion are given in Table 2.

Implementation of DMA

For the calculation of the atomic multipole moments, three additional multipole interaction terms in spherical tensor notation (52) have been implemented into CHARMM (27). A linear molecule (such as CO) can be oriented in a reference axis system for population analysis such that only the tensor components in z-direction (or in one arbitrary direction) are permanently populated. In the following, $q(00)$, $q(1z)$, $q(20)$, and $q(30)$ represent the charge and the z-components of the dipole, quadrupole, and octopole moment of atom sites a and b; R is the distance between the two atoms and
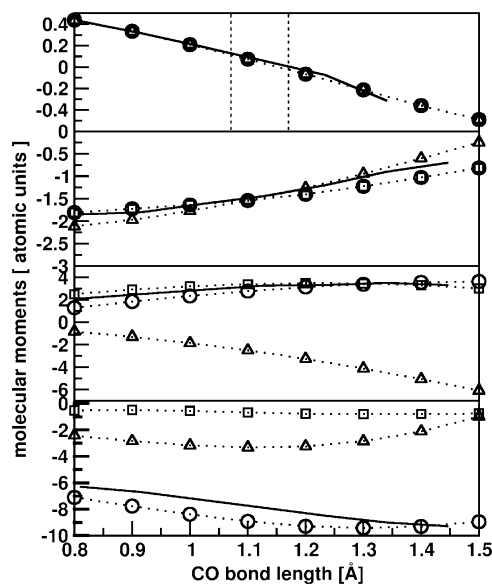

FIGURE 2 Molecular moments for the three electrostatic models for seven C-O bond lengths between 0.8 and 1.4 Å. (From *top* to *bottom*) Dipoles, quadrupoles, octopoles, and hexadecapoles. Model A is represented by triangles, model B by squares, and model C by circles. The solid lines represent the CCSD(T) values (50,51). Dipole and quadrupole moment are in good agreement with the ab initio data for all models. The octopole moment is accurately represented by models B and C; the hexadecapole moment from model C is the only one to qualitatively agree with the CCSD(T) calculations. The vertical lines represent the classical turning points for CO for $v_{CO} = 0$ with the RRKR potential.

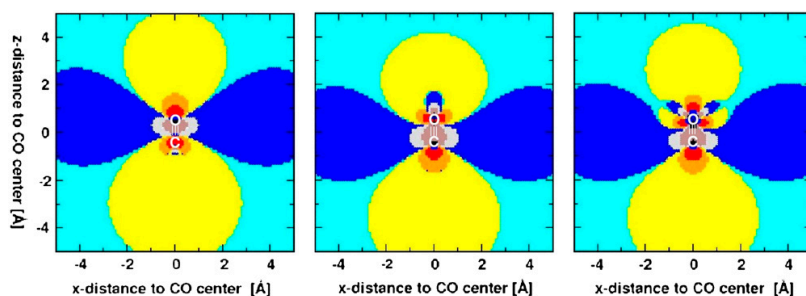FIGURE 3 Electrostatic potentials for each of the three CO models. Model A (*left*), model B (*middle*), and model C (*right*). The colors mark the energy in kcal/mol of a probe charge of $1e$ interacting with the CO molecule, with the oxygen at (0,0,0.564) Å and the carbon at (0,0,–0.564) Å. Contours (in kcal/mol) are drawn for $E > 3$ (*brown*); $1.25 < E < 3$ (*gray*), $0.5 < E < 1.25$ (*blue*), $0 < E < 0.5$ (*cyan*), $0 < E < -0.5$ (*yellow*), $-0.5 < E > -1.25$ (*orange*), and $-1.25 < E < -3$ (*red*).

r_z^a is the direction cosine of a unit vector pointing from a to b. From the distributed multipoles on each site, the overall, molecule-centered multipoles are calculated with respect to the geometric center of the molecule using standard formulae (46). The geometric center is chosen instead of the center of mass to ensure that the electrostatic potential only depends on the electronic structure of a molecule; e.g., ^{14}CO and ^{15}CO should have the same molecular moments.

For the charge-dipole, charge-quadrupole, and charge-octopole interactions the contributions to the electrostatic interaction between site a on the CO molecule (described by DMA) and site b (which is an atom of the surrounding protein described by a point charge) is given by the following equations:

$$U_{(1z00)} = \frac{q(1z)_a q(00)_b r_z^a}{R^2},$$ (3)

$$U_{(2000)} = \frac{q(20)_a q(00)_b (3(r_z^a)^2 - 1)}{2R^3},$$ (4)

$$U_{(3000)} = \frac{q(30)_a q(00)_b (5(r_z^a)^3 - 3r_z^a)}{2R^4}.$$ (5)

For the simulations three different models are considered. Model A includes atomic multipoles up to the dipole on each atom (C and O, i.e., (dipole/dipole) on (C/O)), which allows us to accurately describe the quadrupole
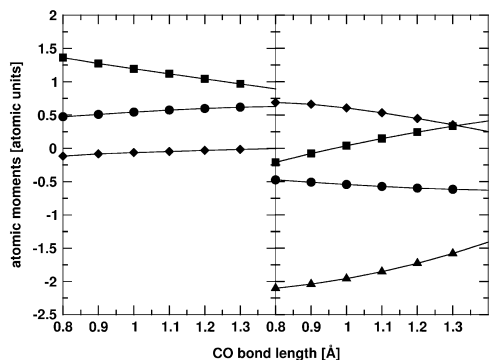

FIGURE 4 Atomic multipole moments for C-O bond lengths between 0.8 and 1.4 Å for model C. The left and right panels display the moments on the carbon and oxygen atom, respectively. Charges are marked by circles, dipoles by squares, quadrupoles by diamonds, and octopoles by triangles. The variation of the atomic moments with bond length is, to a good approximation, linear. This is also the case for the other two electrostatic models (not shown).

moment of the CO molecule. Models B and C include (dipole/quadrupole) and (quadrupole/octopole) moments which give accurate values for the molecular octopole and hexadecapole, respectively. This data is summarized in Table 1. To obtain a more complete picture of the relationship between the convergent multipole expansion models and the trends in the corresponding observables, an additional model (C′) was considered which includes atomic moments up to octopole on both atoms. The molecular moments differ from model C only in the hexadecapole moment, which is $\Phi_{CO} = -6.855\ ea_0^4$ compared to $\Phi_{CO} = -9.025\ ea_0^4$ from model C.

RESULTS

As mentioned in Computational Methods, simulations are carried out for both protonation states of His-64 (His$_\delta$64 and His$_\varepsilon$64). The first part of this section presents results for His$_\delta$64 which can be directly compared to previous simulations which were based on the fluctuating charge model (13,16). In the second part, where comparison with experiment is made, results for His$_\delta$64 and His$_\varepsilon$64 are discussed.

Characterization of electrostatic models A–C

First, the influence of details of the electrostatic models on the localization and IR spectra of CO after photodissociation are presented. For this, eight independent 1-ns trajectories for each electrostatic model (A–C) are compared. Model C′ is not discussed in detail here because no significant differences to model C were found.

Orientational preference and location of the B-state

Initially, the orientation of CO in the docking site (B-state) was analyzed. The three electrostatic models differ mainly in their prediction of the CO-orientation with respect to the Fe atom which can be related to the difference $\delta = |r_{FeC_{CO}} - r_{FeO_{CO}}|$ between the iron-carbon and the iron-oxygen distance. Considering atom-atom distances is equivalent to transforming to spherical polar coordinates (θ, ϕ) but describes the quantity relevant in this work (''which CO-atom is closer to the Fe'') more directly.

The results can be divided into three categories: orientations with carbon closer to iron (Fe···CO substate), orientations with oxygen closer to iron (Fe···OC substate), and

TABLE 2 Distributed multipole parameters for all electrostatic models

		Model A	Model B	Model C	Model C'
Charge[e]*	C	−0.563935[0.512792]	0.270293[0.227539]	0.563418[0.124576]	0.346638[0.073925]
Dipole[ea_0]	C	−0.226197[−0.581893]	0.818143[−0.431636]	1.096170[−0.395194]	0.880978[−0.151487]
	O	−0.955840[0.390026]	−0.180268[0.518313]	0.134258[0.5133167]	−0.104431[0.177545]
Quadrupole[ea_0^2]	C			−0.050725[0.087115]	−0.036699[−0.028279]
	O		0.562133[−0.324187]	0.467227[−0.453366]	0.624407[−0.087891]
Octopole[ea_0^3]	C				0.199193[0.197417]
	O			−1.774760[0.664064]	−1.239752[−0.048944]
Radius[Å]	C	0.500	0.660	0.700	0.800
	O	0.770	0.600	0.525	0.700

The values (α, [β]) refer to the equilibrium CO position and the linear changes (slope) of the moments as a function of the CO distance r (see main text). The units for the slopes are e/a_0, ea_0/a_0, ea_0^2/a_0, and ea_0^3/a_0, respectively. The radii are the relative atomic radii used for Lebedev integration in GDMA.
*Charges and changes of charge with bond length have opposite signs on oxygen.

orientations for which both distances are similar, which includes all distance differences $\delta \leq 0.3$ Å. The results (Fig. 5) show that models B and C clearly prefer the Fe···CO orientation, whereas model A prefers Fe···OC. This result does not only hold for the average orientation over all trajectories, but also for each individual trajectory. Comparison of the multipole moments in Table 1 shows that the major difference between model A and models B and C is the sign of the octopole moment Ω, which appears to determine the preferential orientation of unbound CO with respect to the Fe atom in the B-state. These findings are also in agreement with results from previous studies with a three-point fluctuating charge model with $\Omega = 0$, which makes it an intermediate case between models A and B/C. In MD simulations with the three-point fluctuating point charge model a less clear preferential orientation of CO in the B-state was found (16). The sensitivity of the results with respect to the criterion for distinguishing the substates was checked by using $\delta \leq 0.2$ Å, and no significant differences were found.

In addition to the relative orientation of the CO with respect to the heme-Fe atom it is also of interest to consider the CO probability distribution in the B-state. These distributions were analyzed in two ways. First, the center of mass of CO was projected onto the heme plane. These distributions (not shown) are similar to those from previous studies (16), with slight variations for the individual trajectories. All models find the docking site in a region between 3.5 to 4.5 Å away from the Fe atom and lined by residues His-64, Phe-43, and Val-68, which is consistent with experimental data from which it is known that shortly after photodissociation the CO molecule migrates to a metastable position (docking site or B-state) above the heme plane. Furthermore, the location of the CO molecule within the active site was characterized by considering the distance R between the geometric center of the CO and the Fe atom (see Fig. 5). All three models agree qualitatively for the distributions $p(R)$ with the exception that model C shows a higher proportion of large Fe–CO separations. Closer inspection of the trajectories reveals that these states correspond to CO positions in the Xenon-4 pocket. Models A–C and the previously developed fluctuating point charge model have a fluctuating quadrupole and correctly find the docking site whereas earlier work with a fixed three-point charge model showed some deficiencies in locating the B-state (16,53). In the x-ray structures the distance between the Fe atom and the geometric center of CO is 3.96 Å (PDB code 1DWS) (8) and 3.31 Å (PDB code 1AJH) (9), respectively.

In conclusion, it is found that a fluctuating quadrupole moment correctly describes the location of the B-state and
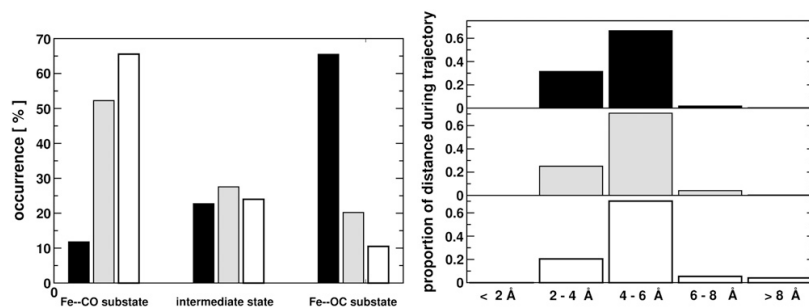

FIGURE 5 (*Left panel*) Orientations of the CO molecule for the three electrostatic models: model A (*solid*), model B (*shaded*), and model C (*open*). The Fe-C and Fe-O distances are compared at each step and divided into three categories: carbon closer to iron (Fe···CO substate), oxygen closer to iron Fe···OC substate), and distance differences $\delta <$ 0.3 Å (intermediate states). (*Right panel*) Distances of CO center of mass to the iron atom, calculated as the average of the Fe-C and Fe-O distance. The numbers indicate the distances at the center of each bin. The distances at ~8 Å correspond to positions in the Xenon-4 pocket.

the sign of the octopole moment differentiates between the Fe⋯CO and Fe⋯OC orientation.

Infrared spectra of dissociated CO

Structural data of unligated CO in Mb is quite rare (8,54,55), whereas the spectroscopic characterization provides more detailed information about the electrostatic environment of the ligand (11,24,25,56). Furthermore, because models B and C give essentially identical geometrical properties with respect to ligand orientation and the location of the B-state, it is of interest to investigate whether IR spectra are sensitive to higher-order multipole moments on the ligand. The experimental spectrum of photodissociated CO is split into two bands with different intensities, separated by 10 cm^{-1} and extending over ~20 cm^{-1} (full-width at half maximum, taken for each fitted Gaussian function separately). The more intense of the two bands (the one at higher frequency) is shifted to the red by 10 cm^{-1} relative to free CO (24).

In the following, the IR spectra from simulations with the three different electrostatic models (see Fig. 6) are discussed separately. The IR spectra are calculated from the autocorrelation function of the (fluctuating) dipole moment function along the trajectory according to Computational Methods (see above). Thus, the spectra displayed are directly comparable to experimental spectra and not power spectra. Differences between the models are characterized in terms of the shift with respect to free CO at 2183 cm^{-1}, the splitting of the

two bands, the bandwidths, and the relative band intensities. The results are summarized in Table 3.

Model A

The simplest model includes atomic multipoles up to the dipole on each atom. For this multipole model the Fe⋯OC orientation is preferred (see Fig. 5). The spectrum averaged over eight trajectories (*black line* in Fig. 6) is broad, extends from 2139 cm^{-1} to 2178 cm^{-1} (full-width at half maximum with respect to three Gaussian functions fitted to the spectrum) and does not show a clear splitting. The dominant peak is centered at ~2172 cm^{-1}. Together with the structural information from Fig. 5 (clear preference of Fe⋯OC over Fe⋯CO), this signal is tentatively assigned to the Fe⋯OC state.

Model B

This model includes atomic multipoles up to quadrupole. The frequency range is more narrow (full-width at half maximum from 2167 cm^{-1} to 2186 cm^{-1}) and the average spectrum shows a clear splitting. The vibrational band at higher frequency is not shifted with respect to free CO, in contrast to experiment (24). The splitting of the average spectrum is composed of the different shifts of the individual spectra and the spectra of single trajectories are either not split or show small splittings of ~4–10 cm^{-1}.

Model C

This model includes atomic multipoles up to octopole. The full-width at half maximum extends from 2153 cm^{-1} to 2184 cm^{-1}, which is intermediate between models A and B. The average spectrum shows a splitting and a shift of both bands compared to free CO, in qualitative agreement with the experimental data. Quantitatively, the shift is smaller (4 cm^{-1}) than the experimentally observed one (10 cm^{-1}). A more detailed analysis with additional data is given in
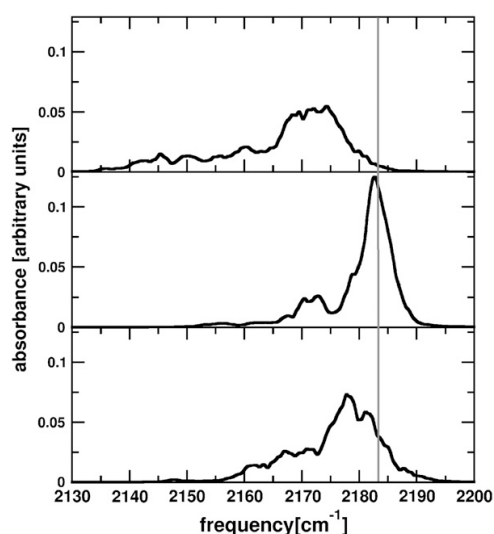

FIGURE 6 Average IR spectra over eight trajectories of 1 ns for each of the electrostatic models. The spectrum for model A (*top*) shows no clear splitting; broadening and red shift are too large. Model B (*center*) shows a splitting but no red shift compared to free CO. Model C (*bottom*) shows values for shifts and splitting in agreement with experiment. The shaded line indicates the frequency of free CO (2183 cm^{-1}).

TABLE 3 Characterization of the CO infrared spectra for the different electrostatic models

	His$_\delta$64			His$_\varepsilon$64	
	Model A	Model B	Model C	Model C	Exp. (24)
Splitting [cm^{-1}]	—	10.6	10.7 (13.8)	8.6	10
Shift [cm^{-1}]	11	0	4 (5)	3	10
Width [cm^{-1}]	39	19	31 (30)	32	32
Rel. intensities [%]	—	19	37 (46)	85	67

For His$_\delta$64 (model C), the number in brackets are averages over 16 trajectories, whereas for models A and B, the averages are over eight independent trajectories. For His$_\varepsilon$64 the averages are over 24 trajectories. The splitting is calculated from the fit to three Gaussian functions and the shift is reported for the band at higher frequency relative to free CO. Widths are full-widths at half maximum evaluated from the fit to three Gaussian functions. The relative band intensity is the maximum intensity of the band at lower frequency relative to the maximum intensity of the band at higher frequency.

the next section. In contrast to model B, the average splitting is the result of the splittings and shifts in the spectra of each individual trajectory, with splittings varying between 7 and 15 cm^{-1}.

Model C'

This model includes moments up to octopole on both atoms. It differs from model C only in the magnitude of the hexadecapole moment. With a splitting of ≈ 14 cm^{-1}, a red shift of 6 cm^{-1}, and a width of 33 cm^{-1}, the spectrum is very similar to the one from model C (see Table 3). The deviations are within the range of statistical errors estimated from comparing the results from 8 and 16 trajectories for model C (Table 3). This suggests that the spectra are converged with respect to the multipole order on the ligand. Also, these values are in quite favorable agreement with the experimental data.

Additional evaluation for the most accurate electrostatic model

For model C, which most accurately describes higher multipole moments of CO, eight additional trajectories for His$_\delta$64 each 1 ns in length have been evaluated to better converge the average spectra. Based on this data, the relation between the dynamics of the CO molecule and the IR spectra has been analyzed. To characterize the differences between protonation at His$_\delta$64 and His$_\varepsilon$64, additional simulations (24 trajectories of 1 ns length) were carried out for the His$_\varepsilon$64 protonation with model C.

