

*chapter three*

# FORCE FIELD PARAMETERS

*from  
quantum  
chemical  
calculations*

## CONTENTS

3.1	<i>Multipole derived charge analysis</i>	27-39
3.2	<i>IntraFF force constants</i>	40-57

## SUMMARY

*Two methods are presented to obtain force field parameters that are needed in classical mechanics calculations. A new charge analysis is presented that provides atomic charges that reproduce by construction the atomic and therefore also the molecular multipole moments, and a new method to obtain from quantum chemical Hessian matrices force constants for the bonding interactions in classical mechanics is discussed.*

## 3.1

**Multipole derived charge analysis***A charge analysis derived from an atomic multipole expansion*

The electrostatic potential from the charge density  $\rho(r_i)$  of a molecule is obtained in a point  $r_s$  as<sup>a</sup>:

$$V_C(r_s) = \int \frac{\rho(r_i)}{|r_s - r_i|} dr_i \quad (1)$$

However, this equation is not well suited for standard use in molecular simulations. One could simplify the integral by using a one-center multipole expansion, but this can not be applied for larger molecules, since one needs to go to high order multipoles. Associated with it is a large convergence radius, outside which the expansion is valid. A better method (for instance the Distributed Multipole Analysis<sup>167,168</sup>) is to use multipoles located at several centers (usually the atoms) with two advantages: the order of multipoles needed is smaller and the convergence radius is per center (and smaller). However, even these methods are cumbersome to use in simulations: all multipole moments should be transformed from the local (i.e. the one in which they were obtained) to the global coordinate system (i.e. the one in the simulation), which involves a number of subsequent [3x3]-transformations for each moment (1 for dipoles, 2 for quadrupoles, etc.). A further disadvantage is that in Molecular Dynamics simulations for the calculation of the forces the derivative of the potential is needed, making it even more cumbersome.

A further simplification is therefore worthwhile: taking only (atomic) charges into consideration. This reduces not only the number of interactions per pair of centers (from 10-35 to 1, depending on the order of multipoles located at the centers), but removes also the need for transforming the moments. The potential is then obtained from the charges  $q_i$  at positions  $r_i$  as:

$$V_C(r_s) = \sum_i \frac{q_i}{|r_s - r_i|} \quad (2)$$

Therefore, assigning charges to atoms is an important problem in computational chemistry<sup>1</sup>, especially in view of constructing *accurate* force fields for simulations using Molecular Dynamics or Monte Carlo techniques<sup>160,161</sup>. Our interest is in the field of biochemistry, in particular copper proteins<sup>8</sup>, where the standard force field parameters, if present, are not accurate enough to model the chemical processes going on in the active site of the proteins. This is especially true for the study of properties like the redox potential, where one wants to see the behavior of the protein when an electron is removed from the system.

Many methods to obtain atomic charges from a quantum chemical calculation are available (detailed surveys can be found elsewhere<sup>1,169,170</sup>), where there is a general preference in the molecular modeling world for *Potential Derived* (PD) charges<sup>1,171</sup>. However, associated with this class of charge analyses are several major drawbacks, all of which are related more or less to the need of a grid of points where the quantum chemical potential is fitted by the electrostatic potential from the charges. It results in:

- a strong dependence on the choices made for determining the grid:

---

<sup>a</sup> In atomic units (see appendix A.1 at the end of the thesis)

- an arbitrary change in parameters results in different atomic charges
- some methods are even orientation-dependent<sup>172,173</sup>
- a strong dependence on the method how to fit the charges:
  - underdetermined set of equations, multiple "solutions" possible
  - numerical (in)stability
  - great uncertainties in assigning charges to atoms *buried* within the molecule, since the potential outside the molecule is mainly determined by the atoms near the surface<sup>1</sup>
  - an enormous increase in CPU-time, since lots of  $1/r$  terms resulting from charge-gridpoint distances should be evaluated many times

Several authors have extended these PD-methods by introducing constraints on the atomic charges; usually the *molecular* multipole moments up to some order should be conserved. This seems to make the methods less grid-dependent, but they are still limited to a relatively small number of atoms (approx. 20; depending on the order of the multipoles<sup>171</sup>) and still suffer from the other drawbacks.

We present here a new charge analysis based on ideas used for the *Dipole Preserving Charge analysis*<sup>174</sup>, but formulated in another way and using more accurate *atomic* multipoles. There are three stages involved in this method: we start by writing the molecular charge density as a sum of atomic densities. Subsequently, from these atomic densities a set of atomic multipoles can be defined, which can be used to get the electrostatic potential outside the charge distribution. Finally, these atomic multipoles are reconstructed by using a scheme that distributes charges over all atoms to reproduce these multipoles *exactly*. Therefore, this method does not suffer from the drawbacks of the PD-methods.

An important feature to notice here further, is the advantage of using an *atomic* over *molecular* multipole expansion. Close to any atom the electrostatic potential is mainly determined by the charge distribution around that atom; or within the *atomic* multipole expansion, by the *atomic* multipoles near to that point. This is one of the main advantages over using molecular multipoles only, when one needs to go to high orders to get a good representation of the potential in that point (if at all). Moreover, since the *molecular* multipole moments (up to order X) are reproduced necessarily by the *atomic* multipoles (up to order X), our charges have the nice feature that they do not only represent the *atomic* but also the *molecular* multipoles.

This section consists of three parts: the first deals with how to get from a molecular density to a set of atomic multipoles, in the second we explain the distribution scheme, while in the third, we give some results of the method.

#### *Atomic multipoles from a multipole expansion*

A molecular charge density  $\rho$  is usually obtained in a basis set expansion (with atom indices  $A, B$ , basis functions  $\varphi_j$ , basis function indices  $i, j$  and  $P_{ij}^{AB}$  element  $ij$  of the density matrix):

$$\rho = \sum_{AB} \rho_{AB} = \sum_{ABij} P_{ij}^{AB} \varphi_i^A \varphi_j^B \quad (3)$$

which defines the density as a sum of atom pair densities  $\rho_{AB}$ . Next, we fit these atom pair densities by using atomic functions  $f_i$  (without specific details about the functions to keep the discussion general; our choice and motivation will be specified in the **ADF implementation** part):

$$\tilde{\rho}_{AB} = \sum_i d_i^A f_i^A + \sum_j d_j^B f_j^B \quad (4)$$

where the coefficients can be obtained from minimization of the density differences:

$$\left( \sum_{AB} \rho_{AB} \right)^2 d\mathbf{r} \quad (5)$$

The total density can now be written as a sum of atomic densities:

$$\tilde{\rho} = \sum_{AB} \tilde{\rho}_{AB} = \sum_{AB} \sum_i d_i^A f_i^A + \sum_j d_j^B f_j^B = \sum_{Ai} c_i^A f_i^A = \sum_A \tilde{\rho}_A \quad (6)$$

Furthermore, using these functions  $f_i$ , the electrostatic potential in a point  $\mathbf{s}$  is given by

$$V_C(\mathbf{r}_s) = \sum_A V_C^A(\mathbf{r}_s) = \sum_A \sum_{i \in A} c_i^A \int \frac{f_i^A(\mathbf{r})}{|\mathbf{r} - \mathbf{r}_s|} d\mathbf{r} \quad (7)$$

Next, an atomic multipole expansion of the  $r^{-1}$  term can be used:

$$V_C^A(\mathbf{r}_s) = \sum_l \sum_{m=-l}^{m=l} \frac{4\pi}{2l+1} \frac{M_{lm}^A Z_{lm}(\hat{\mathbf{R}}_{sA})}{R_{sA}^{l+1}} \quad (8)$$

with the real spherical harmonics  $Z_{lm}$ <sup>175</sup> and the multipole "moments"  $M_{lm}^A$  where  $R_{sA}$  is the distance vector from nucleus A to a point  $r_s$ . These "moments" can be obtained from the coefficients as:

$$M_{lm}^A = \sum_{i \in A} c_i^A \int f_i^A(\mathbf{r}_2) \mathbf{r}_2^l Z_{lm}(\mathbf{r}_2) d\mathbf{r}_2 \quad (9)$$

These multipole "moments" are not equal to the multipoles from the Buckingham convention (see **Computational details**), but they can be easily transformed into those by inserting the Cartesian expressions for the  $Z_{lm}$ 's.

### Atomic multipole derived charge analysis

We start from the *atomic* multipoles as they are obtained from the multipole expansion. Then for each atom, we reconstruct the set of multipoles (up to some order X) located on that atom by *redistributed* charges  $q_{s,A}$  on *all* atoms that preserve (up to order X) that particular set of multipoles. The multipole moments represented by the charges  $q_{s,A}$  are obtained with the position vectors relative to atom A,  $r_{is,A}$ , as:

$$\begin{aligned} Q_A^{repr} &= \sum_s q_{s,A} \\ \rho_{i,A}^{repr} &= \sum_s q_{s,A} r_{is,A} \\ \rho_{jk,A}^{repr} &= \sum_s q_{s,A} \left( \frac{3}{2} r_{js,A} r_{ks,A} - \frac{1}{2} \delta_{jk} r_{s,A}^2 \right) \\ \rho_{lmn,A}^{repr} &= \sum_s q_{s,A} \left( \frac{5}{2} r_{ls,A} r_{ms,A} r_{ns,A} - \frac{1}{2} \delta_{lm} r_{ns,A} r_{s,A}^2 - \frac{1}{2} \delta_{nl} r_{ms,A} r_{s,A}^2 - \frac{1}{2} \delta_{mn} r_{ls,A} r_{s,A}^2 \right) \end{aligned} \quad (10)$$

When the number of atoms is larger than the total number of multipole moments (per atom) to be reconstructed, there is of course more than one way to distribute the charges. We therefore use a weight function that falls off rapidly to keep the *atomic* multipoles as local as possible, i.e. as close as possible to the atom where the multipoles are located:

$$w_s = \exp\left(-\frac{|r_s - r_A|}{d_A}\right) \quad (11)$$

where  $d_A$  is the distance from atom  $A$  to its nearest neighbor,  $\lambda$  an exponential prefactor and  $w_s$  the weight for atom  $s$ , when distributing the multipole moments of atom  $A$ .

