



# A Charge Analysis Derived from an Atomic Multipole Expansion

MARCEL SWART, PIET TH. VAN DUIJNEN, JAAP G. SNIJDERS

*Theoretische Chemie, Materials Science Centre, Rijksuniversiteit Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands*

*Received 12 February 2000; accepted 28 June 2000*

**ABSTRACT:** A new charge analysis is presented that gives an accurate description of the electrostatic potential from the charge distribution in molecules. This is achieved in three steps: first, the total density is written as a sum of atomic densities; next, from these atomic densities a set of atomic multipoles is defined; finally, these *atomic* multipoles are reconstructed *exactly* by distributing charges over all atoms. The method is generally applicable to any method able to provide *atomic* multipole moments, but in this article we take advantage of the way the electrostatic potential is calculated within the Density Functional Theory framework. We investigated a set of 31 molecules as well as all amino acid residues to test the quality of the method, and found accurate results for the molecular multipole moments directly from the DFT calculations. The deviations from experimental values for the dipole/quadrupole moments are also small. Finally, our Multipole Derived Charges reproduce both the *atomic* and *molecular* multipole moments *exactly*. © 2000 John Wiley & Sons, Inc. J Comput Chem 22: 79–88, 2001

**Keywords:** density functional theory; charge analysis; multipole expansion; force field

Correspondence to: M. Swart; e-mail: m.swart@chem.rug.nl

Contract/grant sponsors: The Netherlands Foundation for Chemical Research, The Netherlands Organization for Scientific Research, and Unilever Research Vlaardingen

This article includes Supplementary Material available from the author upon request or via the Internet at <ftp.wiley.com/public/journals/jcc/suppmat/22/79> or <http://journals.wiley.com/jcc/>

## Introduction

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

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

however, this is not well suited for standard use in molecular simulations. One could simplify this by using a one-center multipole expansion, but this cannot be applied for larger molecules, because one needs to go to high order multipoles. Associated with this is a large convergence radius, outside which the expansion is valid. A better method (for instance, the Distributed Multipole Analysis<sup>1,2</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 smaller (and per center). 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). This involves a number of subsequent  $[3 \times 3]$ -transformations for each moment (one for dipoles, two for quadrupoles, etc.). 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 also removes 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>3</sup> especially in view of constructing *accurate* force fields for simulations using Molecular Dynamics or Monte Carlo techniques.<sup>4,5</sup> Our interest is in the field of biochemistry, in particular copper proteins,<sup>6</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 while studying properties like the redox potential, where one wants to see the behavior of the protein when an electron is removed from the system. The first objective of this study is to obtain atomic charges that give an *accurate* representation of the electrostatic potential outside the

quantumchemical system under study, in our case, the active site of copper proteins.

Many methods to obtain atomic charges from a quantumchemical calculation are available (detailed surveys can be found elsewhere<sup>3,7,8</sup>), where there is a general preference in the molecular modeling world for *Potential-Derived* (PD) charges.<sup>3,9</sup> However, associated with this class of charge analyses are several major drawbacks all of which are related more or less to its need of a grid of points where the quantumchemical potential is fitted by the electrostatic potential from the charges. This results in: (a) a strong dependence on the choices made for determining the grid; (b) an arbitrary change in parameters result in different atomic charges; (c) some methods are even orientation-dependent;<sup>10,11</sup> (d) a strong dependence on the method how to fit the charges; (e) underdetermined set of equations, multiple “solutions” possible; (f) numerical (in)stability; (g) great uncertainties in assigning charges to atoms *buried* within the molecule, because the potential outside the molecule is mainly determined by the atoms near the surface;<sup>3</sup> and (h) an enormous increase in CPU-time, because 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>9</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>12</sup> but formulated in another way and using more accurate *atomic* multipoles. There are three stages involved in this method: first, we write the molecular charge density as a sum of atomic densities. Second, 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. Third, 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 thing 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, because 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 article 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 Multipole Expansion

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

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

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 Results section):

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

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

$$\int |\rho_{AB} - \tilde{\rho}_{AB}| d\tau \quad (5)$$

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

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

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

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

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

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

with the real spherical harmonics  $Z_{lm}^{13}$  and the multipole “moments”  $M_{lm}$  where  $\mathbf{R}_{sA}$  is the distance vector from nucleus  $A$  to a point  $\mathbf{r}_s$ . These “moments” can be obtained from the coefficients as:

$$M_{lm}^A = \sum_{i \in A} c_i^A \int f_i^A(\mathbf{r}_2) \mathbf{r}_2^l Z_{lm}(\mathbf{r}_2) d\tau_2 \quad (9)$$