Relationship between CO orientation and absorption band intensities

It was already previously noted that extracting splittings in the wavenumber range in the IR spectrum from MD simulations alone may be difficult (15) and obscured through sampling of a multitude of substates which give rise to slightly different vibrational frequencies (26). Thus, in the following, an attempt is made to correlate structural aspects with spectroscopic features.

His$_\delta$64

Each of the 16 trajectories samples both substates (Fe···CO and Fe···OC) but with a different probability density for the two states according to the criterion (magnitude of δ) discussed in Orientational preference and location of the B-state. Thus, the 16 trajectories for model C have been divided into three groups. The first group (group I, least preference for Fe···CO) contained seven trajectories with <60% of the Fe···CO population, group II had six trajectories with between 60 and 70%, and group III (high Fe···CO preference) had three trajectories with >70% in the Fe···CO substate. Averages of the IR spectra have been calculated over all individual groups and three Gaussian functions were found to faithfully represent the average spectrum. The average over

trajectories with dominant Fe···CO preference (group III) shows mainly one broad band, whereas the average over group I (less pronounced Fe···CO preference, and therefore larger contributions of the Fe···OC substate) shows a clear splitting into two bands. The main peaks of the two states are separated by 13 cm^{-1}. In Fig. 7, the average spectrum from group I is compared to the average over all trajectories. Thus, the signal at lower wavenumber which only appears if the Fe···OC state is sufficiently populated, is assigned to the B$_2$ state. These findings are in accord with the identification of the two bands as the Fe···CO (B$_1$) and Fe···OC (B$_2$) substate (21,22) and are at variance with the recently published identification of the lower band as the Fe···CO substate and the higher band as the Fe···OC substate (see also Discussion) (23).

His$_\varepsilon$64

The analysis of the orientations of CO with respect to iron for His$_\varepsilon$64 showed a different distribution of the substates. The population for all three electrostatic models is shifted toward larger contributions of the Fe···OC substate, with model C showing only a moderate preference for the Fe···CO orientation. The 24 trajectories were again divided into three groups as follows:

Group I (least preference for Fe···CO) contains six trajectories with <30% Fe···CO.
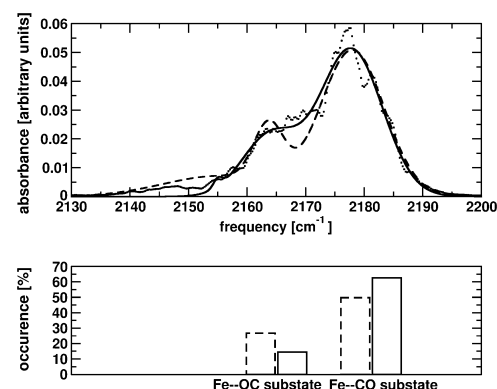Group II (intermediate) contains 12 trajectories.

FIGURE 7   His$_\delta$64: Relation between CO orientation with respect to iron and IR spectra. Average spectra are calculated over the seven trajectories with <60% of Fe···CO substate (group I, *dashed line*) and over all trajectories (*solid line*). The dotted line represents the raw (unfitted) data for the average over all trajectories. The sum of three Gaussian functions is fitted to the averages. The comparison shows that the shoulder at 2163 cm^{-1} of the average over all spectra can be identified with the Fe···OC substate, since it appears more pronounced in the dashed spectrum, where only trajectories with larger proportion of the Fe···OC substate are taken into account. The average percentage of each substate for the evaluated group I and for all trajectories is shown in the figure below. Dashed and solid lines in the diagram correspond to dashed and solid lines in the spectrum.

Group III (highest preference for Fe···CO) contains the remaining six trajectories with >42% of Fe···CO.

The averaged CO infrared spectra for each group are shown in Fig. 8. From this it is found that trajectories with a stronger preference for Fe···CO contribute more to the upper band (at 2180 cm^{-1}), whereas trajectories with a stronger preference for the Fe···OC orientations contribute more to the lower band (at 2172 cm^{-1}). This agrees with the findings for the His$_\delta$64 protonation state. The feature at ~2155 cm^{-1} can be related to CO positions close and above the iron center with high contributions of the Fe···OC substate.

DISCUSSION AND CONCLUSION

In this study the influence of higher multipole moments on the dynamics and spectroscopy of photodissociated CO in the B-state of myoglobin has been systematically investigated. The CO molecule is described by atom-centered multipole moments on the carbon and the oxygen atom which were derived from a distributed multipole analysis (DMA). The protein environment, on the other hand, is treated with conventional point charges. For CO, three increasingly detailed charge models were used which differ in their ability to reproduce the experimentally known molecular multipole moments. Model A correctly describes the dipole and quadrupole moment, whereas model B reproduces the octopole moment and model C gives a satisfactory hexadecapole moment (with some deficiencies for the octopole moment).

The results from the MD simulations show that all models with fluctuating moments which correctly describe the
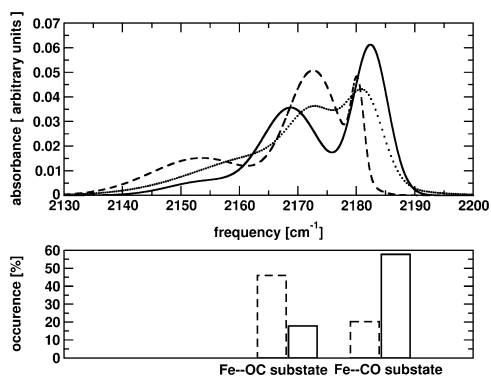

FIGURE 8   His$_\delta$64: Relation between CO orientation with respect to iron and IR spectra. Averages are calculated over six spectra belonging to trajectories with the highest population of Fe···CO substate (*solid line*), six spectra belonging to trajectories with the highest population of Fe···OC substate (*dashed*), and over all trajectories (*dotted*). The sum over three Gaussian functions is fitted to each average. The average percentage of each substate for the evaluated groups is shown in the figure below. Dashed and solid lines in the diagram correspond to dashed and solid lines in the spectrum.

quadrupole moment are capable of locating the docking site (B-state) in agreement with experimental data. On the other hand, the average orientation of the CO molecule with respect to the heme-Fe atom (Fe–CO versus Fe–OC) in the B-state sensitively depends upon the sign of the octopole moment Ω. For the simplest model A, the Fe–OC substate is the dominant one, whereas with models B and C, Fe–CO is the more probable state. This is consistent with previous simulations (16) which were based on a fluctuating charge model with $\Omega = 0$ (i.e., intermediate between model A and models B/C) where it was found that the Fe–CO state is more stable.

One particular advantage of molecular dynamics simulations is that the infrared spectrum can be calculated directly from the Fourier transform of the dipole autocorrelation function. Such a procedure includes the full dynamics and treats the energetics and spectroscopy at an equal level. This is not the case for the combined MD/DFT/quantum bound state (22) and the PMM (23) calculations. All models used in this work correctly describe $\mu(r)$ over the range of r which is sampled in the vibrational ground and low excited states (see Fig. 2). Thus, the differences between the infrared spectra can be directly related to how CO interacts with its environment which, in turn, influences the space sampled by the photodissociated ligand. Contrary to previous efforts to relate structural and spectroscopic features (21–23), this work retains a maximum of information about the dynamically changing environment around the ligand. Earlier work related to this question suffers from different shortcomings. In the combined MD/DFT/quantum bound state work the frequency distribution was based on 50 snapshots. Although this number is considerable given the computational effort to determine the CO stretching potentials, some uncertainty remains as to the convergence of the results. However, a statistical analysis suggested that the principal findings should not change dramatically upon increasing the number of snapshots used (22). The PMM study included 21,000 snapshots at the expense of only using a minimal QM region (the CO molecule) which was treated with B3LYP/aug-cc-pVTZ whereas all other atoms were represented by point charges (23). This may influence the CO-vibrations because they depend on how well the interactions between CO and the nearby heme and His-64 residues are described. Such effects can also be seen in this study and are discussed below.

The extensive MD simulations for both protonation states (δ, ε) of His-64 lead to a split IR spectrum except for the simplest charge model A, which is the least reliable one. From grouping the trajectories using model C into those which sample the Fe–CO more (>70%) or less (<60%) extensively it is possible to relate the signal at lower wavenumbers to the Fe–OC conformation and the band at higher wavenumbers to the Fe–CO state. This is the assignment also found from the MD/DFT/quantum bound state calculations. Further (indirect) support for this assignment is provided by the IR spectrum from model A which predominantly samples the Fe–OC state (see Fig. 6). The corresponding IR spectrum

has a large peak in the wavenumber region (2170 cm^{-1}) where models B and C have their low-intensity bands. The calculated splitting from models B and C is between 8.6 and 12.8 cm^{-1}, depending on the protonation state of His-64. This is in very good agreement with experiment which finds a splitting of 10 cm^{-1}. Furthermore, the overall width and shape of the spectra correspond to the experimentally observed ones. The average over all spectra for Model C evaluated for His$_{\delta}$64 and His$_{\varepsilon}$64 is compared to the experimental spectrum (24) in Fig. 9. For His$_{\delta}$64, the intensity of the lower band is slightly weaker than in the experiment, the splitting is too large by ∼3 cm^{-1} and the shift with respect to free CO is ∼5 cm^{-1} smaller. For His$_{\varepsilon}$64, the intensity of the lower band is stronger than in the experiment, the splitting is too small by ∼1 cm^{-1} and the shift with respect to free CO is ∼7 cm^{-1} smaller. The overall widths and lineshapes of the spectra (flat at the red end and steep at the blue end) closely reflect those observed experimentally (see Fig. 9) and found in previous simulations (16) and differs from the almost symmetric lineshape found in the PMM simulations (23) which may be related to the small size of the QM region (CO molecule) and to the fact that the harmonic approximation was used in calculating the CO frequencies. For both His$_{\delta}$64 and His$_{\varepsilon}$64, the relative band intensities correctly find the lower band to be less intense than the band at higher frequency. Overall, the results for His$_{\varepsilon}$64 reflect the experimental data somewhat better than for His$_{\delta}$64. However, the difference between the results for His$_{\delta}$64 and His$_{\varepsilon}$64 does not only depend on the electrostatic model for the CO molecule, but also on the parameterization of the histidine residue. His-64 is described here by a conventional point charge model,
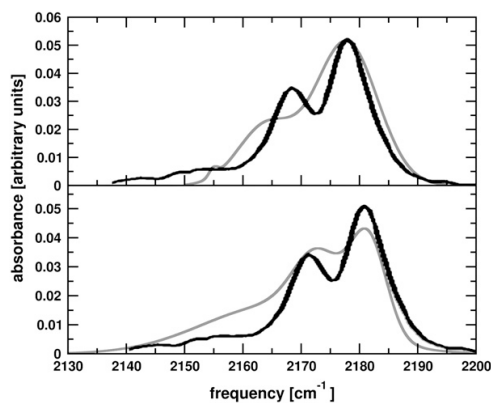
which may be insufficient to capture additional details of the two protonation states. This could be an explanation for the larger difference between the two states observed in previous QM/MM calculations including the CO molecule and the binding site environment in the quantum-mechanical part (22) where His-64 was treated quantum mechanically. To fully understand these discrepancies, more accurate descriptions of the entire binding site will be required.

One further comparison with experimental data can be made from these simulations. Using pump-probe experiments it has been observed that the center frequency of the CO stretch shifts as a function of the time after photodissociation (11,57–59). This was attributed to the memory of the CO molecule of the bound state. Upon photodissociation, the CO fundamental frequency changes from ≈1950 cm^{-1} (bound or A-state) to ≈2130 cm^{-1} (unbound state). This change occurs on a timescale which is considerably longer than the photodissociation process itself. Experimentally, the spectra for times 0.3–100 ps after photodissociation (57) show a blue shift of ≈4 cm^{-1} whereas at longer times (10–100 ns after dissociation) the band shifts by another ≈2 cm^{-1} to the blue (11). From the simulations, the upper band of the IR spectrum at times 100 ps, 200 ps, 500 ps, and 1 ns after photodissociation is found to gradually shift to the blue by 10 cm^{-1}. The center frequency of the upper band is located at 2168 cm^{-1} after 100 ps and shifted to 2178 cm^{-1} at 1 ns, which is consistent with the experimental data for the high-frequency band (57). However, quantitatively the calculated blue shift of 10 cm^{-1} between 100 ps and 1 ns is too large. Possible reasons are the different quality of the stretching potential for bound and unbound CO and the fact that the electrostatic interaction between CO and the surrounding protein influences the stretching frequency. Furthermore, short dipole-dipole autocorrelation functions (for times early after photodissociation) are less reliable and lead to an artificial broadening of the spectra for short timescales. Thus, the uncertainty in the maximum absorption frequency is larger for early times compared to later times.

The IR spectra calculated for the different electrostatic models show a trend toward better agreement with experiment as the accuracy of the electrostatic models is improved. This is observed for the splittings, the red shifts with respect to free CO, the bandwidths, and the relative band intensities (Table 3). A comparison of all these features reveals that overall, model C, as the most accurate electrostatic model, shows the best agreement with experiment. Thus, it is found that using systematically improved descriptions of the electrostatics around CO yields improved calculations for structural and spectroscopic features of the CO dynamics in the main binding site of myoglobin.

FIGURE 9   Comparison of calculated and experimental spectra (Fig. 4 *B* from (24)). The computed spectra (*shaded*) are averages over 16 trajectories for His$_{\varepsilon}$64 (*top*) and 24 trajectories for His$_{\delta}$64 (*bottom*). In both cases, three Gaussian functions have been fitted to the average spectrum. For direct comparison, the maximum of the high frequency band from the calculated and the experimentally observed spectrum (24) is separately aligned for both cases. The width and the splitting for the spectrum with His$_{\varepsilon}$64 better reproduces the experimental data.

REFERENCES

1. Williams, P., J. Cosme, and A. Ward. 2003. Crystal structure of human cytochrome P4502C9 with bound warfarin. *Nature*. 242:464–468.

2. Reibarkh, M., T. Malia, and G. Wagner. 2006. NMR distinction of single- and multiple-mode binding of small-molecule protein ligands. *J. Am. Chem. Soc.* 128:2160–2161.

3. Kraut, J., P. Sigala, B. Pybus, C. Liu, D. Ringe, G. Petsko, and D. Herschlag. 2006. Testing electrostatic complementarity in enzyme catalysis: hydrogen bonding in the ketosteroid isomerase oxyanion hole. *PLoS Biol.* 4:501–519.

4. Campbell, S., N. Gold, R. Jackson, and D. Westhead. 2003. Ligand binding: functional site location, similarity and docking. *Curr. Opin. Struct. Biol.* 13:389–395.

5. Kraut, J. 1988. How do enzymes work? *Science.* 242:533–540.

6. Wolfenden, R. 2003. Thermodynamic and extrathermodynamic requirements of enzyme catalysis. *Biophys. Chem.* 105:559–572.

7. Warshel, A., P. K. Sharma, M. Kato, Y. Xiang, H. B. Liu, and M. H. M. Olsson. 2006. Electrostatic basis for enzyme catalysis. *Chem. Rev.* 106: 3210–3235.

8. Schlichting, I., J. Berendzen, G. Phillips, and R. Sweet. 1994. Crystal structure of photolyzed carbonmonoxy-myoglobin. *Nature.* 371:808–812.

9. Teng, T., V. Srajer, and K. Moffat. 1994. Photolysis-induced changes. *Nat. Struct. Biol.* 1:701–705.

10. Ostermann, A., R. Waschipky, F. Parak, and G. Nienhaus. 2000. Ligand binding and conformational motions in myoglobin. *Nature.* 404:205–208.

11. Schotte, F., M. Lim, A. Jackson, V. Smirnov, J. Soman, J. Olson, G. Phillips, M. Wulff, and P. A. Anfinrud. 2003. Watching a protein as it functions with 150-ps time-resolved x-ray crystallography. *Science.* 300:1944–1947.

12. Srajer, V., Z. Ren, T. Teng, M. Schmidt, T. Ursby, D. Bourgeois, C. Pradervand, W. Schildkamp, M. Wulff, and K. Moffat. 2001. Protein conformational relaxation and ligand migration in myoglobin: a nanosecond to millisecond molecular movie from time-resolved Laue x-ray diffraction. *Biochemistry.* 40:13802–13815.

13. Straub, J. E., and M. Karplus. 1991. Molecular dynamics study of the photodissociation of carbon monoxide from myoglobin: ligand dynamics in the first 10 ps. *Chem. Phys.* 158:221–248.

14. Ma, J., S. Huo, and J. Straub. 1997. Molecular dynamics simulation study of the B-states of solvated carbon monoxymyoglobin. *J. Am. Chem. Soc.* 119:2541–2551.

15. Meller, J., and R. Elber. 1998. Computer simulations of carbon monoxide photodissociation in myoglobin: Structural interpretation of the B states. *Biophys. J.* 74:789–802.

16. Nutt, D. R., and M. Meuwly. 2003. Theoretical investigation of infrared spectra and pocket dynamics of photodissociated carbonmonoxy myoglobin. *Biophys. J.* 85:3612–3623.

17. Nutt, D. R., and M. Meuwly. 2004. CO migration in native and mutant myoglobin: atomistic simulations for the understanding of protein function. *Proc. Natl. Acad. Sci. USA.* 101:5998–6002.

18. Bossa, C., M. Anselmi, D. Roccatano, A. Amadei, B. Vallone, M. Brunori, and A. Di Nola. 2004. Extended molecular dynamics simulation of the carbon monoxide migration in sperm whale myoglobin. *Biophys. J.* 86:3855–3862.

19. Danielsson, J., P. Banushkina, D. Nutt, and M. Meuwly. 2006. Computer simulations of structures, energetics and dynamics of myoglobin···ligand complexes. *Int. Rev. Phys. Chem.* 25:407–425.

20. Meuwly, M. 2007. Using small molecules to probe protein cavities: the myoglobin-XO (X = C, N) family of systems. *Eur. Phys. J.* 141:209–216.

21. Nienhaus, K., J. S. Olson, S. Franzen, and G. U. Nienhaus. 2005. The origin of stark splitting in the initial photoproduct state of MbCO. *J. Am. Chem. Soc.* 127:40–41.

22. Meuwly, M. 2006. On the influence of the local environment on the CO stretching frequencies in native myoglobin: assignment of the B-states in MbCO. *ChemPhysChem.* 10:2061–2063.

23. Anselmi, M., M. Aschi, A. Di Nola, and A. Amadei. 2007. Theoretical characterization of carbon monoxide vibrational spectrum in sperm whale myoglobin distal pocket. *Biophys. J.* 92:3442–3447.