Now, we want the *redistributed* charges to be as small as possible, and at the same time constrain the represented multipoles ( $Q^{repr}$ ,  $\mu^{repr}$ ,  $\nu^{repr}$ ,  $\tau^{repr}$ ) to be equal to the *atomic* multipoles from the multipole expansion ( $Q^{mpe}$ ,  $\mu^{mpe}$ ,  $\nu^{mpe}$ ,  $\tau^{mpe}$ ). We achieve this by minimizing the following function where the constraints are met by using Lagrangian multipliers  $\lambda_A$ ,  $\lambda_{i,A}$ ,  $\lambda_{jk,A}$ ,  $\lambda_{lmn,A}$ :

$$g_A = \sum_s \frac{q_{s,A}^2}{2w_{s,A}} + \lambda_A (Q_A^{mpe} - Q_A^{repr}) + \sum_i \lambda_{i,A} (\mu_{i,A}^{mpe} - \mu_{i,A}^{repr}) + \sum_{jk} \lambda_{jk,A} (\nu_{jk,A}^{mpe} - \nu_{jk,A}^{repr}) + \sum_{lmn} \lambda_{lmn,A} (\tau_{lmn,A}^{mpe} - \tau_{lmn,A}^{repr}) \quad (12)$$

With this choice of function, we ensure that the distribution mainly takes place close to the atom  $A$  where the multipoles are located. After taking the derivative of this function with respect to the redistributed charges  $q_{s,A}$  to find the optimal choice for the charges, the following equation is obtained:

$$q_{s,A} = w_{s,A} \lambda_A + \sum_i \lambda_{i,A} r_{is,A} + \sum_{jk} \lambda_{jk,A} \left( \frac{3}{2} r_{js,A} r_{ks,A} - \frac{1}{2} \lambda_{jk} r_{s,A}^2 \right) + \sum_{lmn} \lambda_{lmn,A} (\dots) \quad (13)$$

that shows clearly that points far away from atom  $A$  (and thus a small weight  $w_s$ ) get a small redistributed charge.

Using the constraints, we obtain for the Lagrangian multipliers:

$$Q_A - \sum_s q_{s,A} = 0 \quad (14)$$

$$Q_A = \lambda_A \sum_s w_{s,A} + \sum_i \lambda_{i,A} \sum_s w_{s,A} r_{is,A} + \sum_{jk} \lambda_{jk,A} \sum_s w_{s,A} \left( \frac{3}{2} r_{js,A} r_{ks,A} - \frac{1}{2} \lambda_{jk} r_{s,A}^2 \right) + \dots$$

$$\mu_{i,A} - \sum_s q_{s,A} r_{is,A} = 0 \quad (15)$$

$$\mu_{i,A} = \lambda_A \sum_s w_{s,A} r_{is,A} + \sum_i \lambda_{i,A} \sum_s w_{s,A} r_{is,A} r_{is,A} + \sum_{jk} \lambda_{jk,A} \sum_s w_{s,A} r_{is,A} \left( \frac{3}{2} r_{js,A} r_{ks,A} - \frac{1}{2} \lambda_{jk} r_{s,A}^2 \right) + \dots$$

$$\nu_{tu,A} - \sum_s q_{s,A} \left( \frac{3}{2} r_{ts,A} r_{us,A} - \frac{1}{2} \lambda_{tu} r_{s,A}^2 \right) = 0 \quad (16)$$

$$\begin{aligned}
 \square_{tu,A} = & \square_A \square_s w_{s,A} \left( \frac{3}{2} r_{ts,A} r_{us,A} \square \frac{1}{2} \square_{tu} r_{s,A}^2 \right) + \square_i \square_{i,A} \square_s w_{s,A} \left( \frac{3}{2} r_{ts,A} r_{us,A} \square \frac{1}{2} \square_{tu} r_{s,A}^2 \right) r_{is,A} \\
 & + \square_{jk} \square_{jk,A} \square_s w_{s,A} \left( \frac{3}{2} r_{ts,A} r_{us,A} \square \frac{1}{2} \square_{tu} r_{s,A}^2 \right) \left( \frac{3}{2} r_{js,A} r_{ks,A} \square \frac{1}{2} \square_{jk} r_{s,A}^2 \right) + \dots
 \end{aligned} \tag{16}$$

and a similar equation for the octupole moments. This results for each atom in a set of linear equations (of size 4 when distributing up to dipole moment [ $Q, \square_i$ ], 10 up to quadrupoles [ $Q, \square_i, \square_{jk}$ ] and 20 up to octupoles [ $Q, \square_i, \square_{jk}, \square_{lmn}$ ]) for the Lagrangian multipliers, which are solved by a standard  $Ax = b$  routine.

Finally, the values obtained for the Lagrange multipliers are used to get the *redistributed* atomic charges which when summed, result in the Multipole Derived Charges (up to some order X):

$$q_s^{MDC} = \square_A q_{s,A} \tag{17}$$

In the following we shall refer to the charges as MDC-D charges if the multipoles are reconstructed up to the dipole moment, MDC-Q if up to quadrupole (recommended to be used) and MDC-O if up to the octupole moment.

## Computational details

### Molecule set

We investigated a set of 31 molecules<sup>a</sup> to obtain a good test of the quality of the method, that were taken predominantly from ref. <sup>171</sup>, and extended with some that were of interest to us. Furthermore, as a second test set, we used all amino acid residues. Since we are, in *this* study, not interested in creating a generally applicable force field for amino acid residues for use in MD simulations of proteins, we took as a model for the amino acid residues a reduced conformation as they appear in proteins. I.e. replacing the  $NH_3^+$  and  $COO^-$  groups by  $NH_2$  and  $CHO$ . In fact, this means we cut off the backbone and replaced it by hydrogens. We are aware of the fact that this is not the standard model being used for getting amino acid residue charges, but we like to keep the model as simple as possible. Moreover, for a few amino acid residues, we also tested the so-called dipeptide model ( $CH_3CONHCHRCONHCH_3$ ) for the side chain  $R$ , which resulted in virtually the same charges in the side-chain.

For all molecules, we first optimized the geometries, then we performed single point energy calculations to get the molecular properties (energies, multipoles). All calculations were carried out within the Density Functional Theory framework<sup>1</sup> with the ADF program<sup>118,176-178</sup> using the Becke-Perdew exchange-correlation<sup>120,121</sup> potential in the TZ2P ( $V$  in ADF terms) basis set. In the following we shall make a distinction between the *fitted* and “*exact*” molecular multipole moments: the former refer to the values obtained from the *fitted* density (or from the MDC charges since they represent them exactly), while the latter refers to the values from the “*exact*” density.

<sup>a</sup> The set of 31 molecules consisted of: benzene, ethylene, methylcyanide, methyl lithium, methanol, methanol anion, methanethiol, methanethiol anion, methane, chlorine, carbonmonoxide, carbondioxide, carbondisulfide, fluorine, hydrogen, formaldehyde, hydrogenchloride, formamide, formic acid anion, hydrogenfluoride, hexafluorobenzene, hydrogen fluoride, lithiumhydride, nitrogen, ammonia, oxygen, dimethylether, hydroxyl anion, phosphine, thiophene, water, water dimer

### Exponential prefactor

This left us the task to find a value for the exponential prefactor  $\alpha$ , which we want to make as large as possible to keep the multipoles local. On the other hand, if we make it too big, the weight function will approach a delta function, leading to hardly any freedom to distribute the charges over the other atoms and numerical instabilities. We took as “optimal” value the largest value where the errors in the represented multipoles (due to machine precision and numerical accuracies) of the amino acid residues set were smaller than the required accuracy of the numerical integration in the calculations ( $10^{-6}$  in these cases), which turned out to be 3.0.

### Buckingham convention

In the literature, several conventions are being used for the multipole moments. In this study, we use the most commonly used Buckingham convention<sup>179-181</sup>. This convention has the following expressions for the multipole moments:

$$\begin{aligned} \mu_i &= \int r_i d\rho \\ \mu_{jk} &= \int \left( \frac{3}{2} r_j r_k - \frac{1}{2} \mu_{jk} r^2 \right) d\rho \\ \mu_{lmn} &= \int \left( \frac{5}{2} r_l r_m r_n - \frac{1}{2} \mu_{lm} r_n r^2 - \frac{1}{2} \mu_{mn} r_l r^2 - \frac{1}{2} \mu_{nl} r_m r^2 \right) d\rho \\ &\dots \end{aligned} \quad (18)$$

In this convention, the electrostatic potential is obtained as follows:

$$V_C^A(r_s) = \frac{Q_A}{r_{sA}} + \sum_i \frac{\mu_i r_{sA,i}}{r_{sA}^3} + \frac{1}{2} \sum_{jk} \frac{\mu_{jk} r_{sA,j} r_{sA,k}}{r_{sA}^5} + \frac{1}{6} \sum_{lmn} \frac{\mu_{lmn} r_{sA,l} r_{sA,m} r_{sA,n}}{r_{sA}^7} + \dots \quad (19)$$

Inserting the expressions for the  $Z_{lm}$ ’s in the multipole expansion (eqn. (8)), and rewriting them into Cartesian components, the relation to the Buckingham multipoles can easily be derived.

All *molecular* multipole moments reported here have been obtained relative to the center of mass of the molecule.

### Point charges in non-atomic (dummy) positions

For small systems, a proper charge distribution can not be represented by assigning charges only to atoms<sup>1</sup>. This is most easily seen for a homonuclear diatomic molecule like hydrogen. When using charges on the two atoms only, all methods should result in charges of exactly zero, because of symmetry. Adding a *third* (dummy) point, for instance in the center, should then result in a much better description of the charge distribution within the molecule (see also Allen/Tildesley<sup>160,161</sup>, who use nitrogen as an example and use 5 points with rather large charges). In practice, adding dummy points is only necessary for small molecules, and in all cases adding one point (which is placed in bond midway points in this section) already suffices to reproduce the multipole moments up to the quadrupole moment.

### ADF implementation

The method described in this section is generally applicable, however, we use Density Functional Theory<sup>1</sup> as incorporated in the ADF program<sup>118,176-178</sup> and take advantage of the way the electrostatic potential is being calculated there. The program uses a numerical integration scheme, by employing a



grid around the atoms to do the integration<sup>176</sup>. It uses a basis set of Slater type orbitals<sup>117</sup>:

$$\chi_u(r, \theta, \phi) = r^{n_u} e^{-\zeta_u r} Y_{lm}(\theta, \phi) \quad (20)$$

centered on each atom, with an auxiliary set of (Slater-type) fitfunctions (called *fit set*) to approximate the density  $\rho$  by expansion in these one-center functions  $f_i$  with coefficients  $a_i$ 's:

$$\rho(\mathbf{r}) = \sum_{i \in A} P_{lm}(\theta, \phi) \chi_i(\mathbf{r}) \quad (21)$$

The atomic multipoles are obtained from the coefficients:

$$M_{lm}^A = \sum_{i \in A} a_i \int f_i(\mathbf{r}_2) r_2^l Y_{lm}(\mathbf{r}_2) d\mathbf{r}_2 \quad (22)$$

where the  $a$ -coefficients are obtained from a least squares minimization of the difference  $\chi_{AB}$  between the "exact" and *fitted* density:

$$\chi_{AB} = \int |\rho_{AB} - \rho_{AB}^f|^2 d\mathbf{r} \quad (23)$$

The electrostatic potential is now obtained as<sup>118</sup>:

$$V_C(r_s) = \sum_A V_C^A(r_s) = \sum_A \sum_{i \in A} a_i \int \frac{f_i(r)}{r_s} \frac{1}{r} dr \quad (24)$$

or, by using the expansion of  $\frac{1}{r_s} \frac{1}{r}$  in spherical harmonics and using the exponential form of the STO fit functions (with the exponent  $\zeta_i$  and the principal quantum number  $n_i$ ):

$$\begin{aligned} V_C^A(r_s) &= \sum_l \sum_{m=-l}^{m=l} \frac{4\pi}{2l+1} Y_{lm}(r_{sA}) I_{lm}^A(r_{sA}) \\ I_{lm}^A(r_{sA}) &= \sum_{i \in A} \int \zeta_i^l Y_{lm}(\theta, \phi) a_i I(n_i, l, \zeta_i, r_{sA}) \end{aligned} \quad (25)$$

The function  $I$  is obtained from incomplete Gamma functions, and can be written as the sum of a multipolar and exponentially decaying part<sup>118</sup>:

$$I(n_i, l, \zeta_i, r_{sA}) = \frac{1}{(r_{sA})^{l+1}} \frac{(n_i + l + 1)}{(\zeta_i)^{n_i + l + 2}} + e^{-\zeta_i r_{sA}} J(n_i, l, \zeta_i, r_{sA}) \quad (26)$$

The function  $J$  consists of a power series in  $r_{sA}$ , with  $n_i$  as highest power, and serves as a screening for the short-range behavior of the multipole expansion. Since the expansion with the screening is correct inside the molecule, outside the molecule, where the screening is absent (due to the short range of action), the expansion is also valid and thus gives the correct potential.