These multipole “moments” are not equal to the multipoles from the Buckingham convention (see Computational Details), but they can be transformed into them easily 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 adding charges  $q_{s,A}$  to *all* atoms that preserve (up to order  $X$ ) that particular set of multipoles. The represented multipole moments of the charges  $q_{s,A}$  are obtained with the position vectors relative to atom  $A$ ,  $\mathbf{r}_{is,A}$ , as:

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

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:

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

where  $d_A$  is the distance from atom  $A$  to its nearest neighbor,  $\zeta$  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 ( $Q_A^{\text{repr}}, \mu_{i,A}^{\text{repr}}, \Theta_{jk,A}^{\text{repr}}, \Omega_{lmn,A}^{\text{repr}}$ ) to be equal to the *atomic* multipoles from the multipole expansion ( $Q_A^{\text{MPE}}, \mu_{i,A}^{\text{MPE}}, \Theta_{jk,A}^{\text{MPE}}, \Omega_{lmn,A}^{\text{MPE}}$ ). We achieve this by minimizing the following function where the constraints are met by using Lagrangian multipliers  $\alpha_A, \beta_{i,A}, \gamma_{jk,A}, \Delta_{lmn,A}$ :

$$\begin{aligned} g_A = & \sum_s \frac{q_{s,A}^2}{2w_{s,A}} + \alpha_A (Q_A^{\text{MPE}} - Q_A^{\text{repr}}) \\ & + \sum_i \beta_{i,A} (\mu_{i,A}^{\text{MPE}} - \mu_{i,A}^{\text{repr}}) \\ & + \sum_{jk} \gamma_{jk,A} (\Theta_{jk,A}^{\text{MPE}} - \Theta_{jk,A}^{\text{repr}}) \\ & + \sum_{lmn} \Delta_{lmn,A} (\Omega_{lmn,A}^{\text{MPE}} - \Omega_{lmn,A}^{\text{repr}}) \end{aligned} \quad (12)$$

With this choice of function, we ensure that the distribution mainly takes place close to the atom  $A$  where the multipoles are located. For the redistributed charges  $q_{s,A}$  we obtain the following equation:

$$\begin{aligned} q_{s,A} = w_{s,A} \left( \alpha_A + \sum_i \beta_{i,A} r_{is,A} + \sum_{jk} \gamma_{jk,A} \left( \frac{3}{2} r_{js,A} r_{ks,A} \right. \right. \\ \left. \left. - \frac{1}{2} \delta_{jk} r_{s,A}^2 \right) + \sum_{lmn} \Delta_{lmn}(\dots) \right) \end{aligned} \quad (13)$$

that clearly shows 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:

$$\begin{aligned} Q_A - \sum_s q_{s,A} &= 0 \quad \Leftrightarrow \\ Q_A = \alpha_A \sum_s w_{s,A} + \sum_i \beta_{i,A} \sum_s w_{s,A} r_{is,A} \\ &+ \sum_{jk} \gamma_{jk,A} \sum_s w_{s,A} \left( \frac{3}{2} r_{js,A} r_{ks,A} \right. \\ &\left. - \frac{1}{2} \delta_{jk} r_{s,A}^2 \right) + \dots \end{aligned} \quad (14)$$

$$\begin{aligned} \mu_{t,A} - \sum_s q_{s,A} r_{ts,A} &= 0 \quad \Leftrightarrow \\ \mu_{t,A} = \alpha_A \sum_s w_{s,A} r_{ts,A} + \sum_i \beta_{i,A} \sum_s w_{s,A} r_{ts,A} r_{is,A} \\ &+ \sum_{jk} \gamma_{jk,A} \sum_s w_{s,A} r_{ts,A} \left( \frac{3}{2} r_{js,A} r_{ks,A} \right. \\ &\left. - \frac{1}{2} \delta_{jk} r_{s,A}^2 \right) + \dots \end{aligned} \quad (15)$$

$$\begin{aligned} \Theta_{tu,A} - \sum_s q_{s,A} \left( \frac{3}{2} r_{ts,A} r_{us,A} - \frac{1}{2} \delta_{tu} r_{s,A}^2 \right) &= 0 \quad \Leftrightarrow \\ \Theta_{tu,A} = \alpha_A \sum_s w_{s,A} \left( \frac{3}{2} r_{ts,A} r_{us,A} - \frac{1}{2} \delta_{tu} r_{s,A}^2 \right) \\ &+ \sum_i \beta_{i,A} \sum_s w_{s,A} \left( \frac{3}{2} r_{ts,A} r_{us,A} - \frac{1}{2} \delta_{tu} r_{s,A}^2 \right) r_{is,A} \\ &+ \sum_{jk} \gamma_{jk,A} \sum_s w_{s,A} \left( \frac{3}{2} r_{ts,A} r_{us,A} - \frac{1}{2} \delta_{tu} r_{s,A}^2 \right) \\ &\times \left( \frac{3}{2} r_{js,A} r_{ks,A} - \frac{1}{2} \delta_{jk} r_{s,A}^2 \right) + \dots \end{aligned} \quad (16)$$