24. Lim, M., T. A. Jackson, and P. A. Anfinrud. 1995. Mid-infrared vibrational spectrum of CO after photodissociation from heme: evidence of a docking site in the heme pocket of hemoglobin and myoglobin. *J. Chem. Phys.* 102:4355–4366.

25. Lim, M., T. A. Jackson, and P. A. Anfinrud. 1997. Ultrafast rotation and trapping of carbon monoxide dissociated from myoglobin. *Nat. Struct. Biol.* 4:209–214.

26. Nutt, D., P. Banushkina, and M. Meuwly. 2005. Computational chemistry for elucidating protein function: Energetics and dynamics of myoglobin-ligand systems. *Chimia (Aarau).* 59:517–521.

27. MacKerell, A. D., Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. E. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* 102:3586–3616.

28. Weiner, S. J., P. A. Kollman, D. A. Case, U. Singh, C. Ghio, G. Alagona, S. Profeta, Jr., and P. Weiner. 1984. A new force-field for molecular mechanical simulation of nucleic-acids and proteins. *J. Am. Chem. Soc.* 106:765–784.

29. Mulliken, R. 1955. Electronic population analysis on LCAO-MO molecular wave functions. *J. Chem. Phys.* 23:1833–1840.

30. Singh, U. C., and P. A. Kollman. 1984. An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* 5:129–145.

31. Besler, B. H., K. M. Merz, and P. A. Kollman. 1990. Atomic charges derived from semiempirical methods. *J. Comput. Chem.* 11:431–439.

32. Stone, A. J., and M. Alderton. 1985. Distributed multipole analysis—methods and applications. *Mol. Phys.* 56:1047–1064.

33. Sokalski, W., and R. Poirier. 1983. Cumulative atomic multipole representations of the molecular charge distribution and its basis set dependence. *Chem. Phys. Lett.* 98:86–92.

34. Mankoo, P. K., and T. Keyes. 2006. Induction model for molecular electrostatics: application to the infrared spectroscopy of CO liquid. *J. Chem. Phys.* 124:204503.

35. Stoll, J., J. Vrabec, and H. Hasse. 2003. A set of molecular models for carbon monoxide and halogenated hydrocarbons. *J. Chem. Phys.* 119: 11396–11407.

36. Kuriyan, J., S. Wilz, M. Karplus, and G. Petsko. 1986. X-ray structure and refinement of carbon-monoxy (Fe-II)-myoglobin at 1.5-Å resolution. *J. Mol. Biol.* 192:133–154.

37. Meuwly, M., O. Becker, R. Stote, and M. Karplus. 2002. NO rebinding to myoglobin: a reactive molecular dynamics study. *Biophys. Chem.* 89:183–207.

38. Huffaker, J. N. 1976. Diatomic molecules as perturbed Morse oscillators. I. Energy levels. *J. Chem. Phys.* 64:3175–3181.

39. Johnson, J., D. Lamb, H. Frauenfelder, J. Muller, B. McMahon, G. Nienhaus, and R. D. Young. 1996. Ligand binding to heme proteins. 6. Interconversion of taxonomic substates in carbonmonoxymyoglobin. *Biophys. J.* 71:1563–1573.

40. Rovira, C., B. Schulze, M. Eichinger, J. D. Evanseck, and M. Parrinello. 2001. Influence of the heme pocket conformation on the structure and vibrations of the Fe-CO bond in myoglobin: a QM/MM density functional study. *Biophys. J.* 81:435–445.

41. Merchant, K. A., W. G. Noid, D. E. Thompson, R. Akiyama, R. F. Loring, and M. D. Fayer. 2003. Structural assignments and dynamics of the A substates of MbCO: spectrally resolved vibrational echo experiments and molecular dynamics simulations. *J. Phys. Chem. B.* 107:4–7.

42. Allen, M. P., and D. J. Tildesley. 1987. Computer Simulation of Liquids. Clarendon Press, Oxford.

43. Berens, P. H., and K. R. Wilson. 1981. Molecular-dynamics and spectra. 1. Diatomic rotation and vibration. *J. Chem. Phys.* 74:4872–4882.

44. Stone, A. J. 2005. Distributed multipole analysis: stability for large basis sets. *J. Chem. Theory Comput.* 1:1128–1132.

45. Lebedev, V., and D. N. Laikov. 1999. Quadrature formula for the sphere of 131th algebraic order of accuracy. *Dokl. Math.* 59:477–481.

46. Stone, A. J., A. Dullweber, O. Engkvist, E. Fraschini, M. P. Hodges, A. W. Meredith, D. R. Nutt, P. L. A. Popelier, and D. J. Wales. 2002. ORIENT: a program for studying interactions between molecules. Available at http://www.stone.ch.cam.ac.uk/programs.html#Orient.

47. Frisch, M. J., G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. E. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, and K. Raghavachari. J. B. F. Resman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. A. Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. 2004. Gaussian 03, Rev. C.01. Gaussian Incorporated, Wallingford, CT.

48. Becke, A. D. 1993. Density-functional thermochemistry. 3. The role of exact exchange. *J. Chem. Phys.* 98:5648–5652.

49. Dunning, T. H., Jr. 1989. Gaussian-basis sets for use in correlated molecular calculations. 1. The atoms boron through neon and hydrogen. *J. Chem. Phys.* 90:1007–1023.

50. Maroulis, G. 1996. Electric polarizability and hyperpolarizability of carbon monoxide. *J. Phys. Chem.* 100:13466–13473.

51. Maroulis, G. 2001. Accurate higher electric multipole moments for carbon monoxide. *Chem. Phys. Lett.* 334:214–219.

52. Stone, A. J. 1996. The Theory of Intermolecular Forces. Clarendon Press, Oxford, UK.

53. Vitkup, D., G. A. Petsko, and M. Karplus. 1997. A comparison between molecular dynamics and x-ray results for dissociated CO in myoglobin. *Nat. Struct. Biol.* 4:202–208.

54. Teng, T., V. Srajer, and K. Moffat. 1997. Initial trajectory of carbon monoxide after photodissociation from myoglobin at cryogenic temperatures. *Biochemistry.* 36:12087–12100.

55. Vojtechovsky, J., K. Chu, J. Berendzen, G. Phillips, R. Sweet, and I. Schlichting. 1999. Crystal structures of myoglobin-ligand complexes at near-atomic resolution. *Biophys. J.* 77:2153–2174.

56. Phillips, G., M. Teodoro, T. Li, B. Smith, and J. Olson. 1999. Bound CO is a molecular probe of electrostatic potential in the distal pocket of myoglobin. *J. Phys. Chem. B.* 103:8817–8829.

57. Kim, S., and M. Lim. 2005. Picosecond dynamics of ligand interconversion in the primary docking site of heme proteins. *J. Am. Chem. Soc.* 127:5786–5787.

58. Lin, M., T. A. Jackson, and P. Anfinrud. 2004. Orientational distribution of CO before and after photolysis of MbCO and HbCO: a determination using time-resolved polarized mid-IR spectroscopy. *J. Am. Chem. Soc.* 126:7946–7957.

59. Anfinrud, P., M. Lim, and T. Jackson. 1994. Femtosecond IR spectroscopy: methods and applications to protein dynamics. *Proc. SPIE Int. Soc. Opt. Eng.* 2138:107–115.

4.2 Atomistic Simulations of CO Vibrations in Ices Relevant to Astrochemistry

N. Plattner and M. Meuwly

Atomistic Simulations of CO Vibrations in Ices Relevant to Astrochemistry

Nuria Plattner and Markus Meuwly*[a]

The experimental absorption band of carbon monoxide (CO) in mixed ices has been extensively studied in the past. The astrophysical interest in this band is related to its characteristic shape, which appears to depend on the surrounding ice structure. Herein, molecular dynamics simulations are carried out to analyze the relationship between the structure of the ice and the infrared (IR) spectrum of embedded CO molecules at different concentrations. Instead of conventional force fields, anharmonic potentials are used for the bonded interactions. The electrostatic interactions are more accurately described by means of fluctuating atomic multipole moments (up to quadrupole). The experimental-ly observed splitting of the CO absorption band (gas phase: 2143 cm^{-1}) into a blue- (2152 cm^{-1}) and a red-shifted (2138 cm^{-1}) signal is also found in the simulations. Complementary atomistic simulations allow us to relate the spectra with the structural features. The distinction between interstitial and substitutional CO molecules as the origin of this splitting is found to be qualitatively correct. However, at increasing CO concentrations, additional effects—such as mutual interactions between CO molecules—become important, and the simplistic picture needs to be revised.

1. Introduction

Water and carbon monoxide (CO) are abundant molecules in inter- and circumstellar grain mantles. They occur in dense zones of the interstellar medium, such as star-forming regions and dark clouds. Both water and CO have been detected in these areas by using infrared spectroscopy, their presence being confirmed by comparison with laboratory work. Interstellar water ice was first identified through a strong band at 3279 cm^{-1} (the symmetric OH-stretching motion). This band was later assigned to water ice by detailed laboratory work.[1] In the case of carbon monoxide, the characteristic absorption band around 2139 cm^{-1} was assigned to solid CO.[2]

The infrared spectrum of CO incorporated in H_2O ice at 10 K shows two features at about 2138 and 2152 cm^{-1}. The two bands are referred to as the nonpolar and polar components, respectively, and are related to CO molecules interacting with species, such as H_2O or NH_3, through (polar) hydrogen bonds and to CO molecules interacting with each other (or with CO_2) through (nonpolar) van der Waals interactions. Upon heating, the high-frequency band disappears and the main band shifts from 2138 to 2136 cm^{-1}.[3,4] The 2152 cm^{-1} band appears to be absent in astronomical spectra, which is attributed to the fact that in interstellar ices, the dangling-OH binding sites either are not accessible to CO molecules or do not exist.[5] Laboratory studies show that the exact frequency and shape of the CO absorption band varies depending on the particular ice environment. The CO absorption band has been measured in ices composed of water, CO, and a third species including methane, CO_2 or acetone.[6–8] Furthermore, CO absorption spectra have also been measured for different CO/water fractions and at different temperatures.[3,6,9] These studies establish that composition and thermodynamic conditions influence the shape, width, and splitting of the CO absorption band.

Herein, we focus on the splitting of the CO absorption band observed at temperatures below 70 K in water-rich ice. Ices of such compositions are known to be amorphous,[3,10] which makes direct crystallographic characterization difficult (if not impossible). Nevertheless, splitting into two discrete bands suggests that despite the lack of global crystal symmetry, a locally defined structure may be possible. Historically, splitting of the CO absorption band has been attributed to CO being either substitutional (i.e. replacing one water molecule), which leads to the red-shifted band, or interstitial (occupation of a void), which should lead to the blue-shifted band.[4] Herein, we use atomistic simulations to characterize the CO vibration in a heterogeneous environment and correlate the spectroscopic signatures with structural features. The splittings and shifts of the infrared bands are sufficiently small to warrant the use of detailed interaction models. It has been shown that representing the electrostatic field around a CO molecule by including higher multipole moments allows one to carry out atomistic simulations, which provide quantitative information, and assist assignments of experimental data.[11–13] Herein, we use the best-available CO model (without modification) in simulations of CO in ices and assess a number of electrostatic representations with rigid and flexible water models. The work is structured as follows: First, we explain the computational methods and simulation systems used. Then, we describe our results on the structures and infrared spectra, together with the radial

[a] N. Plattner, Prof. M. Meuwly
Department of Chemistry, University of Basel
Klingelbergstrasse 80, 4056 Basel (Switzerland)
Fax: (+41) 61 267 48 55
E-Mail:
E-mail: m.meuwly@unibas.ch

CHEMPHYSCHEM

distribution functions, and finally, we discuss our findings in view of the experimental data and draw our conclusions.

Experimental and Computational Methods

Interaction Potentials and Parametrizations: All simulations were carried out using the CHARMM program,[14] with provisions for evaluating electrostatic interactions involving higher distributed multipoles (DMs, see below). The CO bond potential was described by a spectroscopically accurate anharmonic RRKR (rotational Rydberg–Klein–Rees) potential,[15,16] which has the functional form of a distorted Morse potential, and the electrostatic interactions were described by atomic multipole moments based on the distributed multipole analysis (DMA),[17,18] which accurately represent the molecular moments up to hexadecapole[13] (see Table 1).

Table 1. Molecular moments of CO calculated from the distributed multipoles on the C and O atoms. The multipole moments (in atomic units) are evaluated for the equilibrium structure of CO.

	μ [ea_0]	Θ [ea_0^2]	Ω [ea_0^3]	Φ [ea_0^4]
DMA[13]	0.04	−1.51	2.86	−9.03
CCSD(T)[33,34]	0.051	−1.47	3.46	−9.07
experiment[a]	0.048[b]	−1.58[c]	4.15	−8.15

[a] The values vary somewhat between experiments; for detailed information, see refs. [33,34]. [b] The experimental values converge to this number. [c] This value varies between −1.4 and −1.68 ea_0^2 for different experiments.

For the water molecule, both a rigid (TIP3P[19]) and a flexible [the Kumagai, Kawamura, and Yokokawa (KKY) potential[20]] model were used. As was previously reported, the KKY potential parameters need to be reparametrized to reproduce the water gas-phase frequencies.[21] In the simulations with TIP3P, the intramolecular degrees of freedom in water were constrained by using the SHAKE algorithm.[22] For the electrostatic interactions, two alternatives were used, namely, TIP3P-standard point charges and DMs including point charges plus additional atomic moments up to quadrupole on oxygen, and up to dipole on hydrogen (see Table 2). Four different models (intramolecular degrees of freedom/electrostatics) were considered: A) TIP3P/TIP3P, B) TIP3P/DMA, C) KKY/TIP3P, and D) KKY/DMA. It should be noted that model C is imbalanced because the charges are not parametrized for use with flexible water. It has been shown that in the case of flexible bonds, the nonbonded parameters need to be adapted.[23] Nevertheless, this model was used to test whether markedly different results were found by using an imbalanced representation for the water molecules. All simulations were carried out with periodic boundary conditions

Table 2. Electrostatic moments for the water multipole model. The values correspond to a spherical tensor notation.

	Charge [e]	Dipole [ea_0]		Quadrupole [ea_0^2]	
		q10	q11s	q20	q22c
oxygen	−0.700	−0.131	–	−0.030	−0.800
hydrogen	0.350	0.010	−0.010[a]	–	–

[a] Opposite signs on the two hydrogen atoms.

and a simulation box with dimensions: $35 \times 31 \times 29$ Å. For the rigid water models (namely, A and B), the integration time step was 1 fs. This time was reduced to 0.4 fs for the simulations carried out using the flexible water models (C and D).

The infrared (IR) spectrum was calculated from the Fourier transform of the dipole-moment autocorrelation function $C(t)$, which was accumulated over 2^n time origins, where n is an integer such that 2^n corresponds to between 1/3 and 1/2 of the trajectory, with the time origins being separated by 1 fs. The function $C(t)$ was then transformed to $C(\omega)$, thereby using a fast Fourier transform with a Blackman filter to minimize noise.[24] The final infrared absorption spectrum $A(\omega)$ was then calculated from Equation (1):

$$A(\omega) = \omega\{1 - \exp[-\omega/(kT)]\}C(\omega) \tag{1}$$

where k is the Boltzmann constant and T is the temperature in Kelvin.

To compare the calculated band positions and their shifts with the experiments, we calculated reference stretching frequencies for gas-phase CO using the RRKR potential. The calculation included ten trajectories for free CO with time steps of 0.4 and 1 fs. The free-CO stretching band was found at 2187 cm^{-1} in the simulations carried out using a time step of 1 fs (for models A and B) and at 2172 cm^{-1} when a time step of 0.4 fs was used (models C and D). Deviations between the ten trajectories were small (0.2 cm^{-1}). The offsets of 44 cm^{-1} (between 2187 cm^{-1} obtained in the simulations with $\Delta t = 1$ fs and the experimental value of 2143 cm^{-1} [25]) and 29 cm^{-1} (between the simulated value of 2172 cm^{-1} obtained with $\Delta t = 0.4$ fs and the experimental value) have two different origins. First, simulations of CO in argon[26] have shown that integration of the equations of motion using the velocity Verlet algorithm with a time step of $\Delta t = 1$ fs systematically shifts the frequency by 16 cm^{-1}. When smaller time steps are used, the error and frequency shift decrease. The remaining 28 cm^{-1} (for $\Delta t = 1$ fs) can be attributed to the classical treatment, which for a Morse oscillator [$V(r) = D_e(1 - \exp[-\alpha(r-r_e)])^2$], has an energy-dependent fundamental frequency ω_E [Eq. (2)]:

$$\omega_E = \alpha\sqrt{\frac{2(D_e - E_{tot})}{\mu}} \tag{2}$$

where E_{tot} is the total energy of the particular simulation, D_e is the dissociation energy, r_e is the equilibrium bond length and α describes the steepness of the repulsive wall. All IR spectra discussed below are compared with the calculated spectrum for free CO (2187 or 2172 cm^{-1}).

Structural Setups: The simulations were started from different initial structures. To test the hypothesis that CO molecules at interstitial and substitutional sites are responsible for the two main frequencies, CO molecules were positioned in a hexagonal ice structure. This setup is referred to as initial structure (I). Hexagonal ice was chosen as a starting point, since it is the most probable structure for pure water ice in the low-temperature and low-pressure region of the phase diagram. Other structures of water ice include cubic ice (metastable) and ice XI, of which the formation is kinetically very unfavorable in pure water ice. The CO positions were chosen either at the position of a water molecule (substitutional) or in voids between the water molecules (interstitial). This setup was made for one and for five CO molecules of either type (two and ten CO molecules in total).

More realistic—that is, amorphous—structures were generated as follows:

II) Starting from a hexagonal ice structure containing 1024 water molecules, between 40 and 60 CO molecules (to reflect the H$_2$O:CO ratio used in the experiments) were positioned in a random fashion.

III) Starting from the structures generated in (I), all the water molecules within a distance $\rho \in$ 0, 1.5, 1.8, and 2.3 Å of any CO molecule were removed. These structures were equilibrated for 50 ps at 20 K. This setup leads to H$_2$O:CO ratios between 10:1 and 20:1. Such small amounts of CO ensure that the environment of each CO molecule is primarily composed of water molecules, so that spectral features can be predominantly related to CO–water interactions. Experimental data corresponding to such ratios are described in the literature.[6] This initial structure was equilibrated during 50 ps at 20 K.

IV) The structure obtained in (III) with $\rho = 1.8$ was equilibrated by simulated annealing: After heating to 200 K, the structure was equilibrated for 50 ps at 150 K (setup IV.1), and 50 K (setup IV.2). The initial ordering of the water molecules is rapidly lost during the heating process; structural elements start to reform locally at low temperatures.