A lot of effort has been put in constructing appropriate *fit sets*, and the current standard sets are very well qualified to reproduce electrostatic potentials inside the molecule accurately. Usually, the set of fit functions is larger than the set of basis functions, and sufficiently large to give a good description

of the electrostatic potential within the molecule. However, comparing the *fitted* molecular moments with the “*exact*” values from the basis functions, we found some differences, which were due to fit set incompleteness. To remedy this effect, we constructed a new set of fit functions by adding p and d-functions for the atoms involved in our test set of molecules, except for hydrogen that was not altered at all. With this new set, the differences between the “*exact*” and *fitted* molecular multipole moments were much smaller (see Table 3.1.1 for values averaged over the set of 31 molecules). From this Table, it can be seen that while the new fit set gives better *molecular* multipole moments, there is hardly any influence on the total energy.

TABLE 3.1.1. COMPARISON OF *STANDARD* AND *NEW* FIT SET. ABSOLUTE DIFFERENCES BETWEEN “*EXACT*” AND *FITTED* MOLECULAR MULTIPOLE MOMENTS, IN TOTAL ENERGY AND FIT TEST DIFFERENCE, ALL AVERAGED OVER THE SET OF 31 MOLECULES (A.U.)

	<i>standard</i>		<i>new</i>	
	average	maximum	average	maximum
$Q_{\text{tot}}$	0.000	0.000	0.000	0.000
$\square_k$	0.004	0.039	0.004	0.060
$\square_y$	0.002	0.023	0.001	0.014
$\square_z$	0.038	0.303	0.026	0.258
$\square_{xx}$	0.490	3.640	0.105	0.267
$\square_{xy}$	0.047	0.807	0.035	0.413
$\square_{xz}$	0.000	0.000	0.000	0.000
$\square_{yy}$	0.373	1.788	0.102	0.251
$\square_{yz}$	0.000	0.000	0.000	0.000
$\square_{zz}$	0.685	3.577	0.139	0.502
Fit test	5.05·10 <sup>-5</sup>		3.95·10 <sup>-5</sup>	
Energy difference	0.029 kcal/mol			

#### *Extensions of the implementation*

The standard *fit* sets in the ADF program have been constructed in order to minimize the difference between the “*exact*” and *fitted* density, thereby obtaining from both accurate electrostatic potentials in the system under study. However, no specific constraints were put on the functions to conserve the molecular multipole moments. Moreover, in the standard versions of ADF, the expectation value of only the dipole moment (from the “*exact*” density) is being calculated. We implemented the calculation of the expectation value of the quadrupole moment tensor, in order to check the represented quadrupole moments from both the “*exact*” and *fitted* densities.

In Table 3.1.1, we give the absolute differences between the *molecular* multipole moments from the “*exact*” density on one hand, the *fitted* density on the other, averaged over the set of 31 molecules. We report the values for both the standard fit set (TZ2P basis set, V in ADF terminology) as well as a new one, where we added p- and d-functions to the fit set for the first row-atoms. We also give the values of the fit test for both *fit* sets, as well as the averaged absolute energy differences between the two *fit* sets. These numbers give a clear indication that there is hardly any influence on the total energy and “*exact*”/*fitted* density difference. There is however a clear improvement of the represented multipole moments, with an average absolute deviation of 0.10 au for the quadrupole moments.

TABLE 3.1.2 COMPARISON OF MOLECULAR MULTIPOLE MOMENTS (A.U.)

<i>Multipole</i>	<i>DFT Exact</i>	<i>DFT Fitted</i>	<i>Experimental</i>
<i>benzene</i>			
$\square_{xx} = \square_{yy}$	2.78	2.79	3.23
$\square_{zz}$	-5.56	-5.59	-6.47 ± 0.37
<i>carbonmonoxide</i>			
$\square_z$	-0.07	-0.09	-0.04
$\square_{xx} = \square_{yy}$	0.76	0.74	
$\square_{zz}$	-1.52	-1.48	-1.86 ± 0.22
<i>carbondisulfide</i>			
$\square_{xx} = \square_{yy}$	-1.35	-1.58	
$\square_{zz}$	2.71	3.15	3.17 ± 0.22
<i>fluorine</i>			
$\square_{xx} = \square_{yy}$	-0.38	-0.33	
$\square_{zz}$	0.76	0.66	0.56
<i>hydrogen</i>			
$\square_{xx} = \square_{yy}$	-0.25	-0.31	-0.24
$\square_{zz}$	0.51	0.61	0.47 ± 0.03
<i>hexafluorobenzene</i>			
$\square_{xx} = \square_{yy}$	-2.99	-2.78	-3.53
$\square_{zz}$	5.99	5.56	7.07 ± 0.37
<i>hydrogenfluoride</i>			
$\square_z$	-0.69	-0.69	-0.72
$\square_{xx} = \square_{yy}$	-0.87	-0.95	
$\square_{zz}$	1.74	1.91	1.75 ± 0.02
<i>nitrogen</i>			
$\square_{xx} = \square_{yy}$	0.561	0.490	0.76
$\square_{zz}$	-1.122	-0.979	-1.52
<i>oxygen</i>			
$\square_{xx} = \square_{yy}$	0.071	0.065	0.20
$\square_{zz}$	-0.142	-0.131	-0.39
<i>water</i>			
$\square_z$	-0.71	-0.73	-0.73
$\square_{xx}$	-1.82	-1.93	-1.86 ± 0.01
$\square_{yy}$	1.91	2.01	1.96 ± 0.01
$\square_{zz}$	-0.09	-0.08	-0.10 ± 0.02

## Results

### *Multipole moments*

Although this section is strictly speaking not concerned with the accuracy of calculating molecular multipole moments, at least it should be established that Density Functional Theory is able to provide reasonable values for them. As the atomic charges are constructed to represent these molecular moments, in order to get reliable charges the molecular multipoles should be correct. In Table 3.1.2, a few calculated molecular multipole moments are given with the corresponding experimental values. The calculated values are given both resulting from the “*exact*” and *fitted* density to be able to compare the difference between them also, and are obtained in the TZ2P-newfit basis set with the Becke<sup>120</sup>-Perdew<sup>121</sup> exchange-correlation potential.

For benzene, the first non-vanishing multipole moment is (due to symmetry) the quadrupole moment. The values for the axes lying in the benzene plane are equal, and are minus half times the perpendicular value (also due to symmetry). The experimental value of  $-6.47$  a.u. for the perpendicular value  $\square_{zz}$  has been obtained by induced-birefringence experiments, and is reasonably well described by the DFT values of  $-5.55$  (“exact” density) and  $-5.59$  a.u. (*fitted* density), if the newfit set is used. Benzene is a good example why the standard fit sets are not always adequate for use with the new charge analysis presented here. In the standard fit set, the  $\square_{zz}$  value from the “exact” density doesn’t change much compared to the newfit set ( $-5.60$  a.u.); the value from the *fitted* density on the other hand differs considerably ( $-9.19$  a.u.). The dipole moment of carbon monoxide has in the past served as a test case for assessing the quality of wavefunction based methods, as it was shown that for instance Hartree-Fock predicts a dipole moment that points in the wrong direction. I.e., the negative end of the molecule was predicted on the side of the oxygen, while in reality it should be on the side of the carbon. In fact, the methods where electron correlation is included explicitly (like for instance MP2, MRSDCI or CASSCF) give a proper qualitative picture. As can be seen from Table 3.1.2, the same is true for the Becke-Perdew xc-potential. In fact, in a recent paper by de Proft et al.<sup>182</sup>, several other DFT potentials were tested and all were shown to give a proper description as well. The same agreement between calculated and experimental values is found for the quadrupole moments of carbondisulfide, fluorine, hydrogen, hexafluorobenzene, nitrogen and oxygen. Hexafluorobenzene is another striking example of the need for improved fit sets, as the value for its  $\square_{zz}$  value from the *fitted* density of the standard TZ2P fit set is  $2.81$  a.u., while the value from the “exact” density is  $6.08$  a.u. in this standard basis set. Although the difference between the “exact” and *fitted* value is a bit larger than for benzene ( $0.43$  a.u., see Table 3.1.2) with the new fit set, it is still a considerable improvement. The calculated values for the dipole and quadrupole moments of hydrogenfluoride and water are again in good agreement with the experimental values.

### Charges

The MDC-q charges of some molecules out of the set of 31 are given in Table 3.1.3. For a few molecules, point charges in non-atomic (dummy) positions have been used as described earlier in this section. In all cases, the molecular multipole moments up to the quadrupole moment have been represented (*by construction*) by these charges. Benzene and hexafluorobenzene show the predicted trend; in benzene, the electrons are more likely to be found on the carbons, while in hexafluorobenzene, they are attracted by the more electronegative fluoride atoms, just as would be expected from the change in sign of the  $\square_{zz}$  value when going from benzene to hexafluorobenzene. There are also some interesting features when comparing the diatomic molecules. Where in the case of chlorine, fluorine and hydrogen, the charge is mainly found in-between the two atoms (as indicated by the charge on the dummy point XX), for carbon monoxide, nitrogen and oxygen, it is found mainly on the atoms. This is of course fully consistent with the normal picture of chemical bonding for the first part (where the charge is mainly found in-between the atoms), while the cases of CO, N<sub>2</sub> and O<sub>2</sub> can be explained by the lone pair electrons that are found on the atoms. Although from the MDC-q charges (including a dummy point) for carbon monoxide, it seems that the more negative end is found near the oxygen, the MDC-d charges do give the correct representation. And as the molecular multipole moments are exactly reproduced up to the quadrupole by the MDC-q charges, also they give a proper description for the dipole moment.