Subsequently, molecular dynamics simulations were run from structures generated according to procedures (II) to (IV). Starting from these initial structures, each trajectory included 20 ps of heating to the target temperature (20 or 60 K), followed by 100 ps of production run in the (*NVT*) ensemble. The results of the simulations performed with 40 CO molecules are qualitatively similar to those obtained with 60 CO molecules and are thus not discussed.

2. Results and Discussion

2.1. IR Spectra

Results obtained for the IR frequencies and the CO molecules corresponding to the different vibrations are shown in Figure 1 for the initial setup (I). For two CO molecules in the ice structure (Figure 1 A), a substitutional CO is responsible for the low-frequency band, whereas an interstitial CO molecule gives the high-frequency band. This agrees with the original hypothesis (see Introduction). As more CO molecules are inserted (see Figure 1 B,C), the frequency bands broaden and split and can no longer be attributed to two clearly distinguishable positions. Some frequencies of interstitial CO molecules are blue-shifted, while some frequencies of substitutional CO molecules are red-shifted, but most frequencies are in-between (the arrows relate frequencies and CO molecules within the ice structure), with no clear differentiation between interstitial and substitutional CO molecules. For higher CO concentrations (namely, 60 CO molecules, see Figure 1, right), individual CO lines are replaced by two broad bands with peaks at 2179 cm^{-1} (red-shifted by 8 cm^{-1} from the free-CO peak) and at 2192 cm^{-1} (blue-shifted by 5 cm^{-1} from the free-CO peak). The initial hexagonal ice structure is still preserved to some extent.

To investigate whether these effects can also be found in experimentally and astrophysically more relevant ice structures, we also carry out simulations for setups II to IV. Figure 2 shows the IR spectrum of CO in ice (with model B) for different H$_2$O:CO mixtures using initial structures generated according to setup III (see above). The mixtures lead to CO concentrations of 5.5, 5.7, 5.9, and 6.9 % and densities of 1.06, 1.03, 1.00, and 0.85 g mL^{-1}. All the spectra are broad and exhibit one or two main peaks. For lower densities, the low-frequency peak is considerably more pronounced, whereas for increasing densities, the absorption is shifted to the blue.

In the case of setup III, the IR spectra obtained from the MD simulations for models A, B, and D are all qualitatively similar. They show a split CO spectrum in which one band is blue-shifted and the other one is red-shifted at 20 K. At 60 K, the intensity of the blue-shifted band is significantly reduced or disappears altogether, which is in agreement with the experiments.[6,9] The intensities and positions of the absorption bands vary somewhat between different setups [(II) to (IV)] for the same model and between the models. Representative results for models B and D and setups III, IV.1, and IV.2 are summarized in Table 3. These results show that the splitting is reproduced for all setups at 20 K. Compared to the experiments, the redshift is generally slightly too large and the blueshift slightly too small. In contrast to models B and D, the splitting disappears in the simulations of model A with setup IV. For model C, no clear splittings are found, and the blueshift of the remaining absorption is either very small or negligible, even at low temperatures.

In summary, interaction models B and D (electrostatics with DMA) yield IR spectra that are in good agreement with the experimental data, whereas model C is found to disagree in all setups. In the case of model A, the results are inconclusive.

2.2. Structural Differences between Different Setups and Water Models

Quantitative characterization of the structural differences between the models (A to D) and setups (I to IV) is achieved by means of radial distribution functions (RDFs). The RDFs that characterize the global structure of the ice are water oxygen–oxygen $g_{OO}(r)$, oxygen–hydrogen $g_{OH}(r)$, and hydrogen–hydrogen $g_{HH}(r)$ distributions. A comparison of $g_{OO}(r)$ for the four different interaction models using setup IV.2 at 20 K is shown in Figure 3 A. The main difference between the water models is the position of the first maximum, which varies between 2.71 and 2.76 Å (see insets in Figure 3). Experimentally, the maximum is found around 2.8 Å at room temperature.[27] The location of the first maximum is dependent on the ice structure, the temperature, and the pressure. In constant-*p* simulations, its location depends on the electrostatics and the bond potential used for water. A comparison between setups III and IV.2 (using the most accurate interaction model, D) is shown in Figure 3 B for simulations at 20 and 60 K. Structural changes are observed between the trajectories at the two temperatures. More pronounced changes occur during simulated annealing, which is shown by the difference between solid and dashed lines in Figure 3 B. The long-range ordering is less pronounced for setup IV.2 than for setup III, which is attributed to the fact that starting from a hexagonal ice structure, not all structural
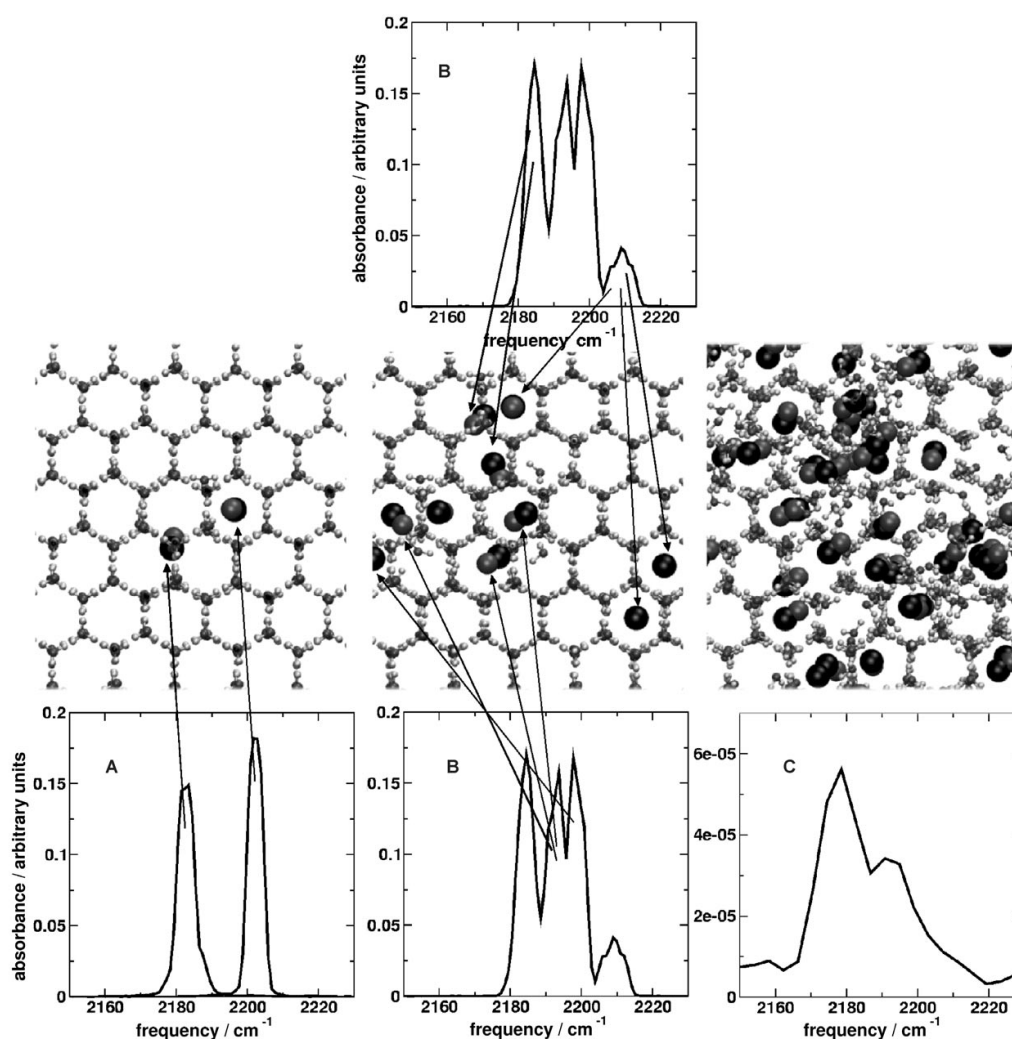
61

M. Meuwly and N. Plattner

Figure 1. Relationship between structure and IR spectrum. Panels A to C show different structures after 100 ps of simulation (at 20 K) together with the corresponding IR spectra. The structure in (A) is obtained by inserting one CO molecule at an interstitial position and another one as replacement for a water molecule; in (B) this is done for five CO molecules of each type; C) is obtained by random insertion of 60 CO molecules.

features of the ice disappear when setup III is used instead of setup IV.2, where simulated annealing destroys most of the long-range order. Overall, the structural differences are small and seem to depend more on the initial setup (see Figure 3 B) than on the water model.

To better understand the origin of the split IR spectra, we analyze the distribution functions of both the water-oxygen and the hydrogen atoms around the CO carbon [i.e. $g_{O(W)C(CO)}(r)$ and $g_{H(W)C(CO)}(r)$] and the CO oxygen [i.e. $g_{O(W)O(CO)}(r)$ and $g_{H(W)O(CO)}(r)$]. The results obtained for models A to D using setup IV.2 are shown in Figure 4 A, B. In the region between 3 and 4 Å, several distinct maxima are found for models B (grey trace in Figure 4 A) and D (grey trace in Figure 4 B). The positions of

these maxima vary not only between different models, but also between different setups for the same models. Thus, the characteristic feature is the existence of defined local structures around the CO molecules rather than their geometry. RDFs for models A and C are much more uniform, which suggests that the water environment is similar for all CO molecules. The same applies to the simulations performed with all models at 60 K. One example for model D is shown in Figure 4 C. Thus, the existence of locally defined structures is most likely to be related to the split CO frequency, which for this particular simulated annealing run is observed only for models B and D. If the $g_{O(W)O(CO)}(r)$ functions are calculated for simulations starting from setup III, models A, B, and D show distinct
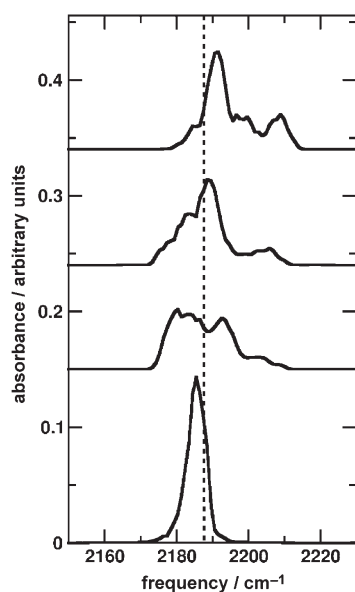
Figure 2. Relationship between structure and frequency. The spectra are evaluated for different CO/water compositions [densities of 1.06 (top), 1.03, 1.00, and 0.85 g mL^{-1} (bottom)] from simulations carried out using model B. The dashed line indicates the position of free CO at 20 K.

Figure 3. Water oxygen–oxygen radial distribution functions (RDFs) between 2.4 and 6 Å. A) Comparison of different water models for setup IV.2 at 20 K. Model A (black), model B (black, dashed), model C (grey), model D (grey, dashed). B) Comparison of RDFs for the most accurate model (i.e. model D). The solid lines represent setup IV.1 and the dashed lines indicate setup III. Grey is used for the trajectories at 20 K and black for the trajectories at 60 K. The insets show the first maximum of the distribution functions in more detail.

peaks (such as the grey traces in Figure 4 A, B) in the RDFs at 20 K, which are partially identified as structural elements of the initial hexagonal ice structure. The corresponding infrared spectra of CO are split for all three simulations. On the contrary, the simulations performed with model C show neither distinct peaks in the RDF nor splitting of the spectrum for any setup.

In summary, a split IR spectrum is observed under two conditions: For models B and D (DMA for electrostatics), splitting always occurs at low temperatures, whereas in the case of models A (TIP3P), B, and D, splitting will only be observed if the simulations start from structures that still contain residual ordering (hexagonal). For model C and all setups, no splitting is observed. Rearrangement of the structures around the inserted CO molecules is dominated by the molecular electrostatics of the water model. Splitting of the CO stretch is observed if locally defined structures around each CO molecule
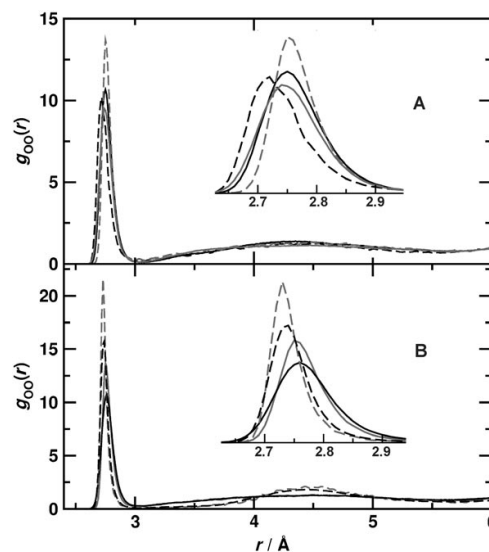
exist and persist during the simulations. These structures can be formed by rearrangement of the ice, if the corresponding water model accurately describes the molecular electrostatics, or they can result from incomplete rearrangement of the initial hexagonal ice, as observed for model A. In the case of model C, the ice rearranges very rapidly and does not form locally defined structures. Consequently, no splitting is observed.

2.3. Structure–Frequency Relationships

Analysis of the RDFs for the different interaction models already shows correlations between an observable splitting of the CO band and the structural elements of the surrounding ice. However, it does not allow us to relate the frequencies of individual CO molecules to the structure of their environment.

		Rigid water + multipoles			Flexible water + multipoles		
	temperature [K]	red[a] [cm^{-1}]	blue[b] [cm^{-1}]	rel. intensity[c] [%]	red[a] [cm^{-1}]	blue[b] [cm^{-1}]	rel. intensity[c] [%]
setup III	20	−11	+8	63	−4	+11	75
(average over three trajectories)	60	−8	–	–	−4	–	–
setup IV.1	20	−6	+5	57	−13	+2	41
	60	−10	–	–	−8	–	–
setup IV.2	20	−13	+2	45	−10	+3	59
(average over three trajectories)	60	−14	–	–	−7	–	–

Table 3. Spectroscopic results obtained from MD simulations for the interaction models B and D with 60 CO molecules (for setups III and IV).

[a] Position of the red-shifted maximum with respect to free CO. [b] Position of the blue-shifted maximum with respect to free CO. [c] Relative absorption intensity of the blue-shift maximum with respect to the red-shift maximum.
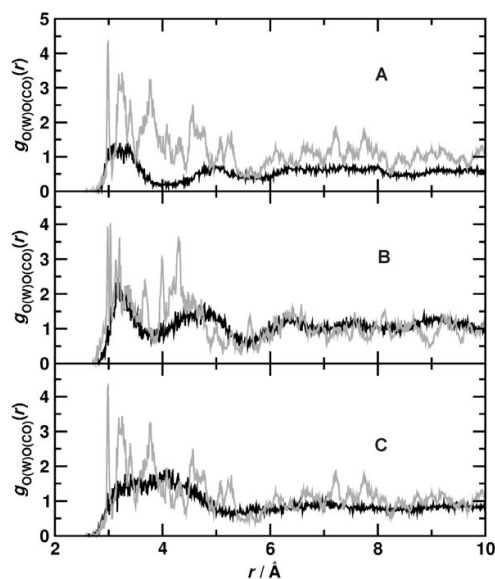
M. Meuwly and N. Plattner

Figure 4. RDFs of water-oxygen atoms around CO-oxygen atoms (between 2 and 10 Å). Panels (A) and (B) show comparisons of different water models for setup IV.2 at 20 K. A) Model A (black), model B (grey). B) Model C (black), model D (grey). C) Comparison of RDFs for the most accurate model (i.e. model D) using setup IV.1 at 20 (grey) and 60 K (black).

Frequency splitting and RDFs indicate the existence of distinguishable local structures around different CO molecules.

To investigate the relationship between splitting of the CO absorption band and the local environment of the CO molecules in more detail, we evaluate the infrared spectrum of each CO molecule individually and correlate it to the average number of hydrogen atoms within a distance ε from the particular CO center of mass. Figure 5 shows an analysis carried out for model B and setup III at 20 K and $\varepsilon = 3.0$ Å. This distance corresponds to an interstitial position in a hexagonal ice structure. For cutoff values between 2.5 and 3.5 Å, the results remain qualitatively the same. However, if ε is too small, the analysis looses its meaning because there are too few nearby hydrogen atoms. The same applies for much larger ε values for which the number of hydrogen atoms becomes essentially constant. As shown in Figure 5, a positive correlation is found between the number of hydrogen atoms and the absorption frequency. Spectra of CO molecules with a larger number of nearby hydrogen atoms are shifted to the blue.

3. Conclusions

We have shown that the experimentally observed splitting of the CO absorption band into a blue- and a red-shifted part can be rationalized based on MD simulations. The local ice structure around the CO molecules is influenced by the representation and accuracy of the electrostatic interactions, in particular by the inclusion of higher multipole moments; for example, the differences between water–water RDFs obtained from sim-
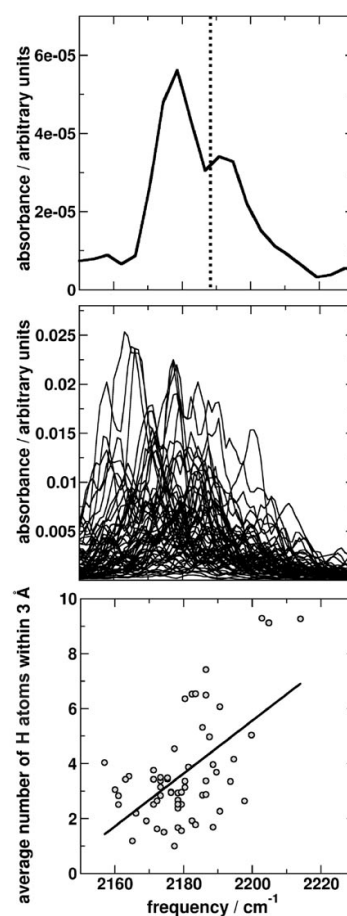

Figure 5. Correlation between the CO frequencies and the number of hydrogen atoms within 3 Å. The evaluation is carried out for model B. A) Spectrum of CO in ice after 20 ps equilibration, evaluated over 100 ps at 20 K. B) Superimposed power spectra for all the 60 CO molecules. C) Correlation between the CO frequencies and the average number of hydrogen atoms within $\varepsilon = 3$ Å (the solid line is a linear regression). The correlation persists even if the three points at the largest blue shifts are excluded from the linear regression.

ulations with different electrostatics (point charges versus DMA) are comparable to those obtained with different bond potentials. These results are different from those obtained for water–CO RDFs, where the influence of electrostatics is dominant. Splitting of the absorption band is related to both the residual structure of the surrounding ice and the positioning of the CO molecule relative to the ice lattice. Electrostatics that include higher multipole moments on the CO molecule, together with a spectroscopically accurate RRKR potential, were recently shown to quantitatively explain the splittings, shifts, and intensity ratios of photodissociated CO in myoglobin.[13] It is important to stress that this same CO model was used in the present work (without any changes). Thus, the model for CO proposed earlier appears to be applicable to the study of the

64

energetic and spectroscopic properties of carbon monoxide in different and heterogeneous environments.