For a molecule with a non-singlet ground state (like for instance oxygen), the charge density can be obtained for both spins separately. E.g., for oxygen with two alpha electrons in excess (triplet state), one obtains a different charge density for the alpha and beta electrons. Resulting from these are also

two sets of *atomic* multipoles; one for the alpha electrons and one for the beta electrons. From these we can extract *alpha* and *beta* Multipole Derived Charges, which when summed result in the atomic charge again. However, the difference between the two densities is the spin density, and as such, the difference between the two sets of charges, the spin density charges. For oxygen, the total charge on the atoms is small (-0.049), but the spin density charge large (-1.007), which is fully consistent with the triplet state with one alpha electron located on one atom, and one on the other.

TABLE 3.1.3 MDC-Q CHARGES FOR SOME OF THE MOLECULES OF THE SET OF 31

Molecule	Charges	Molecule	Charges
<i>benzene</i>		<i>hydrogen</i>	
qC	-0.123	qH	0.610
qH	0.123	qXX	-1.220
<i>ethylene</i>		<i>hexafluorobenzene</i>	
qC	-0.250	qC	0.094
qH	0.125	qF	-0.094
<i>methylcyanide</i>		<i>nitrogen</i>	
qC-Me	0.692	qN	-0.452
qH	-0.157	qXX	0.904
qC	0.203	<i>ammonia</i>	
qN	-0.425	qN	-0.444
<i>chlorine</i>		qH	0.148
qCl	0.367	<i>oxygen</i>	
qXX	-0.734	qO	-0.049
<i>carbondioxide</i>		qXX	0.098
qC	0.574	<i>water</i>	
qO	-0.287	qO	0.458
<i>carbonmonoxide</i>		qH	0.892
qC	-0.537	qXX	-1.121
qO	-0.852	<i>water dimer</i>	
qXX	1.389	qH-donor	0.902
<i>carbendisulfide</i>		qXX-donor	-1.162
qC	-0.351	qO-donor	0.264
qS	0.185	qH-donated	0.699
<i>fluorine</i>		qXX-donated	-0.748
qF	0.186	qO-acceptor	0.560
qXX	-0.371	qH-acceptor	0.930
		qXX-acceptor	-1.188

For carbondioxide and carbendisulfide, the sign of the quadrupole moment is reversed; i.e., whereas  $Q_{zz}$  is negative for carbondioxide, for carbendisulfide it is positive. This is reflected by the charge of the carbon atom, which is positive for the former and negative for the latter. The larger electronegativity of oxygen over sulphur plays a decisive role also here. The molecular dipole and quadrupole moment of water can not be represented exactly by charges on the atoms only; putting charges on dummy positions on both bond midway-points is however sufficient to do this. This results in rather large charges in the dummy positions, which are again indicative for the chemical bonding taking place. For the *hydrogen bonded* water dimer, depicted in Figure 3.1.1 with the donor on the left

and the acceptor on the right, the point charges in the dummy positions were also included. This enables a comparison with the values for water itself. Not surprisingly, the charges for the hydrogen (and corresponding dummy point charges) not involved in the hydrogen bond do not change significantly. The charge on the hydrogen involved in the hydrogen bond (the so-called *donated* hydrogen) changes dramatically. Whereas the total of the hydrogen and dummy charge is around  $-0.23$  in water, for the *donated* hydrogen this drops down to  $-0.05$ ; the difference is mainly flowing to the oxygen of the donor molecule, whose positive charge drops down by almost the same amount.

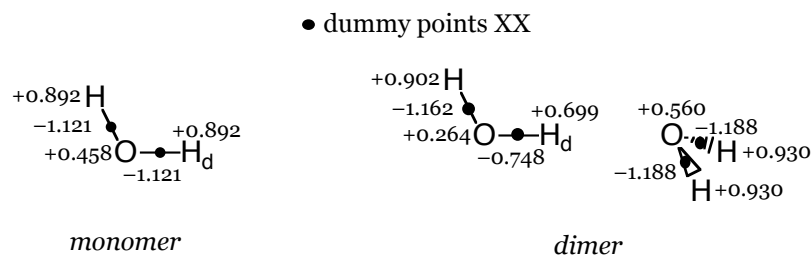


FIGURE 3.1.1 WATER MONOMER AND DIMER WITH MDC-Q CHARGES

The charges for the amino acid residue models *in vacuo* we have considered are given in Table 3.1.4 for the amino acids that are found in the active site of wildtype azurin, to be able to compare them with the values found within the active site (Section 6.1).

TABLE 3.1.4 AMINO ACID RESIDUE CHARGES FOR CYS, GLY, HIS AND MET

Atom	Charge	Atom	Charge	Atom	Charge	Atom	Charge
<i>cysteine</i>		<i>glycine</i>		<i>histidine</i>		<i>methionine</i>	
N	-1.028	N	-0.798	N	-0.882	N	-0.882
H	0.299	H	0.237	H	0.287	H	0.332
H	0.350	H	0.232	H	0.291	H	0.280
CA	0.461	CA	1.076	CA	0.505	CA	0.361
HA	-0.139	HA2	-0.249	HA	-0.087	HA	-0.103
CB	0.398	HA3	-0.247	CB	0.440	CB	0.496
HB2	-0.106	C	0.316	HB2	-0.109	HB2	-0.113
HB3	-0.117	O	-0.517	HB2	-0.178	HB3	-0.203
SG	-0.886	H	-0.051	CG	0.062	CG	0.176
C	0.319			ND1	-0.613	HG2	-0.033
O	-0.536			CE1	0.382	HG3	-0.057
H	-0.014			HE1	0.081	SD	-0.298
				NE2	-0.537	CE	-0.142
				HE2	0.381	HE1	0.079
				CD2	-0.035	HE2	0.104
				HD2	0.132	HE3	0.089
				C	0.390	C	0.367
				O	-0.471	O	-0.434
				H	-0.039	H	-0.021

## Conclusions

The charge analysis presented here gives an accurate description of the charge distribution when obtained with an appropriate set of fit functions as we have introduced in this paper. There are a few advantages of the method: the atomic multipoles that are reconstructed exactly by our Multipole Derived Charges, give by construction the best fit to the electrostatic potential around the molecule. Furthermore, with the appropriate set of fit functions the molecular multipole moments from the fitted density can be made arbitrarily close to the ones from the “exact” density. Our Multipole Derived Charges reproduce the fitted multipoles exactly, except in few cases (like diatomics) when there are not enough degrees of freedom to reproduce all multipoles. Usually it is sufficient to add one extra point, e.g. in the bond middle point of a diatomic molecule.

The molecular multipole moments, both from the fitted and the “exact” density, have a small deviation from experimental values. Since the MDC charges reproduce the fitted values exactly, the MDC charges give an accurate representation of the charge distribution in the molecule and result in a good electrostatic potential.



## IntraFF force constants

Obtaining reliable force constants for use in classical Molecular Dynamics simulations

Over the last decades, many investigations have focused on proteins containing metal atoms, both from an experimental and theoretical point of view (see Section 1.2). The theoretical studies comprise mainly classical Molecular Dynamics<sup>160</sup> (MD) studies, but in the last years it has become possible to treat (reasonable models of) the active sites of these proteins accurately by quantum chemical methods<sup>1</sup>. However, at the present time, it is still not possible to treat the complete protein with quantum chemical (QC) methods nor to perform reliable MD simulations (of at least 1 ns) with (part of) the active site treated at the QC level. For the time being, one therefore has to treat the active site completely classically. This raises a problem, because the interactions of metal atoms are difficult to generalize in terms of a simplified classical force field<sup>1</sup>. We resolve this issue by performing a QC calculation prior to the MD simulation to obtain force field parameters that can be used subsequently in the simulation and which are typical for the particular system under study. The necessary parameters can be divided into two groups: *bonded* and *non-bonded*. The latter consist of Lennard-Jones parameters (which have already been parameterized in the GROMOS96 force field<sup>127</sup>) and atomic fractional charges. The charges can be obtained accurately with Density Functional Theory calculations by using the Multipole Derived Charge analysis<sup>183</sup>. We are therefore left with obtaining parameters for *bonded* interactions.

The most commonly used way to treat these bonded interactions classically is by using a harmonic potential for bonds and bond angles:

$$U(C) = \frac{1}{2} K_C (C - C_{eq})^2 \quad (1)$$

where  $C_{eq}$  is the equilibrium value of the coordinate (bond/angle), and  $K_C$  a force constant typical for that particular bond/angle. A small force constant implies great flexibility in the coordinate while a large value is found for stiff bonds/angles. For dihedral angles  $\phi$  one normally takes a periodic function:

$$U(\phi) = K_\phi \left( 1 + \cos(n\phi - \phi_{shift}) \right) = K_\phi \left( 1 + \cos\left(n\left(\phi - \phi_{eq}\right) - \pi\right) \right) \quad (2)$$
$$\phi_{shift} = n\phi_{eq} + \pi$$

where  $n$  stands for the periodicity of the dihedral angle. Note that GROMOS96<sup>127</sup> uses a quartic potential (see **Appendix 3.2**) for bonds, which was introduced only to avoid the square root operation in the calculation. The deviations from the harmonic potential however show up only at high energies (i.e. at large deviations from the equilibrium position) that are normally never reached in simulations. The IntraFF method can also be used to obtain the force constant for this quartic potential.

Seminario proposed a method (FUERZA)<sup>184</sup> to get the force constants (FC's) directly from the Hessian matrix (a matrix of second derivatives of the energy with respect to atomic coordinates) as obtained in QC calculations. The method is invariant for the choice of internal coordinates, but the particular formulas he used for getting the force constants are not related directly to the classical



energy expressions. They result for instance in low values for bending constants, which is corrected for by adding extra *nonphysical* bonds. In his definition, two atoms are bonded if the energy increases when one of the two is moved in any direction, no matter how far apart the atoms are. For instance, the hydrogen atoms in water are bonded with a force constant of 31 kcal/mol/Å<sup>2</sup>. Although it is true that these hydrogens are “bonded”, the interaction between them has already been accounted for by the angle (bending) interaction.

In this study we report a new method (IntraFF) for obtaining force constants from QC calculations, that is based on ideas by Seminario<sup>184</sup>, but with formulas taken from the classical energy expressions as used in the force fields, in which the total energy is a sum of contributions from separate internal coordinates (like bonds, angles and dihedrals). Each of these contributions consists of a force constant  $K_C$  and a  $F_C$  part:

$$U_{total} = \sum_c U_c = \sum_c K_c F_c \quad (3)$$

The  $F_C$  part depends on the coordinates of **only** those atoms that are involved in the internal coordinate  $C$ . Likewise, the Hessian ( $3N \times 3N$  matrix for a system of  $N$  atoms containing the second derivatives of the energy with respect to the atomic Cartesian coordinates) is also a sum of contributions from the separate internal coordinates. The force constants  $K_C$  are obtained from the quantum chemical Hessian; they could be obtained by fitting the Hessian from the force field (FF) to the quantum chemical Hessian, but by doing so one already has to make a choice for the coordinates to be included in the force field, and the force constant values will depend on this choice. Instead, in the IntraFF method a direction vector is chosen for each internal coordinate  $C$ , which is characteristic for the coordinate; if the atoms involved in a coordinate are displaced by an infinitesimal amount along the direction vector of the coordinate, it will lead to a force working on the atoms, which is determined by the relevant part of the (QC or FF) Hessian. By minimizing the difference between the QC and FF infinitesimal forces, an expression is obtained in which the force constants are uniquely determined (see eq. (14)).