The current simulations suggest that the proposal of substitutional and interstitial CO molecules as being responsible for the splitting of the CO absorption band has to be replaced. We found that the main spectroscopic features were driven by gradual changes in the local ice environment around each CO molecule and the formation of defined local structures. According to the original proposal, the spectrum in Figure 1 B should consist of two main absorption bands of comparable intensity corresponding to the five interstitial and five structural CO molecules, which is not the case. Furthermore, the analysis of the water-oxygen RDFs around the CO molecules [i.e. $g_{O(W)C(CO)}(r)$ and $g_{O(W)O(CO)}(r)$, see Figure 4] shows that a split CO spectrum can result from a distribution of geometries, which cannot be simply attributed to two categories (substitutional and interstitial). Also, the analysis presented in Figure 5 shows gradual changes in the local structure (instead of two clearly distinguishable CO positions in the ice). These observations suggest that the picture of structural and interstitial water molecules cannot fully explain the observed spectrum.

Carefully parametrized force fields—with physically meaningful multipole moments for the electrostatic interactions—provide quantitatively correct information and novel insight from atomistic simulations. This is particularly true for the correlation between infrared absorptions and structural features of the system. The present analysis of the absorption spectrum of CO in ice is a starting point for a better understanding of the infrared spectra of doped ices[9,28,29] and other mixtures, such as water with methane, CO_2, and methanol, which are known to be abundant in this type of ice structures.[28,30–32]

Acknowledgements

Keywords: astrochemistry · distributed multipole analysis · IR spectroscopy · interstellar ice · molecular dynamics

[1] K. M. Merrill, R. W. Russel, B. T. Soifer, *Astrophys. J.* **1976**, *207*, 763–769.
[2] B. T. Soifer, R. C. Puetter, R. W. Russel, S. P. Willner, P. M. Harvey, F. C. Gillett, *Astrophys. J.* **1979**, *232*, L53–L57.
[3] S. A. Sandford, L. J. Allamandola, *Icarus* **1988**, *76*, 201–224.
[4] A. Al-Halabi, H. J. Fraser, G. J. Kroes, E. F. van Dishoeck, *Astron. Astrophys.* **2004**, *422*, 777–791.
[5] H. J. Fraser, M. P. Colling, J. W. Dever, M. R. S. McCoustra, *Mon. Not. R. Astron. Soc.* **2004**, *353*, 59–68.
[6] M. E. Palumbo, G. Strazzulla, *Astron. Astrophys.* **1993**, *269*, 568–580.
[7] P. A. Gerakines, W. A. Schutte, J. M. Greenberg, E. F. van Dishoeck, *Astron. Astrophys.* **1995**, *296*, 810–818.
[8] M. P. Collings, J. W. Dever, H. J. Fraser, M. R. S. McCoustra, D. Williams, *Astrophys. J.* **2003**, *583*, 1058–1062.
[9] B. Schmitt, J. M. Greenberg, R. J. A. Grim, *Astrophys. J.* **1989**, *340*, L33–L36.
[10] P. Jenniskens, D. F. Blake, *Science* **1994**, *265*, 753–756.
[11] D. R. Nutt, M. Meuwly, *Biophys. J.* **2003**, *85*, 3612–3623.
[12] M. Meuwly, *ChemPhysChem* **2006**, *7*, 2061–2063.
[13] N. Plattner, M. Meuwly, *Biophys. J.* **2008**, *94*, 2505–2515.
[14] A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. E. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, *J. Phys. Chem. B* **1998**, *102*, 3586.
[15] J. N. Huffaker, *J. Chem. Phys.* **1976**, *64*, 3175–3181.
[16] J. N. Huffaker, *J. Chem. Phys.* **1976**, *64*, 4564–4570.
[17] A. J. Stone, *The Theory of Intermolecular Forces*, Clarendon, Oxford, **1996**.
[18] A. J. Stone, *J. Chem. Theory Comput.* **2005**, *1*, 1128–1132.
[19] W. L. Jorgensen, J. D. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926.
[20] N. Kumagai, K. Kawamura, T. Yokokawa, *Mol. Simul.* **1994**, *12*, 177–186.
[21] C. J. Burnham, J. C. Li, M. Leslie, *J. Phys. Chem. B* **1997**, *101*, 6192–6195.
[22] W. van Gunsteren, H. Berendsen, *Mol. Phys.* **1977**, *34*, 1311.
[23] Y. Wu, H. Tepper, G. Voth, *J. Chem. Phys.* **2006**, *124*, 024503.
[24] M. P. Allen, D. J. Tildesley, *Computer Simulation of Liquids*, Clarendon, Oxford, **1987**.
[25] G. E. Ewing, *J. Chem. Phys.* **1962**, *37*, 2250–2256.
[26] P. Berens, K. Wilson, *J. Chem. Phys.* **1981**, *74*, 4872–4882.
[27] A. K. Soper, *Chem. Phys.* **2000**, *258*, 121–137.
[28] J. Bouwman, W. Ludwig, Z. Awad, K. I. Oeberg, G. W. Fuchs, E. F. van Dishoeck, H. Linnartz, *Astron. Astrophys.* **2007**, *476*, 995–1003.
[29] A. G. G. M. Tielens, A. T. Tokunaga, T. R. Geballe, F. Baas, *Astrophys. J.* **1991**, *381*, 181–199.
[30] P. A. Gerakines, J. J. Bray, A. Davis, C. R. Richey, *Astrophys. J.* **2005**, *620*, 1140–1150.
[31] M. Bernstein, D. Cruikshank, S. Sandford, *Icarus*, **2005**, *179*, 527-534.
[32] K. I. Oeberg, H. J. Fraser, A. C. A. Boogert, S. E. Bisschop, G. W. Fuchs, E. F. van Dishoeck, H. Linnartz, *Astron. Astrophys.*, **2007**, *462*, 1187-1198.
[33] G. Maroulis, *J. Phys. Chem.* **1996**, *100*, 13466–13473.
[34] G. Maroulis, *Chem. Phys. Lett.* **2001**, *334*, 214–219.

4.3 MD simulations using distributed multipole electrostatics: Structural and spectroscopic properties of CO- and methane-containing clathrates

N. Plattner, T. Bandi, J.D. Doll, D.L. Freeman and M. Meuwly

Taylor & Francis
Taylor & Francis Group

INVITED ARTICLE

MD simulations using distributed multipole electrostatics: Structural and spectroscopic properties of CO- and methane-containing clathrates

Nuria Plattner[a], Tobias Bandi[a], J.D. Doll[b], David L. Freeman[c] and Markus Meuwly[a]*

[a]*Department of Chemistry, University of Basel, CH-4056 Basel, Switzerland;* [b]*Chemistry Department, Brown University, Providence, RI, USA;* [c]*Chemistry Department, University of Rhode Island, Kingston, RI, USA*

Molecular dynamics (MD) simulations are used to investigate the properties of empty and methane- and carbon-monoxide-containing hydrates. Intermolecular interactions are described by force fields including a combination of anharmonic bond potentials and accurate molecular electrostatics based on distributed multipoles. It is found that structural and spectroscopic properties and small differences between the systems can be correctly described with such an approach. In particular, spectra in the acoustic region (low-frequency lattice modes) delicately depend on the electrostatic interactions between the molecules for clathrates with different guests. Analysis of the radial distribution functions demonstrates that the short-range order is dominated by steric effects, whereas the long-range order is more sensitive to electrostatic interactions. Predictions are made for the acoustic modes of CO-containing clathrates and for the structural differences in the hydrogen radial distribution function between the CO- and methane-containing clathrates.

Keywords: intermolecular forces; distributed multipoles; clathrates; atomistic simulations

1. Introduction

Clathrate hydrates are formed by water molecules which arrange in an ice-like structure around guest molecules or atoms. A large variety of clathrates with different guest molecules and atoms have already been observed in nature or have been synthesized [1]. Methane hydrates are of special interest, because they have been detected in large amounts on the sea floor [2]. These natural methane storages are important on one hand as a potential energy source, on the other hand as a possible environmental hazard in the case of decomposition of the hydrates. Technologically, CO hydrates have been studied in the context of separation of industrial flue gases [3]. Furthermore, they are of interest for the modeling of planetary atmospheres, since CO is a very abundant molecule in space. On the Saturn moon Titan, decomposition of CO and methane clathrates has been suggested as the source of observed methane and CO emissions [4,5].

The structures of clathrate hydrates can be divided into three groups according to the arrangement of small and large water cages. These groups are called type I, type II and hexagonal structure type [1]. The structures of both methane and CO hydrates have been characterized by X-ray techniques and have been

shown to form preferentially a type I structure [6,7]. For CO, the formation of structure type I is rather unexpected, since similar diatomic molecules such as N_2 and O_2 are known to form structure type II, and thermodynamic models have made contradictory predictions, favouring in one case structure II [8], and in another case structure I [9]. The crystalline structure of clathrates suggests that the physical properties of this material are similar to those of ice, but it was found that, in contrast to ice, the thermal conductivity is very low and the temperature dependence is more comparable to amorphous material than to crystal [10]. The thermal energy in solids is transported via the lattice vibrations, which are characterized in clathrates by strong couplings between the guest and the host lattice [11].

To investigate in more detail the influence of the guest molecule on the host lattice, methane and CO were chosen in this study for comparison. Particular emphasis is placed on spectroscopic and structural properties as the thermodynamics have already been investigated [12–14]. Due to their different molecular properties and the unexpected finding of CO favouring the same structure as methane, the influence of these two guest molecules on the host lattice can be expected to lead to significant changes in the lattice behaviour.

*Corresponding author. Email: m.meuwly@unibas.ch

67

N. Plattner et al.

Based on the validation of methane, predictions for the CO hydrate system are possible. Comparative experimental and computational studies of different guest molecules in clathrates of the same structure type have already been performed for methane and xenon [15] and for methane and empty clathrates [12].

The importance of anharmonic bond potentials and accurate electrostatics for atomistic simulations of clathrate hydrates has been demonstrated in previous work [13,14,16]. In the present study, Distributed Multipole Analysis [17] is employed to analyse and derive suitable descriptions of the molecular electrostatics from them. They are then further used in atomistic simulations of clathrates under realistic experimental conditions and structural and spectroscopic signatures are derived from them. The work is structured as follows. First, the methods and simulation techniques are discussed. Then, results are presented and discussed in the final section.

2. Computational methods

All simulations were carried out with the CHARMM program [18]. Interaction terms for atomic multipole moments up to rank two for all molecular symmetries and up to rank three for linear molecules were implemented additionally based on the published terms [19]. To describe the orientations of the atomic multipole moments with respect to the global Cartesian coordinate system, reference axis systems are assigned to each molecule and the interactions are transformed in each step into the global axis system.

For methane, CO and water, distributed multipole moments were obtained from GDMA [20] based on electron densities calculated with GAUSSIAN [21]. For water and methane, multipoles based on SCF, DFT and MP2 wave functions with different basis sets (6-31G* and aug-cc-pVTZ) were calculated. As a test, the difference between the electrostatic potentials outside the Van-der-Waals radius of the molecules based on the atomic multipole moments and the *ab initio* electrostatic potential based on the SCF and MP2 density, respectively, were evaluated. The quality of the fields was assessed by evaluating the electrostatic potential on a $101 \times 101 \times 101$ grid in the x-, y- and z-directions within a cube of dimensions $0.1 \times 0.1 \times 0.1$ nm. For water, multipoles from different methods lead to comparable fields and those used in the simulations (from B3LYP/aug-cc-pVTZ, see Table 1) fit the electrostatic fields to within 3% (SCF/6-31G*) and 18% (MP2/aug-cc-pVTZ), respectively. It is worth pointing out that multipoles from different methods provide electrostatic fields that agree to within a few percent. Furthermore, it was found that multipole moments from DFT methods for CO converge reliably with basis set size. For methane, the differences between the fields from the wave function and those from the multipole moments are larger [22]. Multipole moments from MP2/aug-cc-pVTZ calculations (see Table 2) reproduce the

Table 1. Electrostatic moments for the water multipole model. The values correspond to spherical tensor notation.

	Charge (e)	Dipole (ea_0)		Quadrupole (ea_0^2)	
		$q10$	$q11s$	$q20$	$q22c$
Oxygen	−0.700000	−0.131000		−0.030000	−0.800000
Hydrogen	0.350000	0.010000	−0.010000[a]		

Note: [a]Opposite signs on the two hydrogens.

Table 2. Electrostatic moments for the methane multipole model. The values correspond to spherical tensor notation. Signs and directions differ for different hydrogens and according to the definition of the molecular reference axis system. The values shown here correspond to a carbon at the origin and a hydrogen in the negative x- and z- and the positive y-direction; α corresponds to x, β to y, γ to z.

	Charge (e)	Dipole (ea_0)		Quadrupole (ea_0^2)			
		$q1_\gamma$	$q1_\alpha$	$q2_{\alpha\gamma}$	$q2_{\beta\gamma}$	$q2_{\alpha\beta}$	$q2_{\alpha^2-\beta^2}$
Carbon	−0.360000						
Hydrogen	0.090000	−0.061124	−0.086376	0.222918	−0.011851	−0.030930	0.074024

Note: Due to the different molecular symmetries, additional indices are required for methane; for example, $q10$ from water corresponds to $q1_\gamma$ in methane.

electrostatic field derived directly from the wavefunction to within 30%, whereas for SCF/6-31G* the difference is 70%. The multipole parameters for CO from previous work [22] are given in Table 3.

For the bond and angle parameters of the water molecule, the flexible KKY (Kumagai, Kawamura and Yokokawa) potential [23] was used. As previously reported, the KKY potential parameters have to be re-parametrized to represent the water stretching and bending frequencies [24]. The KKY potential has already been successfully applied to methane clathrates [16].

The simulations for the empty, CO- and methane-containing clathrate structures were carried out in the *NVT* ensemble with periodic boundary conditions in a simulation box with dimensions 2.406 nm in each dimension. This corresponds to $2 \times 2 \times 2$ clathrate unit cells of structure type I. For CO-containing clathrates there are no structural data on the occupancy of the cages, but since the composition is similar to methane clathrates [7], it has been assumed that each cage is occupied by one CO molecule. In methane hydrate, each cage was found to be occupied by one CH_4 molecule [15]. A timestep of 0.4 fs was chosen to follow the flexible hydrogen motion. For validation, additional simulations for methane hydrate were carried out and analysed: one employed the *NpT* ensemble with $p = 20$ bar, and the other used a larger system, containing $3 \times 3 \times 3$ (*NVT* ensemble) unit cell. An example of a unit cell used in the simulations is shown in Figure 1. For both validation simulations, the spectra (not shown) agree with those from the setup used for the production runs.

Infrared spectra can be calculated from the Fourier transform of the dipole moment autocorrelation function $C(t)$ [25]. The latter is defined as $C(t) = \langle \boldsymbol{M}(0)\boldsymbol{M}(t)\rangle/\langle \boldsymbol{M}(0)\boldsymbol{M}(0)\rangle$, where $\boldsymbol{M}(t)$ is the total dipole moment of the system at time t. $C(t)$ is accumulated over 2^n time origins, where n is an integer such that 2^n corresponds to between 1/3 and 1/2 of the trajectory, with the time origins separated by 0.8 fs. $C(t)$ is then transformed to yield $C(\omega)$ using a Fast Fourier Transform with a Blackman filter to minimize noise [26]. The final spectrum $A(\omega)$ is then calculated from

$$A(\omega) = \omega\{1 - \exp[-\hbar\omega/(kT)]\}C(\omega), \qquad (1)$$

where k is the Boltzmann constant and T is the temperature in Kelvin.

The acoustic modes in the low-frequency region give clearly distinguishable but rather weak peaks. Spectra from the dipole moment autocorrelation function provide information on any motion in a system that changes the dipole moment in the *x*-, *y*- or *z*-direction. For translations, such motion is either associated with (1) the translation of the oxygen (or hydrogen) atoms which are not pure translations of an entire water molecule, and (2) translation of one or a few water molecules relative to the other water molecules in the lattice. It is observed that the translational region of the water IR spectrum consists of a broad background on which distinct lines are superimposed. In clathrates, the three translational bands occur due to avoided crossings between the host lattice and the guest lattice motion [11]. A correlated motion of the waters in the host lattice should lead to a change of the dipole moment. The infrared spectrum is dominated by the lattice vibrations. Alternatively, spectral features can also be calculated
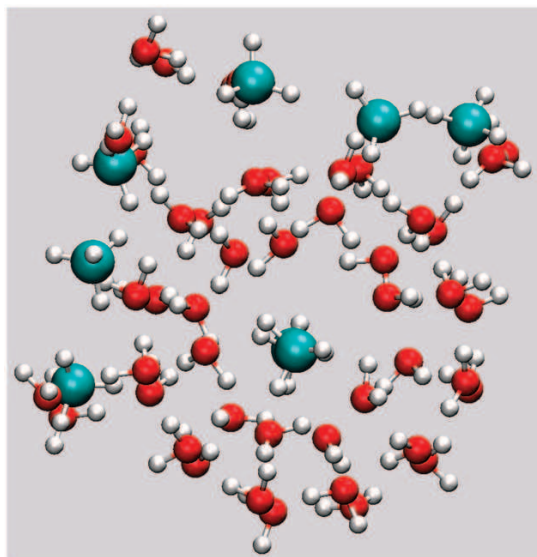

Figure 1. Methane clathrate unit cell; snapshot after 100 ps of dynamics at 100 K in a system of $2 \times 2 \times 2$ unit cells with periodic boundary conditions.

Table 3. Distributed multipole parameters for CO models. The values refer to the equilibrium CO separation (112.8 pm).

	Charge (e)	Dipole (ea_0)	Quadrupole (ea_0^2)	Octopole (ea_0^3)
Carbon	0.563418	1.096170	−0.050725	
Oxygen	−0.563418	0.134258	0.467227	−1.774760

from power spectra of interatomic distances or from the Fourier transform of the velocity autocorrelation function, which provides the density of states. The power spectra can be compared with earlier studies of methane hydrate, where other force fields were used [13,14]. It should be noted that only the spectra from the dipole-moment autocorrelation function provide intensity information. For example, the signal that appears from a distance–distance autocorrelation function may not occur in $I(\omega)$ calculated from the dipole-moment autocorrelation function for symmetry reasons [27,28].

3. Results and discussion

3.1. *Validation for CH₄–D₂O and CD₄–H₂O*

To validate the computational approach, low-frequency (INS) spectra are calculated and compared with the experimental frequencies for CD_4–H_2O and CH_4–D_2O at 150 K. The experimental INS frequencies for CD_4–H_2O are found at 42.7, 58.9 and 87.9 cm^{-1}, whereas for CH_4–D_2O they are located at 43.5, 61.3 and 80.6 cm^{-1} [29]. For both systems, averaged spectra were calculated based on five independent, 100 ps trajectories.

To calculate the averaged power spectra corresponding to the O–O vibrations, the distance-autocorrelation function was evaluated for 33 uniformly distributed and randomly selected oxygen–oxygen distances r_{OO}. The number of 33 pairs is arbitrary and tests were carried out with fewer O–O pairs and different cutoff distances. It was found that the main features in the INS spectra are captured if five or more O–O pairs are used and for cutoff distances $r_{OO} \geq 1$ nm (see Figure 2). To analyse the spectra, averages over a subset of the longest distances were additionally evaluated. The criterion to include a distance was $r_{OO} \geq 2$ nm. Such spectra indicate correlated motions in the entire host lattice. To analyse and compare the power spectra with the IR spectrum, a sum of three Gaussian functions was fitted to the signal between 20 and 100 cm^{-1}. The results are shown in Figure 2. From the fit of the Gaussian functions, three center frequencies are found at 40, 62 and 88 cm^{-1} for CD_4–H_2O and at 45, 67 and 85 cm^{-1} for CH_4–D_2O. This analysis shows that the trends in frequency shifts are reproduced by the calculated spectra, with the lower two frequencies of the CH_4–D_2O shifted to the blue with respect to CD_4–H_2O (by 5 cm^{-1} from the simulations compared with 0.8 and 2.4 cm^{-1} from

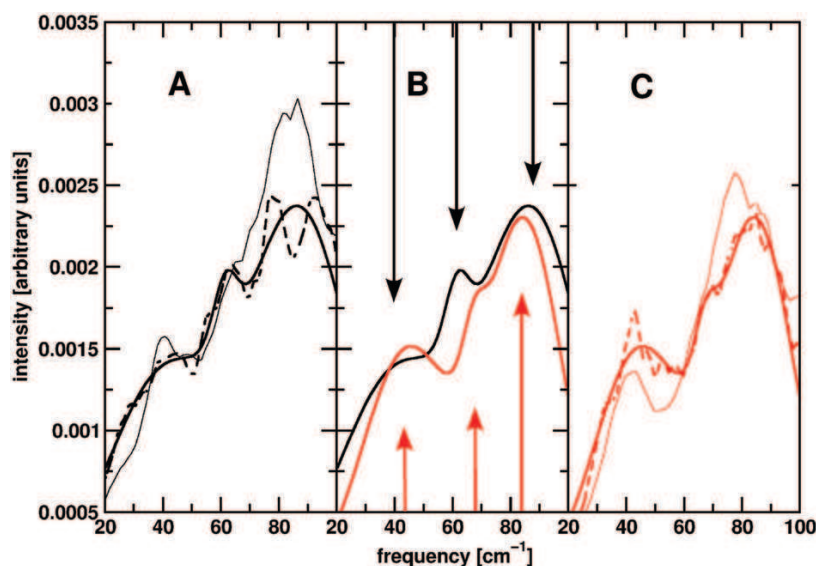Figure 2. INS spectra in the translational region for CD₄–H₂O (black lines) and CH₄–D₂O (red lines). Averages over power spectra from 33 (thin lines) and five (dashed lines) oxygen–oxygen distance time series for OO distances larger than 2 nm. The set of 33 distances contains the entire range of oxygen–oxygen distances between 0.3 and 2.7 nm, whereas the set of five is the subset containing only the five longest distances. The calculated INS spectra are fits of a sum of three Gaussian functions to the power spectra averaged over five runs. (A) Analysis for CD₄–H₂O, (C) analysis for CH₄–D₂O and (B) comparison of the INS spectra of CD₄–H₂O and CH₄–D₂O to clarify the shifts of the bands. The centers of the Gaussian functions are indicated with arrows in black for CD₄–H₂O and red for CH₄–D₂O.

experiment, respectively) and the highest frequency shifted to the red (by $3\,cm^{-1}$ from the simulations and by $7.3\,cm^{-1}$ from experiment, respectively). Thus small spectral changes of the order of a few wavenumbers can be reproduced using the methods described above. To corroborate these results, averaged power spectra from the velocity-velocity (v, v)-autocorrelation functions of all water oxygens were analysed for the two systems. The results are shown in Figure 3. The relative positions of the three center frequencies compare well with those from the distance correlation functions, with deviations between 0 and $4\,cm^{-1}$. From the fit of three Gaussian functions, they are found at 42, 61 and $87\,cm^{-1}$ for CD_4–H_2O and at 46, 63 and $85\,cm^{-1}$ for CH_4–D_2O. Thus, IR spectra and spectral signatures from (v, v)-autocorrelation functions give results in close agreement with each other.

To assess the overall structural features of the two systems, the radial distribution functions (rdfs) were evaluated (Figure 4). The comparison shows that the two rdfs are virtually identical except for small differences ($\leq 2\,pm$) in the position of the first maximum of the g_{OO} and g_{OH} rdfs and the fact that the width of the first peak at short range for the CD_4–H_2O system is narrower and more pronounced. The integral over the first peak of g_{OO} is, however,
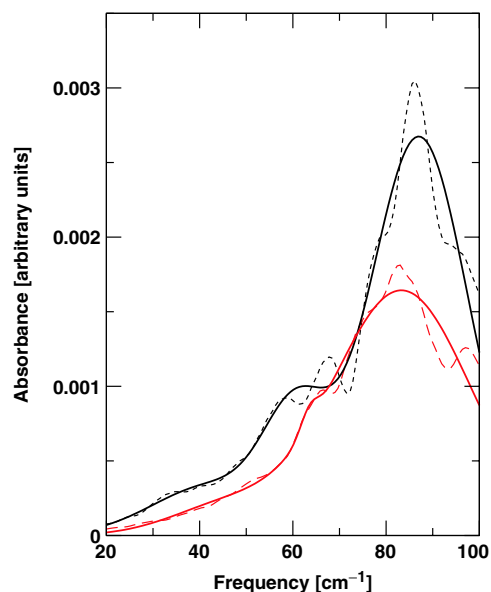
identical for the two systems. Higher peaks have previously been attributed to enhanced abilities to form hydrogen bonds [13]. The position of the first peak is at 0.277 nm, which agrees well with previous work, as does the asymmetry in the first peak [12]. From the definition of a rdf it is expected that different isotopomers have identical radial distribution functions, because, for an *NVT* ensemble, rdfs are strictly mass-independent. Thus, additional tests were carried out. For longer simulations (500 ps instead of 100 ps) for CD_4–H_2O, g_{OO} converges towards g_{OO} from simulations on CH_4–D_2O (see inset of Figure 4). For the second peak, convergence is essentially complete, whereas for the first peak, convergence is slow. This suggests that 500 ps of simulation for the present setup (flexible molecules with multipolar electrostatics) may not be sufficient to stabilize the water g_{OO} rdf. One reason for the slow equilibration may be mode locking [13]. The maxima of the present g_{OO} rdfs are 8.6
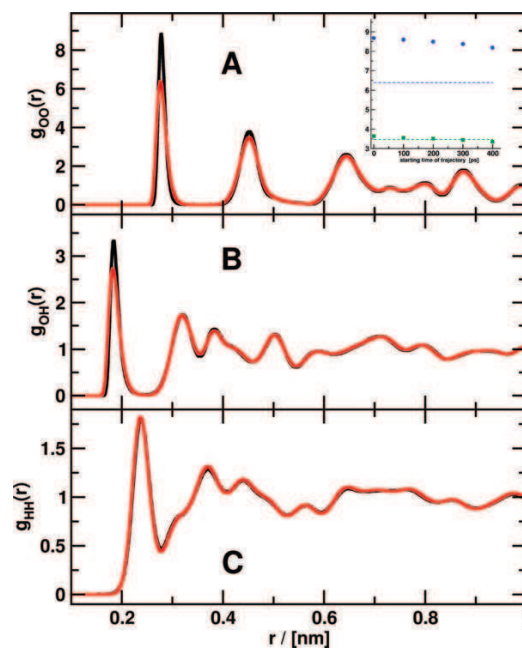
Figure 3. Translational spectra calculated from (v, v)-autocorrelation functions for CD_4–H_2O (black lines) and CH_4–D_2O (red lines). Shown are averages over five independent trajectories. Dashed lines are the raw data and solid lines are sums over three Gaussian functions fitted to them.

Figure 4. Radial distribution functions for the CD_4–H_2O (black) and CH_4–D_2O (red) clathrates. The evaluations are averages over trajectories at 150 K. The distributions are from water molecules only. Oxygen–oxygen distributions are shown in panel (A), oxygen–hydrogen in (B) and hydrogen–hydrogen in (C). The inset in panel (A) shows the average height of the first and second peak for g_{OO} from 500 ps simulations on CD_4–H_2O at different times. Dashed lines are the height of g_{OO} from simulations on CH_4–D_2O. As can be seen, the second peak is converged by ≈ 300 ps, whereas the first peak is not converged by 500 ps.

and 6.4 (at 150 K), respectively, compared with values of between 6.1 and 7 (at 200 K) [13], ≈7 [12] (at 145 K), and 6.5 to 7.2 (at 200 K) [14]. The temperature is known to influence the maximum peak height by up to 25% in the temperature range from 265 K to 348 K (liquid water) [30]. Thus, peak heights between 7.0 and 5.5 are expected at 200 K, which was also found from test simulations (*NVT* ensemble). However, additional studies on this are required to arrive at a more definitive answer. To the best of our knowledge, detailed studies of equilibration times in different ensembles ((*NVE*), (*NVT*) or (*NpT*)) have not thus far been carried out involving flexible water molecules.

3.2. *Comparison of methane hydrate with CO and empty hydrates*

Experimental spectral data for methane-containing hydrates are available from inelastic neutron scattering (INS) measurements in the low-frequency region [29] and from Raman spectra which report on the high-frequency region [31]. For methane, CO and empty hydrates, five trajectories of 100 ps each were run at 150 K. The pressures in the *NVT* simulations for the three systems were evaluated at 150 K and were found to be −0.8 kbar for methane hydrate, −0.7 kbar for empty hydrates and −1.6 kbar for CO hydrate. Such pressures for empty and methane-containing hydrates agree with previous work on the two systems and

do not lead to mechanical instabilities as pressures are considerably lower than the bulk modulus of ice (≈100 kbar) [12]. In more recent work the pressures in methane hydrates were calculated for a range of water models and temperatures [14]. It was found that, at 125 K, the AMOEBA and COS/G2 force field gave pressures of −1.8 kbar and −2.4 kbar, respectively. For 200 K, the pressures approach 0 kbar. For methane and CO hydrates, five additional runs were carried out at 200 K for comparison and all systems were found to be stable. Finally, one trajectory for methane hydrate was extended to 500 ps for convergence tests and two *NpT* simulations were carried out and analysed for $p = 20$ bar.

In Figure 5, the translational spectra of the different clathrate systems are compared based on IR and power spectra. For empty clathrates, no distinct peaks are found in the low-frequency region. The methane hydrate shows frequencies similar to CD_4–H_2O, namely at 42, 62 and 90 cm^{-1}. For CO hydrate, the lattice frequencies are red-shifted compared with the methane-containing clathrates and are found at 40, 61 and 76 cm^{-1} based on the analysis of the oxygen–oxygen distances, and at 44, 63 and 80 cm^{-1} based on the velocity autocorrelation functions (see Figure 6). The IR intensity is largest for CO hydrates.

It is instructive to compare the power spectrum and the infrared spectrum from the dipole moment autocorrelation function. For the CO-containing
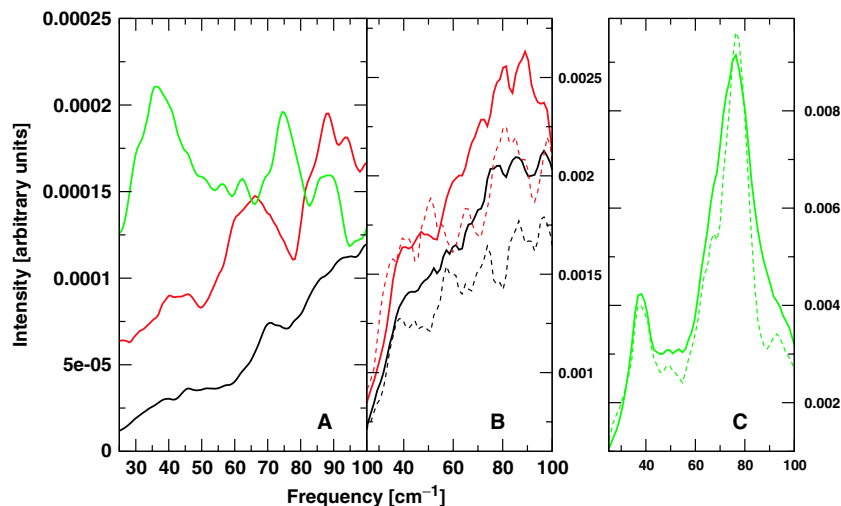

Figure 5. Comparison of IR and power spectra in the translational region for different clathrate systems: empty hydrate (black), methane hydrate (red), CO hydrate (green). The IR spectra are shown in panel (A), the power spectra for methane and empty hydrates in (B), and the CO hydrate power spectra in (C). The solid lines in the power spectra show the average over all 33 oxygen–oxygen distances, the dashed line averages of distances longer than 2 nm. A separate panel is chosen for CO due to the greater intensity.

clathrate (Figures 5(A) and (C)), both maxima in the power spectrum at 40 and 76 cm^{-1} have counterparts in the IR spectrum (at 39 and 80 cm^{-1}). This is in agreement with studies on other systems (D$_2$O [32], H$_5$O$_2^+$ [28] or N$_2$H$_7^+$ [27,33]), where it was shown that power spectra can assist in assigning infrared-active bands to atomic motions. However, no intensity information can be obtained from the power spectra. This is more clearly visible if the infrared and power spectra for methane-containing clathrates are considered. There, the spectral feature at 65 cm^{-1} in Figure 5(A) is not visible in the power spectrum. It should, however, be remembered that the power spectrum is only evaluated over internuclear distances, whereas the infrared spectrum is also sensitive to angular motions of all molecules involved. Comparing the infrared and power spectra for empty clathrates with the methane-containing clathrate, it is demonstrated that little information concerning intensities can be extracted from the power spectra.

In other regions of the water IR spectrum the methane hydrate shows a clear difference to the other two systems in the intensity distribution. Infrared lines for the O–H stretch (3200 to 3700 cm^{-1}) and bending vibrations (1600 to 1700 cm^{-1}) are more intense,

whereas in the lattice vibrational region (600 cm^{-1}) the intensity is reduced (see Figure 7). The splitting of the OH-stretch frequency region into an absorption due to OH motion at about 3600 cm^{-1} and a coupled O–H region at lower frequencies is also observed in experimental spectra with a splitting of 500 cm^{-1} [31]. Quantitatively, the observed splitting of 300 cm^{-1} is too small, which is most likely related to the water model used. This is in line with investigations of ice Ih where the shift of the coupled OH-stretch frequency was found difficult to quantitatively reproduce with classical force fields [34]. For different water models, frequency shifts ranging from 100 to 600 cm^{-1} between liquid water and ice Ih have been observed. An important reason for the difficulties in capturing the experimental redshift are quantum effects. It has been shown that, for example for the TIP3FP model, the mean O–H stretch frequency shifts from 3490 cm^{-1} to 3354 cm^{-1}, if path-integral methods are used instead of classical methods [34]. More recently, quantum, classical and ring-polymer simulations of infrared spectra of H$^+$(H$_2$O)$_2$ were presented [35]. This work showed, in particular, that the spectra in the O–H stretching region from fully quantum simulations are very similar to those from classical MD simulations and ring-polymer MD. In particular, all these approaches find a split O–H spectrum between 3500 and 4000 cm^{-1}.