Therefore, just like the FUERZA scheme of Seminario<sup>184</sup>, our method is invariant for the choice of coordinates used. The equations for harmonic potentials applied to molecules in their equilibrium structure will be derived first. Then, we will discuss the specific details for the direction vectors needed for these equations, and the harmonic equations will be extended for molecules in a non-equilibrium structure. Subsequently, the equations for an anharmonic potential are presented, followed by the a normal mode analysis part and finally the results for some molecules.

A final subject that should be addressed here concerns the nomenclature of the different coordinate systems that are in use. Although the term *internal coordinates* has been used for many different kinds of coordinate sets, we use it for the *molecular* internal coordinates such as bonds, angles and dihedral angles. The coordinate system consisting of the normal mode vectors will be referred to as *normal mode coordinates*. Together with the *Cartesian coordinates* these sets of coordinates are complementary and sufficient.

## Harmonic potentials for molecules in equilibrium geometries

First, we will have to make the connection between the quantum chemical description and the classical force field description. This can be achieved if we expand the energy  $U$  to second order in a Taylor series about the atomic coordinates  $r$ . For this we need a direction vector  $d = \delta r$  which has a special

meaning in our analysis (more details will be given later on); it will be shown that it should be representative for a certain *internal* coordinate. For the moment however, we will keep the discussion general and let the vector be arbitrary. We then get the following expression for the Taylor series (up to second order):

$$U(r + d) = U(r) + d^T G(r) + \frac{1}{2} d^T H(r) d \quad (4)$$

where  $G(r)$  stands for the gradient and  $H(r)$  for the Hessian, with for atoms  $i, j$  the elements:

$$\begin{aligned} G_{i,p} &= \frac{U(r)}{r_{i,p}} \\ H_{i,p;j,q} &= \frac{\partial^2 U(r)}{r_{i,p} r_{j,q}} \\ & \quad r_{i,p} \quad \{x_i, y_i, z_i\} \end{aligned} \quad (5)$$

By definition, an equilibrium geometry is characterized by a zero gradient. Thus, the Taylor expansion leaves us with an expression for the change in energy that contains only a quadratic term (just like the classical harmonic expression [eq. (1)]):

$$U(r_{eq} + d) - U(r_{eq}) = \frac{1}{2} d^T H(r_{eq}) d \quad (6)$$

In principle, we could try and fit the Hessian matrix of the classical force field to the quantum chemical Hessian matrix. However, our method would then not be invariant to the choice of internal coordinates any longer, since we would have to choose which internal coordinates are used in the fit.

It would be more convenient to obtain the force constant for a given *internal* coordinate as if it were an independent parameter, like it is used also in most of the popular force fields. This can be achieved by looking at the change in gradient if a certain *internal* coordinate is changed by an infinitesimal amount, as proposed by Seminario<sup>184</sup>. For this, we expand not the energy but the gradient in a Taylor series, and obtain a formula for the gradient change  $\delta G$  (called *infinitesimal* gradient) that is linear in both the Hessian and the direction vector:

$$G(r + d) - G(r) = \delta G = H(r) d_c \quad (7)$$

This expression is valid for both the classical (FF) and the quantum chemical (QC) Hessian, and results in the classical ( $\delta G_{CL}$ ) and quantum chemical ( $\delta G_{QC}$ ) *infinitesimal* gradients. At this point, we have used the direction vector  $d_c$  with its special meaning as will be discussed later on, since it should be representative for the internal coordinate we are interested in. We will delay the discussion about the form of this direction vector until later, since its specific details are not important for formulating the analysis, as long as the vector is representative for a certain *internal* coordinate. Now, we want to find a value for the force constant such that the difference between the FF and the QC *infinitesimal* gradient is minimized. This is achieved by using a least squares fit of the difference between the *infinitesimal* gradients:

$$G = \frac{1}{2} (\delta G_{CL} - \delta G_{QC})^T (\delta G_{CL} - \delta G_{QC}) \quad (8)$$

The best value for the force constant  $K_C$  is then obtained if the derivative of this least squares function with respect to the force constant is zero:

$$\frac{G}{K_C} = 0 \quad (\delta G_{CL} - \delta G_{QC})^T \frac{[\delta G_{CL} - \delta G_{QC}]}{K_C} = 0 \quad (\delta G_{CL} - \delta G_{QC})^T \frac{[\delta G_{CL}]}{K_C} = 0 \quad (9)$$

The last part results from the fact that only the classical Hessian (and therefore the classical *infinitesimal* gradient) depends on the force constants. The derivative of this *infinitesimal* classical gradient with respect to  $K_C$  is obtained from the derivative of the classical Hessian with respect to the force constant. I.e., substituting the expression for the infinitesimal gradient (eq. (7)) into the last part of eq. (9), and taking the direction vector  $d_C$  outside the derivative expression, we obtain:

$$(\delta G_{CL} - \delta G_{QC})^T \frac{[H_{CL}(r)d_C]}{K_C} = 0 \quad (\delta G_{CL} - \delta G_{QC})^T \frac{H_{CL}(r)}{K_C} d_C = 0 \quad (10)$$

Now, we have to establish how the classical Hessian depends on the force constant  $K_C$ . In the harmonic approximation, the classical Hessian has the following form:

$$H_{r_{i,p}, r_{j,q}}^{CL} = \sum_C K_C (C - C_{eq}) \frac{^2 C}{r_{i,p} r_{j,q}} + \frac{C}{r_{i,p}} \frac{C}{r_{j,q}} = \sum_C K_C H_C^* \quad (11)$$

$r_{i,p} \quad \{x_i, y_i, z_i\}$

It is a sum over products of the force constants  $K_C$  of the internal coordinates times a matrix that depends only on the coordinates of the atoms involved in that internal coordinate. The derivative of the total Hessian with respect to the force constant  $K_C$  of one coordinate is then **independent of the other coordinates**, and independent of  $K_C$ :

$$\frac{H_{CL}}{K_C} = H_C^* \quad (12)$$

After putting this result into eq. (10), the only term in it that still depends on the value of  $K_C$  is the *infinitesimal* gradient  $\delta G_{CL}$ . We make it more explicit by combining the outcome of eq. (11) with eq. (7):

$$\delta G_{CL} = H_{CL} d_C = \{K_C H_C^*\} d_C = K_C (H_C^* d_C) = K_C V_C \quad (13)$$

It results in a multiplication of the scalar  $K_C$  with a vector  $V_C$ , which can be combined with eq. (10) to finally get an expression for the force constant  $K_C$ :

$$\begin{aligned} (\delta G_{CL} - \delta G_{QC})^T [H_C^* d_C] &= 0 \quad \delta G_{CL}^T [H_C^* d_C] = \delta G_{QC}^T [H_C^* d_C] \\ K_C V_C^T [H_C^* d_C] &= \delta G_{QC}^T [H_C^* d_C] \quad K_C = \frac{\delta G_{QC}^T [H_C^* d_C]}{V_C^T [H_C^* d_C]} = \frac{[H_{QC}^* d_C]^T [H_C^* d_C]}{[H_C^* d_C]^T [H_C^* d_C]} \end{aligned} \quad (14)$$

Although we have used the harmonic potential as an example here, this expression can in fact be used for any potential for which the Hessian can be written as a product as in eq. (11). Therefore, eq. (14) is also valid for the quartic GROMOS96<sup>127</sup> potential as well as the cosine function that is normally used for proper dihedrals (see **Appendix 3.2** for the expressions of the classical Hessian of the potentials described in this paper).

### Direction vectors

Before going on to the expressions for getting force constants of molecules in a non-equilibrium position, one issue should be addressed. It involves the direction vectors (of unit length; in principle we can take a vector of arbitrary length, since it is involved in both the numerator and denominator of eq. (14), but for simplicity we take this direction vector as a unit vector) belonging to the *internal* coordinates for which we want to obtain force constants and/or equilibrium values. As shown in the harmonic potential part, the difference between the classical and quantum chemical *infinitesimal* gradients in the direction of this vector is minimized (see eq. (8)); but no specific description has been given yet what this vector looks like. The only thing stated so far, is that the direction vector  $d_C$  should be representative for the *internal* coordinate we are interested in. In this part, we shall deal with this issue.

As the interactions for a given *internal* coordinate are expressed in terms of a scalar value, the issue does not arise when using the coordinate in a classical force field. Hence, no clearcut formulations are readily available for it. We will have to make a choice for the direction vector based on the positions of the atoms involved in the *internal* coordinate. The most simple approach is to take the vector which will induce the largest increase in the value of the *internal* coordinate, which is given by the gradient of the coordinate. For example, for a bond between two (homonuclear) atoms, the direction vector is given in Figure 3.2.1 as  $\delta r$ :

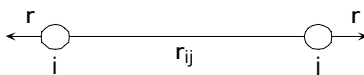


FIGURE 3.2.1. SCHEMATIC REPRESENTATION OF BOND DIRECTION VECTOR

Only in the case of a homonuclear dimer is this the one and only choice. For instance, if we take a heteronuclear dimer, we have to take the different atomic masses into consideration by using some kind of weight factor. But in what way, is not yet clear. We could argue that since heavy atoms move less than light atoms, the latter should have a higher weight; based on the expression for the kinetic energy, it would lead to a weight that is equal to the inverse square root of the atomic mass. But by looking at the force, we could argue that the weight should be equal to the inverse of the mass. Moreover, when using the latter choice for the direction vector, neither the center of mass nor the moments of inertia of the molecule change. Therefore, this choice is the preferred one, leading to the direction vector of a given *internal* coordinate  $C$  as:

$$d_{C,i} = w_i (C)_i \quad (15)$$

where the weight for atom  $i$  is given as:

$$w_i = \frac{1}{M_i} \quad (16)$$

In order to check that this is indeed the most appropriate choice for the direction vector, we have tested a few other choices; we looked at the gradients, vibrational frequencies and fit residues (eq. (8)) for a number of molecules. For “normal” bonds, i.e. not involving metal atoms, the choice given in eq. (15) gives indeed the best agreement between quantum chemical and classical properties (gradients, vibrational frequencies, fit residues). However, as will be shown in the **Results** part, when looking at bonds involving metal atoms, more appropriate direction vectors can be used where only the metal atom is allowed to move.