To characterize better the structural features of the clathrates (empty, CO- and CH$_4$-containing), different radial distribution functions for the water molecules were calculated (see Figure 8). The overall structure of the clathrate is defined by the location of the maxima in the g_{OO} rdf (not shown), which is nearly identical for all systems investigated. Thus, the overall clathrate structure as defined by the positions of the water oxygen atoms is less affected by the presence or absence of different guests. Larger differences are found for rdfs involving hydrogen atoms, in particular g_{HH}. The hydrogen–hydrogen distribution changes depending upon the presence or absence and the identity of the guest molecule. The most pronounced difference is observed for the hydrogen distribution when methane is present, whereas it is little affected by the presence of the CO molecules. Considering the cumulative sum of $g_{HH}(r)$ it is found that, for the methane-containing clathrate, the rdf differs from the CO-containing and the empty clathrate between 0.25 and 0.3 nm. At 0.25 nm, the cumulative sum is lower by about 20% compared with the other two clathrates. The missing density is shifted to the radial region at 0.3 nm, which can be explained by slightly different orientations in the hydrogen-bonded network. This is most likely caused by the larger volume of the methane
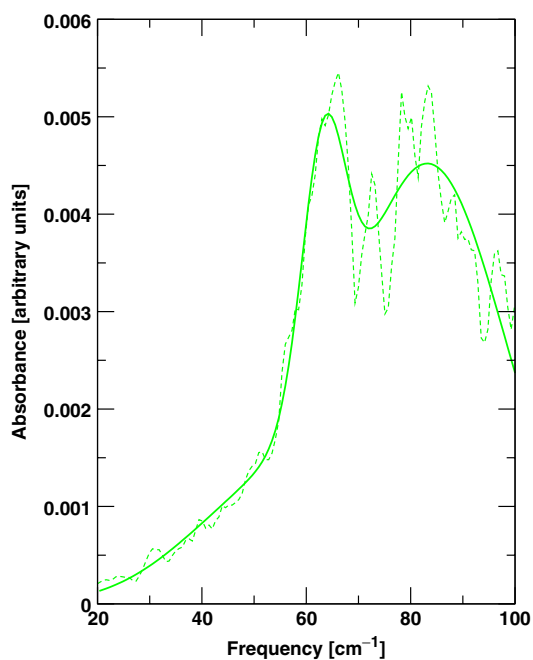

Figure 6. Translational spectra calculated from (v, v)-auto-correlation functions for CO hydrate averages over five trajectories. Dashed line: raw data; solid line: sum over three Gaussian functions fitted to the data.
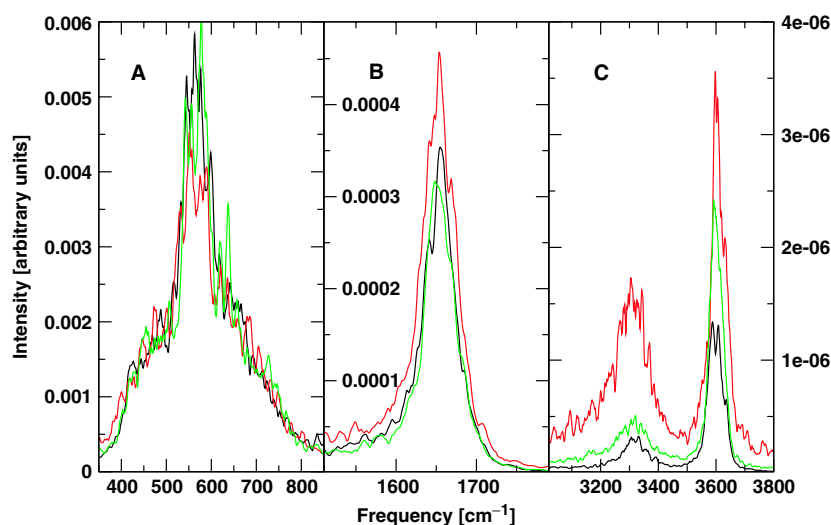
73

Figure 7. Intensities for different hydrates in the lattice vibrational (A), water bend (B) and OH-stretch (C) region. Black lines are for empty hydrates, red lines for methane hydrates, and green lines for CO hydrates.
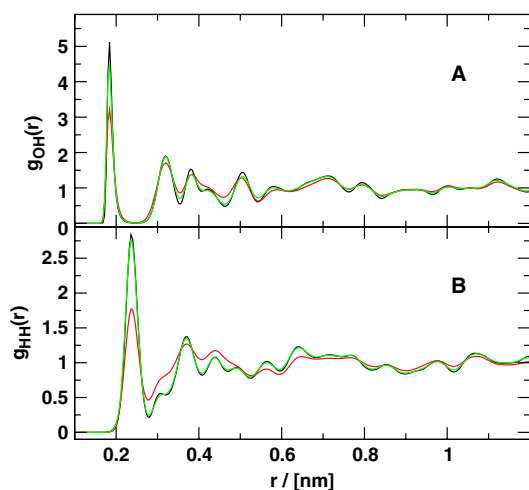

Figure 8. Radial distribution functions g_{OH} (A) and g_{HH} (B). Empty clathrates are shown in black, methane clathrates in red, CO clathrates in green. The results are averages over five trajectories at 150 K.

molecules and by the electrostatic repulsion between the water-H and methane-H atoms. Given that $g_{HH}(r)$ from simulations on CD_4–H_2O and CH_4–D_2O (see Figure 4(C)) is converged, this observation should not be affected by the slow convergence found for g_{OO}.

In addition to (NVT), simulations in the (NpT) ensemble were carried out for methane clathrates at 20 bar (with identical time step and other simulation

parameters as reported in the methods section). As mentioned above, simulations in the (NVT) ensemble lead to pressures of $p \approx -1$ kbar. In the (NpT) simulations the size of the periodic box changed from 2.406 nm to 2.401 nm, where it equilibrated. Rdfs from the two different ensembles (not shown) do not change the location and width of the distribution functions. It is interesting to note that the differences between rdfs from runs in the (NVT) and (NpT) ensembles for the present simulation conditions are considerably smaller than the differences obtained from simulations with different water models [13,14].

The guest-induced structural differences shown by spectra and radial distribution functions above were further analysed by changing the electrostatic models of the guest molecule. For methane and CO hydrate the influence of electrostatic interactions of the guest molecule with the environment was assessed by removing the charges on the guest molecules and only retaining Van-der-Waals interactions. For all system compositions (empty, CO- and methane-containing) and representations (with and without electrostatics) the IR spectra of the water network were calculated from the water dipole moment autocorrelation function and are compared in Figure 9. The largest change is found between systems with and without electrostatics on the guest molecule. Spectra without multipole moments on the guest molecules (dashed lines) are closely related to the INS of the empty clathrate (black): the spectra are largely unstructured. This clearly contrasts with simulations in which the
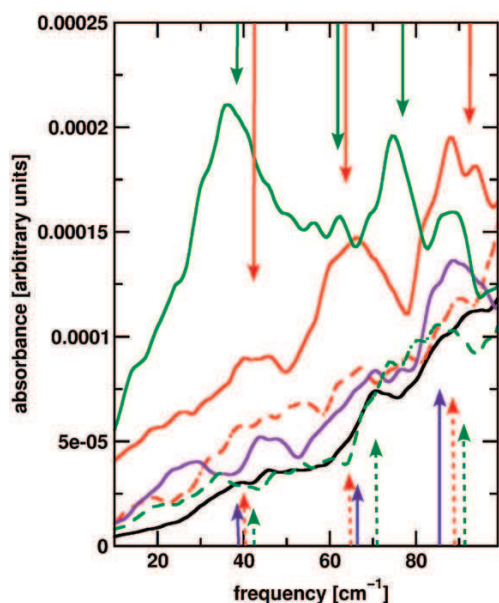
Figure 9. Comparison of IR spectra in the acoustic region of the spectrum for different system compositions and representations of the electrostatics. Colour coding: empty clathrate (black), CH_4 (red), CH_4 without electrostatics (red dashed), CD_4 (violet), CO (green), CO without electrostatics (green dashed). The arrows indicate the frequencies obtained from the corresponding power spectra for comparison (see text).

electrostatics of the guest molecules are described with multipoles up to atomic quadrupole moments (solid lines in Figure 9). Spectra from simulations using the CO model without electrostatics are blue-shifted with respect to CO with electrostatics. For methane, the change due to electrostatics is smaller and can only be observed for the middle frequency, which is blue-shifted.

4. Conclusions

In the present work, DMA-derived multipole moments have been used to investigate the dynamics and spectroscopy of CO- and CH_4-containing clathrates. A comparison of the experimental INS spectra for CH_4-D_2O and CD_4-H_2O clathrates with the simulated spectra shows that a force field based on a combination of anharmonic bond potentials and accurate molecular electrostatics (from distributed multipoles) is able to correctly describe differences in the spectroscopy of clathrate systems. This agrees with previous studies on CO in myoglobin [22] and CO in ices [36], where subtle details of the CO-infrared

spectroscopy (splittings and shifts) could be captured using electrostatics beyond the point-charge approximation.

Based on a comparison of the present results with experimental data from INS, the representation of the electrostatic interactions used here gives quantitatively correct results that are also comparable to those from the polarizable AMOEBA model [14]. However, it should be noted that the latter simulations were carried out at 200 K, whereas the experimental data [29] were recorded at 150 K, the temperature used in the present simulations.

The analysis of the spectra in the acoustic region for the different guest molecules shows an important influence of the molecular electrostatics on this feature, with modulations by the shape defined by the Van-der-Waals radius ('size and shape') of the guest. The structure of the clathrate defined by the oxygen distribution appears to be most influenced at short range by the shape of the guest and at medium and long distances by the electrostatics of the guest molecule.

References

[1] E.D. Sloan, *Clathrate Hydrate of Natural Gas* (Marcel Dekker, New York, 1998).
[2] K.A. Kvenvolden, Org. Geochem. **23**, 997 (1995).
[3] A. Mohammadi, R. Anderson, and B. Tohidi, AIChE J. **51**, 2825 (2005).
[4] G. Tobie, J. Lunine, and C. Sotin, Nature **440**, 61 (2006).
[5] K. Baines, P. Drossart, M. Lopez-Valverde, *et al.*, Planet. Space Sci. **54**, 1552 (2006).
[6] D. Davidson, Y. Handa, C. Ratcliffe, *et al.*, Nature **311**, 142 (1984).
[7] D. Davidson, M. Desando, S. Gough, *et al.*, Nature **328**, 418 (1987).
[8] S.L. Miller, in *Ices in the Solar System*, edited by J. Klinger, D. Benest, A. Dollfus, *et al.* (Kluwer Academic, Dordrecht, 1985).
[9] J. Lunine and D. Stevenson, Astrophys. J. Suppl. Ser. **51**, 2825 (2005).
[10] J. Tse and M. White, J. Phys. Chem. **92**, 5006 (1988).
[11] J. Tse, V. Shpakov, V. Murashov, *et al.*, J. Chem. Phys. **107**, 9271 (1997).

[12] J.S. Tse, M.L. Klein, and I.R. McDonald, J. Phys. Chem. **87**, 4198 (1983).

[13] N. English and J. Macelroy, J. Comp. Chem. **24**, 1569 (2003).

[14] H. Jiang, K. Jordan, and C. Taylor, J. Phys. Chem. B **111**, 6486 (2007).

[15] J. Baumert, C. Gutt, V. Shpakov, *et al.*, Phys. Rev. B **68**, 174301 (2003).

[16] H. Itoh and K. Kawamura, Ann. NY Acad. Sci. **912**, 693 (2000).

[17] A.J. Stone and M. Alderton, Mol. Phys. **56**, 1047 (1985).

[18] A.D. MacKerell Jr, D. Bashford, M. Bellott, *et al.*, J. Phys. Chem. B **102**, 3586 (1998).

[19] A.J. Stone, *The Theory of Intermolecular Forces* (Clarendon Press, Oxford, 1996).

[20] A.J. Stone, J. Chem. Theor. Comp. **1**, 1128 (2005).

[21] M.J. Frisch, G.W. Trucks, H.B. Schlegel, *et al.*, Gaussian 03, Revision B.01 (2004).

[22] N. Plattner and M. Meuwly, Biophys. J. **94**, 2505 (2008).

[23] N. Kumagai, K. Kawamura, and T. Yokokawa, Molec. Simul. **12**, 177 (1993).

[24] C. Burnham, J. Li, and M. Leslie, J. Phys. Chem. B **101**, 6192 (1997).

[25] D.A. McQuarrie, *Statistical Mechanics* (University Science Books, Mill Valley, CA, 2000).

[26] M.P. Allen and D.J. Tildesley, *Computer Simulation of Liquids* (Clarendon Press, Oxford, 1987).

[27] M. Meuwly and M. Karplus, J. Chem. Phys. **116**, 2572 (2002).

[28] S. Lammers and M. Meuwly, J. Phys. Chem. A **111**, 1638 (2007).

[29] J. Baumert, Ph. D. thesis, Christian-Albrechts University, 2003.

[30] P. Ren and J.W. Ponder, J. Phys. Chem. B **108**, 13427 (2008).

[31] C. Tulk, J. Ripmeester, and D. Klug, Ann. NY Acad. Sci. **912**, 859 (2000).

[32] M. Aida and M. Dupuis, J. Molec. Struct. **633**, 247 (2003).

[33] A. Fouqueau and M. Meuwly, J. Chem. Phys. **123**, 244308 (2005).

[34] C. Burnham, G. Reiter, J. Mayers, *et al.*, PCCP **8**, 3966 (2006).

[35] X.C. Huang, S. Habershon, and J.M. Bowman, Chem. Phys. Lett. **450**, 253 (2008).

[36] N. Plattner and M. Meuwly, Chem. Phys. Chem. **9**, 1271 (2008).

5 Conclusions

Distributed multipole moments provide a possibility to systematically improve the accuracy of electrostatic interactions in atomistic force fields. The DMA algorithm and other population analysis methods have been optimized in the past for convergence of the multipole expansions and accuracy of the resulting potential. For the parametrization of atomistic force fields, it is important to find an optimal tradeoff between the accuracy of the expansion and the computational cost, which is related to the number of multipole moments per atom and to the total number of multipole sites. Comparison of the multipole potential to the corresponding *ab initio* potential allows to optimize the number of multipole moments per molecule by testing different truncations of the multipole series for each atom.

Distributed multipole expansions up to rank two are in most cases sufficiently accurate. The difference of the potential at this level to *ab initio* is within a few percent, which is significantly more accurate than point charge potentials. Even lower differences to *ab initio* would probably not be meaningful, since most of these small details are changing due to the polarization by surrounding molecules.

For small or rigid molecules, it is straightforward to assign multipole parameters. For larger molecules with several conformations in contrast, the parametrization is more involved. Atomic multipole moments are highly dependent on the conformation. The same is true for point charges which are mostly used in atomistic force fields, but for multipole moments, the entire accuracy gain with respect to point charges may be lost upon conformational change and the additional computational cost is not justified anymore. Therefore, multipole moments for larger molecules should be averaged over several conformation or additional functions for changing the multipole moments with conformation should be used.

For the implementation of atomic multipole moments into atomistic force fields, the

orientation of the multipole moments needs to be linked to the orientation of the molecules in space. This can be done through the assignment of molecular reference axis systems, which are specific to each type of molecular geometry. The calculation of molecular orientations on the basis of reference axes is straightforward. As a consequence, the multipole interaction energies and derivatives depend on the positions of all the atoms belonging to their reference axis system. Approximate derivatives which include only atomistic terms may be used in some cases, but in general atomistic and molecular terms are required due to the change of the interaction energy upon rotation of the reference axes. This makes the implementation more complicated and less general, since specific terms for each type of reference axis system have to be introduced.

The conservation of the total energy during MD simulations is more problematic for atomic multipole moments than for point charges. For the full expression of the derivatives, the energy conservation is comparable to point charges as long as no cutoffs are used. The use of cutoffs generally introduces errors to simulations, but with atomic multipole moments, the problem is enhanced due to an incomplete compensation of errors.

The application of distributed multipole potentials to several systems shows that the use of more accurate electrostatics leads to a better agreement with experimental data:

- For photodissociated CO in myoglobin, different levels of accuracy of atomic multipole expansions are related to preferential positions and orientations of the CO molecule in the distal Heme pocket during the simulation. The occurrence of the CO orientations with respect to iron can be associated with the two experimentally observed IR bands for CO in myoglobin.

- For mixed CO/water ice systems, the same CO model combined with a multipolar water model provides IR spectra in agreement with experiment. Comparison of ice structures obtained from simulations with multipole potentials to ice structures from simulations with point charge potentials shows that the structure of the amorphous ice differs between the two setups. The amorphous ice structure found with the multipole water model can be related to experimentally observed red- and blueshifted CO IR frequencies.

78

- For methane clathrates, the lattice modes calculated with the atomic multipole force field are in good agreement with experimental INS spectra. The comparison of CO and methane clathrate hydrates provides insight into the differences in the clathrate structure for the two guest molecules.

In summary, the presented methods and applications show that the use of atomic multipole moments in atomistic force fields is complicated and computationally demanding, but the additional effort is justified by the good agreement with experimental data.

6 Future directions

The current version of the atomic multipole implementation in CHARMM can be extended and generalized in several aspects. The application is currently limited by the available geometry types and by the conformational dependence. Nevertheless, a variety of further applications are possible. Diatomic molecules such as O_2, H_2 and N_2 are of interest for several biological systems, e.g. for oxygen diffusion in Hemoglobin.[79] Simulations of for theses systems are mostly carried out without electrostatic interactions for the diatomics[80,81] or with simple point charge models in the case of CO.[82]

For ices of interest to astrochemistry, other ice mixtures have already been investigated experimentally[83–87] and there are open question to which computer simulations with distributed multipole potentials could probably provide answers: structure-frequency relationships in mixed H_2O:CO:CO_2 ice, diffusion of H_2 in amorphous ice, adsorption and desorption of different gas molecules from the ice surface.

The properties of water and ice as a function of the molecular multipole moments could also be investigated in more detail. The problem was studied previously for water by comparison of different empirical water models.[88,89] Atomic multipole moments provide the possibility to systematically vary the water dipole moment and the different components of the water quadrupole moment. For this purpose, the Ewald summation for atomic multipole moments[69] could be added, since the long-range interactions have been shown to be important for simulations of liquid water.[90,91]

An extension of the current implementation to more molecular geometries is straightforward as long as no flexible torsion angles are included. For a generalized implementation, a limited number of reference axis system should be sufficient to cover all molecular geometries. Larger molecules with flexible torsion angles can be assembled based on fragments classified by symmetry. A set of available symmetries should in-

clude linear, triangular planar, trigonal pyramidal, tetrahedral and octahedral. For most molecules, the problem of conformational dependence has to be solved in addition.

Several strategies have been proposed in the literature to solve this problem. Some of these approaches have already been tested. (see Section 2.5 for details.) The additional effort required to implement solutions based on these different strategies into atomistic force fields varies significantly. The consequences of each strategy are outlined in the following:

- If the multipole moments are just averaged over several conformations, no additional terms are required for the implementation. This approach requires only a more elaborate parametrization procedure. The disadvantage is that by averaging the multipole moments over several conformations, some of the accuracy is lost.

- If an additional analytical function for changing the multipole moments with conformation is used, fully accurate derivatives include additional terms depending on this change. The level of complexity for this approach is related to the complexity of the analytical function needed to fit the change of the multipole moments with conformation.

- The most involved but also most accurate approach would be to periodically recalculate the atomic multipole moments *ab initio*. The effort for the calculation could be limited by storing the multipole moments of conformations which have already been visited. In contrast to the other strategies, a fully analytical derivative would not be possible anymore in this case, since the change of the multipole moments would not follow an analytical function. However, this difficulty could possible be overcome by the use of appropriate MD integration schemes, which have already been developed for other simulation methods. In this case, recalculating the multipole parameters periodically could provide further advantages. One of this advantages is that intramolecular polarization could also be taken into account by including parts of the environment of a molecule in the *ab initio* calculations. A similar approach has already been proposed to account for the polarization in a point charge force field.[92] The level of accuracy of MD simulations based on such an approach could be systematically improved, on one hand by the level of theory used for the *ab initio* calculation,

on the other hand by changing the number of steps for recalculating the multipole moments.

From these consideration, it follows that although the current implementation can be applied to a variety of problems, generalization and extension including a periodic reparametrization procedure could provide a much more powerful tool.

Bibliography

1. A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. E. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, 102:3586–3616, 1998.

2. S. J. Weiner, P. A. Kollman, D. A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta Jr, and P. Weiner. A new force-field for molecular mechanical simulation of nucleic-acids and proteins. *J. Am. Chem. Soc.*, 106:765–784, 1984.