### Harmonic potentials for molecules in non-equilibrium geometries

In the harmonic potential part, we discussed how to obtain harmonic force constants for molecules in their equilibrium positions. However, most of computational chemistry is being done on molecules that are not *exactly* in the equilibrium geometry due to numerical accuracy. It would be most helpful to be able to correct the force constants as well as the equilibrium values of the bond lengths/angles for this non-equilibrium character. Within the harmonic approximation this is easily achieved by first getting the force constant as if the molecule were in an equilibrium geometry. Since the second derivative of the harmonic potential is constant and equal to the force constant (in 1D; in 3D, the trace of the Hessian is constant), the value of it does not change if the molecule is put into a non-equilibrium position. The only change occurs for the gradient, which measures how far the molecule is outside its equilibrium position:

$$G(C_{now}) = K_C (C_{now} - C_{eq}) \quad (17)$$

Since the gradient is known from quantum chemistry, the real equilibrium value can be obtained from this expression since all other quantities are known:

$$C_{eq} = C_{now} - \frac{G(C_{now})}{K_C} \quad (18)$$

In three-dimensional notation, this would result in the following expression:

$$C_{eq} = C_{now} - \frac{\mathbf{G}(r_{now})^T C}{K_C (C^T C)^{1/2}} \quad (19)$$

Note that if the gradient is zero, the equilibrium value is simply the current one, like already mentioned earlier.

For a proper dihedral, we do not use a harmonic potential but a cosine function. The equilibrium value can then be obtained by following the same strategy, resulting in:

$$\phi_{eq} = \phi_{now} + \frac{1}{n} \arcsin \frac{-1}{nK_\phi} \frac{G(r_{now})^T \phi}{(\phi^T \phi)^{1/2}} \quad (20)$$

### Anharmonic Frost potential

Until now, we have concerned ourselves with harmonic potentials, which are valid only if one is reasonably close to the equilibrium geometry of the molecule (see Figure 3.2.2). However, sometimes one might want to use an anharmonic potential that is able to give a proper description over a larger bond value range, for instance in the case of the dissociation reaction between any atoms or molecules. The most widely known anharmonic potential is due to Morse<sup>130</sup>. It uses the equilibrium distance  $R_{eq}$ , dissociation energy  $D_e$  and anharmonicity constant  $\beta$ :

$$U_{Morse} = D_e \left( 1 - e^{-\beta(R - R_{eq})} \right)^2 \quad (21)$$

Disadvantageous for this equation is the fact that it is a three parameter equation, while in the strictest sense, we only have two variables to optimize the parameters (first and second derivative: gradient and Hessian).

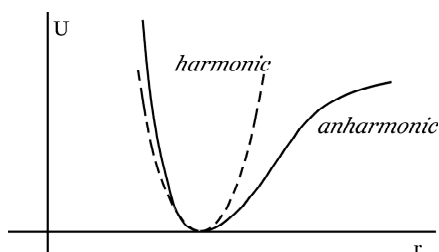


FIGURE 3.2.2. SCHEMATIC COMPARISON OF ANHARMONIC FROST AND HARMONIC POTENTIAL

Varshni<sup>185</sup> has examined several potentials with differing number of parameters, ranging from two to five. One of them is the two parameter Frost potential<sup>131,132</sup>, which has been constructed on semi-theoretical grounds. It is split up in a nuclear and an electronic part, just like normally done in quantum chemistry:

$$U(R) = \frac{N_1 N_2}{R} + U_e(R) \quad (22)$$

$N_i$  stands here for the nuclear charge of atom  $i$ , and  $U_e$  for the electronic part of the potential. Frost and Musulin then defined some theoretical criteria that should be satisfied by a given potential (they used also two less clear ones, which we will not show here):

- 1- the potential energy is the sum of two parts (i.e. nuclear and electronic)
- 2-  $U$  becomes infinite as  $R$  approaches zero
- 3-  $U_e$  is finite at  $R = 0$
- 4-  $U_e \sim -N_1 N_2 / R$  for large  $R$  (due to choice of  $U = 0$  at infinite  $R$ )
- 5-  $U$  must be capable of going through a minimum as  $R$  varies

A simple function for the electronic potential that satisfies these criteria, is given by:

$$U_e = -\frac{N_1 N_2}{R} (1 - e^{-aR}) - b e^{-aR} \quad (23)$$

where  $a, b$  are adjustable parameters. They arrived at the particular form for this function by considering possible products of an exponential with powers of  $R$ , which was the most probable form because the overlap, Coulomb and exchange integrals of quantum-chemical theory are of this type when using Slater type orbitals. The sum of the electronic and nuclear potential energy is then given by:

$$U = e^{-aR} \frac{N_1 N_2}{R} - b \quad (24)$$

The function is plotted with parameters for the hydrogen molecule (from the original article<sup>131</sup>) in Figure 3.2.2. Examining several properties of 23 diatomic molecules, Varshni concluded that the Frost and Morse potential perform equally well<sup>185</sup>.

The question remains however, how to obtain values for the  $a, b$  parameters. For that, we need not only the potential energy, but also the gradient and Hessian (in 1D):

$$\begin{aligned} U &= e^{-aR} \frac{N_1 N_2}{R} - b \\ G &= e^{-aR} \left( ab - \frac{aN_1 N_2}{R} - \frac{N_1 N_2}{R^2} \right) \\ H &= e^{-aR} \left( \frac{2N_1 N_2}{R^3} + \frac{2aN_1 N_2}{R^2} + \frac{a^2 N_1 N_2}{R} - a^2 b \right) \end{aligned} \quad (25)$$

For the equilibrium bond length the gradient is zero; therefore we have an expression for the  $b$  parameter:

$$ab - \frac{aN_1 N_2}{R_{eq}} - \frac{N_1 N_2}{R_{eq}^2} = 0 \quad b = \frac{N_1 N_2}{R_{eq}} + \frac{N_1 N_2}{a R_{eq}^2} \quad (26)$$

We can use this in the equation for the Hessian, which results in an expression for it that depends only on  $a$  and  $R$ :

$$H = N_1 N_2 e^{-aR} \left( \frac{2}{R^3} + \frac{2a}{R^2} + \frac{a^2}{R} - \frac{a^2}{R_{eq}} - \frac{a}{R_{eq}^2} \right) = K \quad (27)$$

However, we also have the value of the Hessian in this point (the force constant  $K$  within the harmonic approximation). Therefore, we can get the parameter  $a$  by minimizing the difference between the actual 2<sup>nd</sup> derivative  $K$  and the Frost 2<sup>nd</sup> derivative  $K$  :

$$a = \frac{1}{2} (K - K)^2 \quad (28)$$

Note, that in order to do this we have to choose a trial equilibrium bond length (which we will optimize later). From eq. (28) we get the value for the parameter  $a$ , which we can use in eq. (26) for the parameter  $b$  at a certain “equilibrium” bond length. With the  $a, b$  parameters, we obtain a value for the gradient within the Frost approximation. If it is equal to the actual quantum chemical gradient, the “trial” equilibrium bond length is correct, else we use the gradient difference to get a better guess of the equilibrium bond length, and repeat this procedure until we have found the correct one (within the Frost approximation).

With the optimized parameters  $a, b$  and the equilibrium bond length, the dissociation energy  $D_e$  is simply obtained by filling in these values in eq. (24). We then get:

$$D_e = -e^{-aR_{eq}} \frac{N_1 N_2}{R_{eq}} - b \quad (29)$$

Note that the minus sign has been added since the dissociation energy is normally defined as a positive quantity, i.e. the energy needed to dissociate a molecule and not as a negative interaction energy as in equation (24).

### Normal mode analysis

An important check on the accuracy of the force field parameters from any method is the ability of them to reproduce the vibrational frequencies (be it from experiments or quantum chemical calculations). They can be obtained from the calculated Hessian in Cartesian coordinates, by transforming it into the mass-weighted Hessian matrix  $H^{mw}$ <sup>186</sup>:

$$H_{ij}^{mw} = \frac{H_{ij}}{\sqrt{M_i M_j}} \quad (30)$$

After diagonalization of this matrix, the eigenvalues  $\varepsilon_i$  can be transformed easily into the vibrational frequencies  $\nu_i$ :

$$\nu_i = \frac{1}{2\pi} \sqrt{\varepsilon_i} \quad (31)$$

This procedure works for any Hessian matrix (both classical and quantum chemical), and therefore we can use it to check the reliability of the classical force field.

One important issue involved in this procedure is the coordinates involved in the normal mode analysis. For a molecule consisting of  $N$  atoms, there are  $3N - 6$  normal modes and vibrational frequencies ( $3N - 5$  for a linear molecule). So far, we have not been concerned about the number of classical *internal coordinates*, but it is obvious that they should be chosen such that the total number is equal to the number of normal modes. However, this is a choice of the person who wants to use the force constants for a classical simulation, and as such does not concern us here in the method development. Normally, the choice for which *internal coordinates* to use is easily made, like in the case of water. One would normally take two O-H bonds and one H-O-H angle, or alternatively three bonds: two O-H and one H-H. Moreover, it is even possible to combine several *internal coordinates* together in one *combined internal coordinate* as are standard in for instance AMBER95 with a few



dihedral angles. Although the total number of actual *internal coordinates* used is larger than the number of normal modes, this is not true when the *combined internal coordinates* are counted.

## Computational details

The method described in this paper is generally applicable to any quantum chemical method able to supply a gradient and Hessian for a certain molecule in a given geometry. In this paper we will report results for some molecules where the quantum chemical data were obtained with Density Functional Theory (DFT)<sup>1</sup> calculations in the ADF program<sup>117,187</sup>. Although the results in this paper are used to test the method, and not the accuracy of DFT, we would like to obtain calculated values that are at least reasonable, to be able to compare the force field parameters obtained with standard force fields like AMBER<sup>126</sup>/GROMOS<sup>127,176</sup>. We therefore performed the DFT calculations in a TZP basis set using the LDA (Local Density Approximation)<sup>1</sup> exchange-correlation potential. In another study, we have investigated the accuracy of several exchange-correlation potentials in a number of basis sets for a set of 19 small molecules. The same set had been used previously by another group<sup>188</sup> to test the accuracy of several post-Hartree Fock methods in combination with different (correlation consistent) basis sets. This enabled us to compare our results not only to experimental data, but also to high-level *ab initio* calculations. We found that (for the TZP basis set we used) the accuracy of the LDA exchange-correlation potential is equal to the normally more accurate but also more CPU expensive Becke<sup>120</sup>-Perdew (BP86<sup>121</sup>) or Perdew-Wang (PW91<sup>189</sup>) exchange-correlation potentials: the average absolute deviations of the bond lengths from experimental data are respectively 1.14 (LDA), 1.24 (BP86) and 1.09 (PW91) pm. Furthermore, it is known that DFT generally gives accurate frequencies at low computational cost<sup>1</sup>, which enables us to treat larger systems at a high computational level.

The geometries of the molecules were optimized until the gradient was smaller than  $1.0 \cdot 10^{-3}$ , a frequency run was performed (without symmetry to obtain the full Hessian matrix), and the force constants obtained. Until now, the IntraFF analysis for obtaining the force field parameters was performed in a standalone program after the ADF calculation. In the near future, it will be added to the ADF program.

## Results

### *Set of twelve molecules*

The set of molecules consists of: *ethylene, ethane, methane, carbon monoxide, hydrogen, hydrogenchloride, hydrogenfluoride, hydrogenperoxide, nitrogen, ammonia, benzene, CFCIBrH, dioxygen and water*. In all cases did the geometry optimization converge within five steps, which took at most one hour of CPU-time on an IBM/RS6000 per molecule. The optimized values for the *internal coordinates* of the molecules are given in Tables 3.2.1 and 3.2.2.