3. W. F. Van Gunstern and H. J. C. Berendsen. Groningen molecular simulation GROMOS library manual. 1987.

4. W. L. Jorgensen and J. Tirado-Rives. The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.*, 110:1657–1666, 1988.

5. M. Karplus and G. A. Petsko. Molecular dynamics simulations in biology. *Nature*, 347:631–639, 1990.

6. W. F. Van Gunstern, F. J. Luque, D. Timms, and A. E. Torda. Molecular mechanics in biology: from structure to function, taking account of solvation. *Ann. Rev. Biophys. Biomol. Struct.*, 23:847–863, 1994.

7. S. Erkoc. Empirical many-body potential energy functions used in computer simulations of condensed matter properties. *Phys. Rep.*, 278:80–105, 1997.

8. D.E. Shaw, M.M. Deneroff, A. O. Dror, J.S. Kuskin, R. H. Larson, J. K. SAlmon, C. Young, B. Batson, K.J. Bowers, J. C. Chao, M. P. Eastwood, J. Gagliardo,

J.P. Grossman, C. Richard, D. J. Ierardi, I. Kolossváry, J. L. Klepeis, T. Layman, C. Mcleavey, M. A. Moraes, R. Mueller, E. C. Priest, Y. Shan, J. Spengler, M. Theobald, B. Towles, and S. C. Wang. Anton, a special-purpose machine for molecular dynamics simulation. *Comm. ACM*, 51:91–97, 2008.

9. P. Maragakis, K. Lindorff-Larsen, R. O. Eastwood, M. P. Dror, J. L. Klepeis, I. T. Arkin, M. O. Jensen, H. Xu, N. Trbovic, R. A. Friesner, A. G. Palmer, and D. E. Shaw. Microsecond molecular dynamics simulation shows effect of slow loop dynamics on backbone amide order parameters of proteins. *J. Phys. Chem. B*, 112:6155–6158, 2008.

10. M. W. Van der Kamp, K. E. Shaw, C. J. Woods, and A. J. Mulholland. Biomolecular simulation and modelling:status, progress and prospects. *J. R. Soc.*, 5:S173–S190, 2008.

11. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.*, 4:187–217, 1983.

12. B. R. Brooks, C. L. Brooks III, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Cao, W. Hodoscek, M.and Im, K. Kuzera, T. Lazaridis, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus. CHARMM: The biomolecular simulation program. *J. Comp. Chem.*, 30:1545–1614, 2009.

13. P. P. Ewald. The calculation of optical and electrostatic grid potential. *Ann. d. Phys.*, 64:253–287, 1921.

14. E. Schrödinger. Quantisierung als Eigenwertproblem. *Ann. d. Phys.*, 79:361–376, 1926.

15. P. A. M. Dirac. A new notation for quantum mechanics. *Proc. Cambr. Phil. Soc.*, 35:416–418, 1939.

16. A. Szabo and N.S. Ostlund. *Modern Quantum Chemistry*. McGraw-Hill, New York, 1989.

17. M. Born and R. J. Oppenheimer. Zur Quantentheorie der Molekeln. *Ann. d. Phys.*, 84:457–484, 1927.

18. R. Car and M. Parrinello. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.*, 55:2471–2474, 1985.

19. P. Carloni, U. Röthlisberger, and M. Parrinello. The role and perspective of ab initio molecular dynamics in the study of biological systems. *Acc. Chem. Res.*, 35:455–464, 2002.

20. J. Aqvist and A. Warshel. Simulations of enzyme reactions unsing valence bond force fields and other hybrid quantum/classical approaches. *Chem. Rev.*, 93:2523–2544, 1993.

21. D. Bakowies and W. Thiel. Hybrid models for combined quantum mechanical and molecular mechanical approaches. *J. Phys. Chem.*, 100:10580–10594, 1996.

22. A. J. Stone. *The Theory of Intermolecular Forces*. Clarendon Press, Oxford, 1996.

23. G. C. Maitland, M. Rigby, E. B. Smith, and W. A. Wakeham. *Intermolecular Forces*. Clarendon Press, Oxford, 1989.

24. A. Warshel, M. Kato, and Pisliakov A. V. Polarizable force fields: History, test cases, and prospects. *J. Chem. Theo. Comp.*, 3:2034–2045, 2007.

25. A. J. Stone. Distributed polarizabilities. *Mol. Phys.*, 56:1065 – 1082, 1985.

26. N. Gresh, A. G. Cisneros, and J.-P. Darden, T. A. Piquemal. Anisotropic, polarizable molecular mechanics studies of inter- and intramolecular interactions and ligand-macromolecule complexes. A bottom-up strategy. *J. Chem. Theo. Comp.*, 3:1960–1986, 2007.

27. M. G. Darley, C. M. Handley, and P. L. A. Popelier. Beyond point charges: Dynamic polarization from neural net predicted multipole moments. *J. Chem. Theo. Comp.*, 4:14351448, 2008.

28. H. Primas and U. Müller-Herold. *Elementare Quantenmechanik*. Teubner, Stuttgart, Germany, 1984.

29. R.S. Mulliken. Electronic population analysis on LCAO-MO molecular wave functions. *J. Chem. Phys.*, 23:1833–1840, 1955.

30. A. J. Stone. Distributed multipole analysis, or how to describe a molecular charge distribution. *Chem. Phys. Lett.*, 83:233–239, 1981.

31. W.A. Sokalski and R.A. Poirier. Cumulative atomic multipole representations of the molecular charge distibution and its basis set dependence. *Chem. Phys. Lett.*, 98:86–92, 1983.

32. A. J. Stone and M. Alderton. Distributed multipol analysis-methods and applications. *Mol. Phys.*, 56:1047–1064, 1985.

33. R.F.W. Bader. *Atoms in Molecules; A Quantum Theory*. Clarendon Press, Oxford, 1990.

34. U. C. Singh and P. A. Kollman. An approach to computing electrostatic charges for molecules. *J. Comp. Chem.*, 5:129–145, 1984.

35. B. H. Besler, K. M. Merz, and P. A. Kollman. Atomic charges derived from semiempirical methods. *J. Comp. Chem.*, 11:431–439, 1990.

36. A. Sawaryn and W.A. Sokalski. Cumulative atomic multipole moments and point charge models describing molecular charge distribution. *Comput. Phys. Commun.*, 52:397–408, 1989.

37. P. L. A. Popelier. *Atoms in Molecules: An Introduction*. Prentice-Hall, Harlow, U. K., 2000.

38. V.I. Lebedev and D. N. Laikov. Quadrature formula for the sphere of 131-th algebraic order of accuracy. *Dokl. Math.*, 59:477–481, 1999.

39. A. J. Stone. Distributed multipole analysis: Stability for large basis sets. *J. Chem. Theo. Comp.*, 1:1128–1132, 2005.

40. D. R. Hartree. The wave mechnics of an atom with a non-coulomb central field. *Proc. Cambr. Phil. Soc.*, 24:89–110, 1928.

41. V. Fock. Approximate method of solution of the problem of many bodies in quantum mechanics. *Z. f. Phys.*, 61:126–148, 1930.

42. J. C. Slater. Note on Hartree's method. *Phys. Rev.*, 35:210–211, 1930.

43. C. C. J. Roothaan. New developments in molecular orbital theory. *Rev. Mod. Phys.*, 23:69–89, 1951.

44. G. G. Hall. The molecular orbital theory of chemical valency VIII. A method of calculating ionization potentials. *Proc. Roy. Soc. (London) A*, 205:541–552, 1951.

45. C. Møller and M. S. Plesset. Note on an approximation treatment for many-electron systems. *Phys. Rev.*, 46:618–622, 1934.

46. P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864–B871, 1964.

47. W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133–A1138, 1965.

48. U. Koch, P.L.A. Popelier, and A.J. Stone. Conformational dependence of atomic multipole moments. *Chem. Phys. Lett.*, 238:253–260, 1995.

49. M. P. Hodges, A. J. Stone, and S. S. Xantheas. Contribution of many-body terms to the energy for small water clusters: A comparison of ab initio calculations and accurate model potentials. *J. Phys. Chem. A*, 101:9163–9168, 1997.

50. L. Joubert and P.L.A. Popelier. The prediction of energies and geometries of hydrogen bonded DNA base-pairs via a topological electrostatic potentialy. *PCCP*, 4:4353–4359, 2002.

51. L. Joubert and P.L.A. Popelier. Improved convergence of the 'atoms in molecules' multipole expansion of electrostatic interaction. *Mol. Phys.*, 100:3357–3365, 2002.

52. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S 2E S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Fo resman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. Gaussian 03, revision c.01. 2004.

53. W. J. Hehre, R. Ditchfield, and J. A. Pople. Self-consistent molecular-orbital methods 12. Further extensions of Gaussian-type basis sets for use in molecular-orbital studies of organic-molecules. *J. Chem. Phys.*, 54:2257–2261, 1972.

54. P. C. Hariharan and J. A. Pople. Influence of polarization functions on molecular-orbital hydrogenation energies. *Theor. Chim. Acta*, 28:213–222, 1973.

55. T. H. Dunning, Jr. Gaussian-basis sets for use in correlated molecular calculations 1. The atoms boron through neon and hydrogen. *J. Chem. Phys.*, 90:1007–1023, 1989.

56. A. D. Becke. Density-funtional thermochemistry 3. The role of exact exchange. *J. Chem. Phys.*, 98:5648–5652, 1993.

57. C. Lee, W. Yang, and R. G. Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron-density. *Phys. Rev. B*, 37:785–789, 1988.

58. N. Plattner and M. Meuwly. Higher order multipole moments for molecular dynamics simulations. *J. Mol. Model.*, 15:687–694, 2009.

59. S.L. Price and A.J. Stone. Electrostatic models for polypeptides: Can we assume transferability ? *JCSF*, 88:1755–1763, 1992.

60. P. L. A. Popelier and R. F. W. Bader. Effect of twisting a polypeptide on its geometry and electron distribution. *J. Phys. Chem.*, 98:4473–4481, 1994.

61. P. Kedzierski and W.A. Sokalski. Analysis of the transferability of atomic multipoles for amino acids in modeling macromolecular charge distribution from fragments. *J. Comp. Chem.*, 22:1082–1097, 2001.

62. N. Plattner and M. Meuwly. The role of higher CO-multipole moments in understanding the dynamics of photodissociated carbonmonoxide in myoglobin. *Biophys. J.*, 94:2505–2515, 2008.

63. U. Koch and A.J. Stone. Conformational dependence of the molecular charge distribution and its influence on intermolecular interactions. *J. Chem. Soc. Faraday Trans.*, 92:1701–1708, 1996.

64. P. Ren and J.W. Ponder. Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations. *J. Comp. Chem.*, 23:1497–1506, 2002.

65. P. G. Karamertzanis and S. L. Price. Energy minimization of crystal structures containing flexible molecules. *J. Chem. Theo. Comp.*, 2:1184–1199, 2006.

66. S. L. Price, C. H. Faerman, and C. W. Murray. Toward accurate transferable electrostatic models for polypeptides: A distributed multipole study of blocked amino acid residue charge distributions. *J. Comp. Chem.*, 12:1187–1197, 1991.

67. C. Hättig and B. A. Hess. Calculation of orientation-dependent double-tensor moments for coulomb-type intermolecular interactions. *Mol. Phys.*, 81:813–824, 1994.

68. Fincham D. Leapfrog rotational algorithms. *Mol. Phys.*, 8:165–178, 1992.

69. M. Leslie. DL MULTI - A molecular dynamics program to use distributed multipole electrostatics models to simulate the dynamics of organic crystals. *Mol. Phys.*, 106:1567–1578, 2008.

70. 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. 2002.

71. N. Kumagai, K. Kawamura, and T. Yokokawa. An interatomic potential model for H_2O: Applications to water and ice polymorphs. *Mol. Sim.*, 12:177–186, 1993.

72. W. Humphrey, A. Dalke, and K. Schulten. VMD - visual molecular dynamics. *J. Molec. Graphics*, 14:33–38, 1996.

73. T. Laino, F. Mohamed, A. Laio, and M. Parrinello. An efficient real space multigrid QM/MM electrostatic coupling. *J. Chem. Theo. Comp.*, 1:1176–1184, 2005.

74. Guidon M., F. Schiffmann, J. Hutter, and J. VandeVondele. Ab initio molecular dynamics using hybrid density functionals. *J. Chem. Phys.*, 128:214104, 2008.

75. Mark E. Tuckerman, Bruce J. Berne, Glenn J. Martyna, and Michael L. Klein. Efficient molecular dynamics and hybrid monte carlo algorithms for path integrals. *J. Chem. Phys.*, 99:2796–2806, 1993.

76. R. W. Hall and B. J. Berne. Nonergodicity in path integral molecular dynamics. *J. Chem. Phys.*, 81:3641–3643, 1984.

77. J. Kolafa. Time-reversible always stable predictorcorrector method for molecular dynamics of polarizable molecules. *J. Comp. Chem.*, 25:335–342, 2003.

78. P. Ren and J. W. Ponder. Polarizable atomic multipole water model for molecular mechanics simulation. *J. Phys. Chem. B*, 107:5933–5947, 2003.

79. J. S. Olson. From O_2 diffusion into red blood cells to ligand pathways in globins. *Protein Rev.*, 9:161–181, 2008.

80. J. Cohen, A. Arkhipov, R. Braun, and K. Schulten. Imaging the migration pathways for O_2, CO, NO and Xe inside myoglobin. *Biophys. J.*, 91:1844–1857, 2006.

81. J. Cohen, K. Kim, P. King, M. Seibert, and K. Schulten. Finding gas diffusion pathways in proteins: Application to O_2 and H_2 transport in CpI [FeFe]-hydrogenase and the role of packing defects. *Structure*, 13:1321–1329, 2005.

82. Ruscio J. Z., Kumar D., M. Shukla, M. G. Prisant, T. M. Murali, and A. V. Onufriev. Atomic level computational identification of ligand migration pathways between solvent and binding site in myoglobin. *Proc. Natl. Acad. Sci.*, 105:9204–9209, 2008.

83. K. M. Merrill, R. W. Russel, and B. T. Soifer. Infrared observations of ices and silicates in molecular clouds. *Astrophys. J*, 207:763–769, 1976.

84. R. W. Dissly, M. Allen, and V. G. Anicich. H_2-rich interstellar grain mantles: an equilibrium description. *Astrophys. J*, 435:685–692, 1994.

85. P. A. Gerakines, W. A. Schutte, J. M. Greenberg, and E. F. van Dishoeck. The infrared band strength of H_2O, CO and CO_2 in laboratory simulations of astrophysical ice mixtures. *Astron. Astrophys.*, 296:810–818, 1995.

86. J. E. Roser, S. Swords, G. Vidali, G. Manico, and V. Pironello. Measurement of the kinetic energy of hydrogen molecules desorbing from amorphous water ice. *Astrophys. J*, 596:L55–L58, 2003.

87. K. I. Oeberg, H. J. Fraser, A. C. A. Boogert, S. E. Bisschop, G. W. Fuchs, E. F. van Dishoeck, and H. Linnartz. Effects of CO_2 on H_2O band profiles and band-strengths in mixed $H_2O:CO_2$ ices. *Astron. Astrophys.*, 462:1187–1198, 2006.

88. J. L. F. Abascal and C. Vega. The water forcefield: Importance of dipolar and quadrupolar interactions. *J. Phys. Chem. C*, 111:15811–15822, 2007.

89. S. W. Rick. A reoptimization of the five-site water potential (TIP5P) for use with Ewald sums. *J. Chem. Phys.*, 120:6085–6093, 2004.

90. M. Belhadj, H. E. Alper, and R. M. Levi. Molecular dynamics simulations of water with Ewald summation for the long range electrostatic interactions. *Chem. Phys. Lett.*, 179:13–20, 1991.

91. K. F. Lau, H. E. Alper, T. S. Thacher, and T. R. Stouch. Effects of switching functions on the behavior of liquid water in molecular dynamics simulations. *J. Phys. Chem.*, 98:8785–8797, 1994.

92. J. L. Banks, H. S. Beard, Y. Cao, A. E. Cho, W. Damm, R. Farid, A. K. K. Felts, T. A. Halgren, D. T. Mainz, J. R. Maple, R. Murphy, D. M. Philipp, M. P. Repasky, L. Y. Zhang, B. J. Berne, R. A. Friesner, E. Gallicchio, and R. M. Levy. Integrated modeling program, applied chemical theory (IMPACT). *J. Comp. Chem.*, 26:1752-1780, 2005.

Appendix

Curriculum vitae

Personal Data

| | Nuria Plattner |
| Email: | Nuria.Plattner@unibas.ch |

Education

12/2005 – 6/2009	**PhD Studies, Department of Chemistry, University of Basel**[*]
	Thesis: Distributed Multipole Moments in Atomistic Force Fields: Implementation and Applications
	Supervisor: Prof. M. Meuwly, University of Basel
	External Examiner: Prof. M. Parrinello, ETH Zürich
09/2001 – 10/2005	**Studies in Chemistry, University of Basel**
	Thesis: An Analysis of the Chemical and 3D Structural Diversity of CDK2 Ligand- Protein Complexes with Implications for Virtual Screening.
	Supervision: Dr. J. Priestle, Novartis, Prof. M. Meuwly, University of Basel
09/1999 – 07/2001	**Studies in Philosophy, University of Basel**
08/1996 – 07/1999	**Teacher Training School, Lehrerseminar Kreuzlingen**

---

[*]Of particular importance for the PhD thesis were the lecures of Prof. H. Huber, Prof. J.P. Maier, Prof. M. Meuwly and Prof. R. Buser, attended during undergraduate and PhD studies.

Publications

- **The Role of Higher CO-Multipole Moments in Understanding the Dynamics of Photodissociated Carbonmonoxide in Myoglobin**
  N. Plattner, M. Meuwly
  Biophys. J. **2008**, *94*: 2505 – 2515

- **Atomistic Simulations of CO Vibrations in Ices Relevant to Astrochemistry**
  N. Plattner, M. Meuwly
  ChemPhysChem **2008**, *9*:1271 – 1277

- **MD simulations using distributed multipole electrostatics: Structural and spectroscopic properties of CO- and methane-containing clathrates**
  N. Plattner, T. Bandi, J.D. Doll, D.L. Freeman and M. Meuwly
  Mol. Phys. **2008**, *106*:1675 – 1684

- **Higher order multipole moments for molecular dynamics simulations**
  N. Plattner, M. Meuwly
  J. Mol. Model. **2009**, *15*:687 – 694

Submitted for publication:

- **Application of Multipolar Charge Models and Molecular Dynamics Simulations to Study Stark Shifts in Non-Uniform Electric Fields**
  M. Devereux, N. Plattner and M. Meuwly

- **A Spatial Averaging Approach to Rare-Event Sampling**
  J. D. Doll, J. E. Gubernatis, N. Plattner, M. Meuwly, P. Dupuis and H. Wang