Probably the most commonly known experimental value for a bond angle known in computational chemistry is the one for the water molecule ( $104.45^\circ$ ). Our computed result of  $104.34^\circ$  is in very good agreement with this value, although it must be mentioned at the same time that the O-H bond is overestimated by some 1.6 pm. The same trend of overestimating bond lengths is seen for some other molecules, like hydrogen (2.0 pm) and ammonia (1.6 pm). On the other hand, the computed results for carbon monoxide, nitrogen and oxygen are in perfect agreement with experiment, and also the agreement for the HNH angle in ammonia is good. Generally speaking we are therefore reasonably close to the experimental geometries.

TABLE 3.2.1. FORCE CONSTANTS  $FC$  (KCAL/MOL/Å<sup>2</sup>) FOR OPTIMIZED BONDS  $R$  (Å), WITH FROST PARAMETERS  $a$  (BOHR<sup>-1</sup>),  $b$  (HARTREE) AND CORRESPONDING DISSOCIATION ENERGY  $D_e$  (eV)

Molecule	Atoms	$R$	$FC$	Amber <sup>126</sup>	$a$	$b$	$D_e$
C2H4	CC	1.324	1554	-	1.101	19.60	9.02
	CH	1.095	760	734	1.016	4.28	4.59
C2H6	CC	1.510	755	620	1.110	16.60	4.56
	CH	1.102	704	680	1.046	4.20	4.07
CH4	CH	1.098	728	680	1.035	4.24	4.28
CO	CO	1.131	2731	-	1.405	29.93	10.08
H2	HH	0.766	740	-	0.788	1.30	5.26
HCl	HCl	1.300	682	-	1.188	9.29	3.48
HF	HF	0.937	1302	-	1.435	7.09	4.29
H2O2	OO	1.443	644	-	1.548	29.02	2.21
	OH	0.981	1061	1106	1.334	6.06	4.01
N2	NN	1.098	3405	-	1.382	31.85	12.75
NH3	NH	1.024	948	868	1.178	5.20	4.42
O2	OO	1.218	1698	-	1.595	35.38	5.25
CFClBrH	CF	1.339	781	734	1.604	26.60	2.47
	CCl	1.755	457	-	1.354	37.61	2.09
	CBr	1.943	379	-	1.414	68.20	1.66
	CH	1.097	743	680	1.026	4.26	4.42
C6H6	CC	1.398	1073	938	1.133	18.34	6.43
	CH	1.090	785	734	1.029	4.20	4.20
H2O	OH	0.974	1118	1106	1.321	6.13	4.27

In order to get vibrational frequencies for the 12 molecules, we calculated the Hessian for each molecule through numerical differentiation of the analytical gradients, followed by a normal mode analysis. This procedure of numerical differentiation has the advantage that one can also obtain the derivative of the dipole moment with respect to the coordinates, and as such the IR intensities of the frequencies. The vibrational frequencies obtained from this procedure are given in Table 3.2.2. For the diatomic molecules, the values are 1582 cm<sup>-1</sup> (O<sub>2</sub>), 2167 cm<sup>-1</sup> (CO), 2395 cm<sup>-1</sup> (N<sub>2</sub>), 2866 cm<sup>-1</sup> (HCl), 4005 cm<sup>-1</sup> (HF), 4161 cm<sup>-1</sup> (H<sub>2</sub>), with intensities 0.0 (O<sub>2</sub>), 72.3 (CO), 0.0 (N<sub>2</sub>), 70.3 (HCl), 145.3 (HF), 0.0 (H<sub>2</sub>) km/mol, respectively. The frequencies are in good agreement with the experimental values<sup>190</sup> of 1580, 2133, 2359, 2989, 4138, 4359 cm<sup>-1</sup>. For water, the DFT frequencies are 1555 [86.3], 3696 [10.6] and 3798 cm<sup>-1</sup> [80.7 km/mol], with the intensities given between square brackets. The frequencies correspond more or less to one bending and two bond stretching modes. This is just a rough splitting of the *internal* coordinate systems, because in reality the normal modes do not correspond exactly to the internal coordinates, but are a mixture of them. For ammonia, we find three high frequency values (3381 [1.1], twice 3513 cm<sup>-1</sup> [10.6]), and three lower values: 968 [160.5], twice 1585 cm<sup>-1</sup> [0.6 km/mol]. A rough division would again lead to the high frequency modes corresponding to bond stretches, and the low values to bending modes.

TABLE 3.2.2. FORCE CONSTANTS  $FC$  (KCAL/MOL/RAD<sup>2</sup>) FOR OPTIMIZED ANGLES AND DIHEDRAL ANGLES  $A$  (°)<sup>a</sup>

<i>Angle</i>	<i>A</i>	<i>FC</i>	<i>Angle</i>	<i>A</i>	<i>FC</i>
<i>C2H4</i>			<i>CFClBrH</i>		
CCH	121.52	176.0	FCl	109.70	384.2
HCCH	0/180	7.1	FCBr	109.01	384.6
<i>C2H6</i>			FCH	110.27	125.3
CCH	111.60	155.8	ClCBr	111.83	555.3
HCCH	60 (± 120)	8.1	ClCH	109.05	112.9
<i>CH4</i>			BrCH	106.94	108.8
HCH	109.47	93.8	<i>C6H6</i>		
<i>H2O2</i>			CCC	120.00	735.6
OOH	100.36	121.1	CCH	120.00	178.7
HOOH	106.76	6.8	<i>H2O</i>		
<i>NH3</i>			HOH	104.34	89.2
HNH	106.65	89.3			

Methane has the following frequencies: three times 1249 cm<sup>-1</sup> [19.9], twice 1480 cm<sup>-1</sup> [0.0], 2951 cm<sup>-1</sup> [0.0] and twice 3080 cm<sup>-1</sup> [7.0 km/mol]. For hydrogenperoxide, the values are 431 [150.2], 956 [0.6], 1285 [108.0], 1362 [0.9], 3618 [62.5] and 3630 cm<sup>-1</sup> [24.1 km/mol]. The latter two would correspond to the OH stretches, and the first one to the HOOH torsion. The other three are some mixture of the OO stretch and the HOO bends. The bond stretches are in this case all lower than the corresponding values found in water. For ethylene, the frequencies are 798 [0.4], 924 [93.1], 937 [0.0], 1029 [0.0], 1184 [0.0], 1319 [0.0], 1390 [14.7], 1652 [0.0], 3038 [5.6], 3050 [0.0], 3117 [0.0] and 3141 cm<sup>-1</sup> [5.4 km/mol]. Finally, for ethane we find 304 [0.0], twice 787 [6.6], 1019 [0.0], twice 1159 [0.0], 1332 [4.7], 1349 [0.0], twice 1425 [0.0], twice 1425 [14.3], 2949 [0.0], 2949 [42.0], twice 3010 [0.0] and twice 3033 cm<sup>-1</sup> [29.2 km/mol].

The computed Hessian matrices for the molecules were used afterwards to get force constants for all possible internal coordinates of the molecules, which are given in Tables 3.2.1 and 3.2.2. Compared to the force constants of AMBER<sup>126</sup> ( $k_{OH}$  1106 kcal/mol/Å<sup>2</sup>) or GROMOS<sup>127</sup> ( $k_{OH}^{quartic}$  2.32·10<sup>7</sup> kJ/mol/nm<sup>4</sup>), our value for water of 1118 kcal/mol/Å<sup>2</sup> (or 2.46 kJ/mol/nm<sup>4</sup>) is rather similar. Our bending constant of 89 kcal/mol/rad<sup>2</sup> is close to the GROMOS value (corresponding to 92 kcal/mol/rad<sup>2</sup>), but rather far off from the AMBER value (200 kcal/mol/rad<sup>2</sup>), which is a surprising deviation on their part. In Table 3.2.3, we report the vibrational frequencies for the molecules from the force constants obtained here, that have on average a deviation of 13 cm<sup>-1</sup> from the DFT values. Here we can also see that our force constants performs well for water, which casts some more doubt on the high bending constant for water in the AMBER force field.

<sup>a</sup>The values of the force constants depend on how many dihedrals are taken into account in the force field; for ethane, it makes a factor 9 difference if all nine possible dihedrals are taken into account or only one; for the values reported here it is assumed that the dihedrals are all taken into account separately

TABLE 3.2.3. FREQUENCIES (CM<sup>-1</sup>) FROM QUANTUMCHEMICAL AND CLASSICAL HESSIANS

QC	FF	QC	FF	QC	FF	QC	FF	QC	FF
<i>ethylene</i>		<i>ethane</i>		<i>hydrogenchloride</i>		<i>nitrogen</i>		<i>benzene</i>	
798	668	304	931	2866	2866	2395	2395	409 <sup>2</sup>	462 <sup>2</sup>
924	854	787 <sup>2</sup>	887 <sup>2</sup>					603 <sup>2</sup>	885 <sup>2</sup>
937	988	1019	1229	<i>hydrogenfluoride</i>		<i>oxygen</i>		624	744
1029	1049	1159 <sup>2</sup>	1054 <sup>2</sup>	4005	4005	1582	1582	716	781
1184	1186	1332	1246					805 <sup>2</sup>	909 <sup>2</sup>
1319	1238	1349	1450	<i>hydrogenperoxide</i>		<i>water</i>		931 <sup>2</sup>	989 <sup>2</sup>
1390	1327	1424 <sup>2</sup>	1420 <sup>2</sup>	431	430	1555	1537	960	896
1652	1789	1425 <sup>2</sup>	1438 <sup>2</sup>	956	920	3696	3703	984	972
3038	3052	2948	2921	1285	1281	3798	3757	996	1182
3050	3085	2950	2931	1362	1343			1016 <sup>2</sup>	1054 <sup>2</sup>
3117	3160	3010 <sup>2</sup>	3026 <sup>2</sup>	3618	3634	<i>CFClBrH</i>		1107	1492
3141	3166	3033 <sup>2</sup>	3034 <sup>2</sup>	3630	3634	215	283	1132 <sup>2</sup>	1254 <sup>2</sup>
						305	386	1297	1696
<i>methane</i>		<i>carbonmonoxide</i>		<i>ammonia</i>		419	451	1324	1730
1249 <sup>3</sup>	1347 <sup>3</sup>	2167	2167	968	1185	641	778	1441 <sup>2</sup>	1644 <sup>2</sup>
1480 <sup>2</sup>	1508 <sup>2</sup>			1585 <sup>2</sup>	1593 <sup>2</sup>	761	946	1574 <sup>2</sup>	1909 <sup>2</sup>
2951	2919	<i>hydrogen</i>		3381	3382	1103	1088	3119	3166
3080 <sup>3</sup>	3085 <sup>3</sup>	4161	4161	3513 <sup>2</sup>	3483 <sup>2</sup>	1154	1328	3125 <sup>2</sup>	3169 <sup>2</sup>
						1256	1439	3137 <sup>2</sup>	3191 <sup>2</sup>
						3054	3078	3148	3210

2) Twice found ;

3) Three times found

For the bonds we report in Table 3.2.1 also the parameters  $a$  and  $b$  for the Frost potential as well as the dissociation energy resulting from these parameters. These energies are in reasonable agreement with experimental values<sup>190</sup> (typical deviation 0.2-0.3 eV, smallest deviation 0.01 eV, largest 1.6 eV).

## Conclusions

We have constructed a new method to obtain force field parameters from quantum chemical calculations for bonding interactions that can be used in molecular simulations by classical mechanics. It is invariant to the choice of internal coordinates, and can be used with any type of quantum chemical method. We have tested several options for direction vectors that are used to describe the internal coordinates, and found the one where the direction vector of maximum increase (or the gradient of the internal coordinate with respect to atomic coordinates) weighted inversely by the atomic masses to give the best performance. The heavy atoms are therefore less important for the force constant, which seems natural since they tend to move less also. The reference data come from Density Functional Theory (DFT) calculations using the LDA exchange-correlation potential in a triple zeta

valence basis set plus polarization functions (TZP), which is at a level sufficiently high to give results that are reasonably close to experimental data. From the computed DFT Hessians we have obtained the vibrational frequencies (and IR intensities) for a dozen molecules, and compared the frequencies from the force field with them. On average the difference between the two sets is 13 cm<sup>-1</sup>.

### Appendix 3.2. Classical force field expressions

#### Harmonic potentials

The energy for any internal coordinate  $C$  using a harmonic potential is given by the following energy expression:

$$U(C_{now}) = \frac{1}{2} K_C (C_{now} - C_{eq})^2 \quad (32)$$

The derivatives of this energy with respect to the atomic coordinates  $r_{p,i}$  are then obtained by applying the chain rule. First, the derivative with respect to the internal coordinate is taken, then the derivative of this internal coordinate with respect to atompair distances ( $R_{ij} = |r_{ij}| = |r_i - r_j|$ ) and finally the derivative of the atompair distance with respect to the atomic coordinates:

$${}_{i,p}U(C_{now}) = G_{i,p} = \frac{U(C_{now})}{C} \frac{C}{R_{ij}} \frac{R_{ij}}{r_{i,p}} = K_C (C_{now} - C_{eq}) \frac{C}{R_{ij}} \frac{R_{ij}}{r_{i,p}} \quad (33)$$

The second derivatives (Hessians) are obtained by the product rule, resulting in two contributions:

$$H_{i,p,j,q} = \frac{G_{i,p}}{r_{j,q}} = K_C \frac{C}{R_{ij}} \frac{R_{ij}}{r_{i,p}} \frac{R_{ij}}{r_{j,q}} + K_C (C_{now} - C_{eq}) \frac{2C}{R_{ij}^2} \frac{R_{ij}}{r_{i,p}} \frac{R_{ij}}{r_{j,q}} \quad (34)$$

The derivatives of a bond  $R$  with respect to the atomic coordinates are given by:

$$\begin{aligned} \frac{R_{ij}}{r_{i,p}} &= \frac{r_{ij,p}}{R_{ij}} \frac{r_{ij}}{r_i} \\ \frac{{}^2R_{ij}}{r_{i,p} r_{j,q}} &= (-1)^{1+\delta_{ij}} \frac{\delta_{pq}}{R_{ij}} - \frac{r_{ij,p} r_{ij,q}}{R_{ij}^3} \end{aligned} \quad (35)$$

For the dihedral angle  $\phi$ , we use the formulations of Blondel and Karplus<sup>191</sup>, who have shown that for any normal dihedral angle value (i.e. all cases except the one where three or more atoms lie on a straight line), the derivatives for the potential of  $\phi_{ijkl}$  can be obtained by using the following vectors:

$$\begin{aligned} V_F &= r_i - r_j & V_A &= V_F \times V_G \\ V_G &= r_j - r_k & V_B &= V_H \times V_G \\ V_H &= r_l - r_k \end{aligned} \quad (36)$$

The derivatives for the four atoms are then given by:

$$\begin{aligned}
\frac{\phi}{r_i} &= -\frac{|V_G|}{(V_A^T V_A)} V_A \\
\frac{\phi}{r_j} &= +\frac{|V_G|}{(V_A^T V_A)} V_A + \frac{(V_F^T V_G)}{(V_A^T V_A)|V_G|} V_A - \frac{(V_H^T V_G)}{(V_B^T V_B)|V_G|} V_B \\
\frac{\phi}{r_k} &= -\frac{|V_G|}{(V_B^T V_B)} V_B - \frac{(V_F^T V_G)}{(V_A^T V_A)|V_G|} V_A + \frac{(V_H^T V_G)}{(V_B^T V_B)|V_G|} V_B \\
\frac{\phi}{r_l} &= +\frac{|V_G|}{(V_B^T V_B)} V_B
\end{aligned} \tag{37}$$

From these equations, one can easily see why this procedure breaks down if three atoms lie on a straight line, since the inner product of either the  $V_A$ - or  $V_B$ -vector will be zero. However, in general such a situation is not likely to occur, since there are bending interactions preventing it from happening. The second derivatives of the dihedral angle with respect to the atomic coordinates are given by first taking the derivatives with respect to the  $V_F, V_G, V_H$ -vectors:

$$\begin{aligned}
\frac{^2\phi}{V_F^2} &= \frac{|V_G|}{(V_A^T V_A)^2} (V_A \quad V_G \times V_A + V_G \times V_A \quad V_A) \\
\frac{^2\phi}{V_H^2} &= \frac{-|V_G|}{(V_B^T V_B)^2} (V_B \quad V_G \times V_B + V_G \times V_B \quad V_B) \\
\frac{^2\phi}{V_F V_G} &= \frac{1}{|V_G|(V_A^T V_A)^2} \left( (V_G^T V_G)(V_A \times V_F) \quad V_A + (V_F^T V_G)V_A \quad (V_A \times V_G) \right) \\
\frac{^2\phi}{V_G V_H} &= \frac{-1}{|V_G|(V_B^T V_B)^2} \left( (V_G^T V_G)(V_B \times V_H) \quad V_B + (V_H^T V_G)V_B \quad (V_B \times V_G) \right) \\
\frac{^2\phi}{V_F V_H} &= 0
\end{aligned} \tag{38}$$

$$\begin{aligned}
\frac{^2\phi}{V_G^2} &= \frac{1}{2|V_G|^3(V_A^T V_A)} (V_G \times V_A \quad V_A + V_A \quad V_G \times V_A) \\
&+ \frac{(V_F^T V_G)}{|V_G|(V_A^T V_A)^2} (V_A \quad V_F \times V_A + V_F \times V_A \quad V_A) \\
&- \frac{1}{2|V_G|^3(V_B^T V_B)} (V_G \times V_B \quad V_B + V_B \quad V_G \times V_B) \\
&- \frac{(V_H^T V_G)}{|V_G|(V_B^T V_B)^2} (V_B \quad V_H \times V_B + V_H \times V_B \quad V_B)
\end{aligned} \tag{39}$$

where  $\otimes$  stands for the tensorproduct:

$$(A \otimes B)_{ij} = A_i B_j \tag{40}$$

Now, with the derivatives of the  $V_F, V_G, V_H$  -vectors with respect to the atomic coordinates:

$$\frac{V_F}{r_i} = 1; \quad \frac{V_F}{r_j} = -1; \quad \frac{V_G}{r_j} = 1; \quad \frac{V_G}{r_k} = -1; \quad \frac{V_H}{r_i} = 1; \quad \frac{V_H}{r_k} = -1 \quad (41)$$

we finally obtain the second derivatives of the dihedral angle as:

$$\frac{\partial^2 \phi}{r_i^2} = \sum_{p,q} \frac{\partial^2 \phi}{V_p V_q} \frac{V_p}{r_i} \frac{V_q}{r_i}; \quad p, q \in \{F, G, H\} \quad (42)$$

By following the same procedure for the angle, and by using the following vectors:

$$\begin{aligned} V_A &= r_i - r_j \\ V_B &= r_k - r_j \end{aligned} \quad V_G = V_A \times V_B \quad (43)$$

we obtain the derivative of the angle  $\theta$  with respect to the atomic coordinates as follows:

$$\begin{aligned} \frac{\theta}{V_A} &= \frac{-1}{(V_A^T V_A) |V_G|} (V_G \times V_A) & \frac{\theta}{r_i} &= \frac{\theta}{V_A} \\ \frac{\theta}{V_B} &= \frac{-1}{(V_B^T V_B) |V_G|} (V_B \times V_G) & \frac{\theta}{r_j} &= -\frac{\theta}{V_A} + \frac{\theta}{V_B} \\ & & \frac{\theta}{r_k} &= \frac{\theta}{V_B} \end{aligned} \quad (44)$$

The second derivative has not been worked out yet, when needed it is obtained by numerical differentiation of eq. (44).

#### Anharmonic Frost potential<sup>131,132</sup>

The energy expression for this anharmonic bond potential is for a bond  $R_{ij}$  between atoms  $i$  and  $j$  as:

$$U(R_{ij}) = e^{-aR_{ij}} \frac{N_i N_j}{R_{ij}} - b \quad (45)$$

The derivatives with respect to the distance  $R_{ij}$  are given by:

$$\begin{aligned} G(R_{ij}) &= e^{-aR_{ij}} \left( ab - \frac{aN_i N_j}{R_{ij}} - \frac{N_i N_j}{R_{ij}^2} \right) \\ H(R_{ij}) &= e^{-aR_{ij}} \left( \frac{2N_i N_j}{R_{ij}^3} + \frac{2aN_i N_j}{R_{ij}^2} + \frac{a^2 N_i N_j}{R_{ij}} - a^2 b \right) \end{aligned} \quad (46)$$



*Quartic GROMOS potential<sup>127</sup>*

In order to avoid the square root operation, GROMOS96 uses quartic potentials for bonds. The expressions for a bond  $R$  are:

$$\begin{aligned}
 U(R) &= \frac{1}{4} K^{quartic} (R^2 - R_{eq}^2)^2 \\
 G(R) &= K^{quartic} R (R^2 - R_{eq}^2) \\
 H(R) &= K^{quartic} (3R^2 - R_{eq}^2)
 \end{aligned}
 \tag{47}$$

*Dihedral potentials*

For the dihedral angle, one normally takes either a harmonic potential, or a periodic one (with periodicity  $n$ ):

$$\begin{aligned}
 U(\phi_{now}) &= K_\phi \left( 1 + \cos(n\phi_{now} - \phi_{shift}) \right) = K_\phi \left( 1 + \cos(n(\phi_{now} - \phi_{eq}) - \pi) \right) \\
 \phi_{shift} &= n\phi_{eq} + \pi
 \end{aligned}
 \tag{48}$$

The derivatives with respect to the dihedral angle are then:

$$\begin{aligned}
 G(\phi_{now}) &= -nK_\phi \sin(n\phi_{now} - \phi_{shift}) \\
 H(\phi_{now}) &= -n^2 K_\phi \cos(n\phi_{now} - \phi_{shift})
 \end{aligned}
 \tag{49}$$