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  $[Q, \mu_i]$ , 10 up to quadrupoles  $[Q, \mu_i, \Theta_{jk}]$  and 20 up to octupoles  $[Q, \mu_i, \Theta_{jk}, \Omega_{lmn}]$ ) 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  $X$ ):

$$q_s^{\text{MDC}} = \sum_A q_{s,A} \quad (17)$$

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>14</sup> to obtain a good test of the quality of the method, that were taken predominantly from ref. 9, and extended with some that were of interest to us. Furthermore, as a second test set, we used all amino acid residues. Because we are, in *this* article, 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  $\text{NH}_3^+$  and  $\text{COO}^-$  groups by  $\text{NH}_2$  and  $\text{CHO}$ . 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  $\text{CH}_3\text{CONHCHRCONHCH}_3$ , 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>3</sup> with the ADF program<sup>15–18</sup> using the Becke–Perdew exchange-correlation<sup>19,20</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 refers to the values obtained from the *fitted* density (or from the MDC charges, because they represent them exactly), while the latter refers to the values from the *exact* density.

### EXPONENTIAL PREFACTOR

This left us the task to find a value for the exponential prefactor  $\zeta$ , 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 accuracy) 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 article, we use the most commonly used Buckingham convention.<sup>21–23</sup> This convention has the following expressions for the multipole moments:

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

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

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

Inserting the expressions for the  $Z_{lm}$ s in the multipole expansion [eq. (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 NONATOMIC (DUMMY) POSITIONS

For small systems, a proper charge distribution cannot be represented by assigning charges only to atoms.<sup>3</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>3</sup> who use nitrogen as an example and use five points with rather large charges). In practice, adding dummy points is only necessary for small molecules, and in all cases adding one point already suffices to reproduce the multipole moments up to the quadrupole moment.

## Results

### DENSITY FUNCTIONAL THEORY IN ADF-PROGRAM

The method described in this article is generally applicable; however, we use the Density Functional Theory<sup>3</sup> as incorporated in the ADF program<sup>15–18</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>15</sup> It uses a basis set of Slater type orbitals:

$$\chi_\mu(r, \vartheta, \varphi) = r^{\eta_\mu-1} e^{-\alpha_\mu r} Z_{lm}(\vartheta, \varphi) \quad (20)$$

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  $a_i$ :

$$\rho(\mathbf{r}) = \sum_{\mu\nu} P_{\mu\nu} \chi_\mu \chi_\nu \approx \sum_i a_i f_i(\mathbf{r}) \quad (21)$$

The atomic multipoles are obtained from the coefficients:

$$M_{lm}^A = \sum_{i \in A} a_i \int f_i(\mathbf{r}_2) \mathbf{r}_2^l Z_{lm}(\mathbf{r}_2) d\tau_2 \quad (22)$$

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

$$\Delta_{AB} = \int |\rho_{AB} - \tilde{\rho}_{AB}|^2 d\tau \quad (23)$$

The electrostatic potential is now obtained as:<sup>18</sup>

$$V_C(r_s) = \sum_A V_C^A(r_s) = \sum_{i \in A} a_i \int \frac{f_i(r')}{|\mathbf{r}_s - \mathbf{r}'|} dr' \quad (24)$$

or, by using the expansion of  $|\mathbf{r}_s - \mathbf{r}'|$  in spherical harmonics and using the exponential form of the STO fit functions (with the exponent  $\alpha_i$  and the principal quantum number  $n_i$ ):

$$\begin{aligned} V_C^A(r_s) &= \sum_l \sum_{m=-l}^{m=l} \frac{4\pi}{2l+1} Z_{lm}(r_{sA}) I_{lm}^A(r_{sA}) \\ I_{lm}^A(r_{sA}) &= \sum_{i \in A} \delta(l, l_i) \delta(m, m_i) a_i I(n_i, l_i, \alpha_i, r_{sA}) \end{aligned} \quad (25)$$

The function  $I$  is obtained from incomplete Gamma functions, and can be written as the sum of a multipolar and exponentially decaying part:<sup>18</sup>

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

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. Because 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 gives thus the correct potential.

A lot of effort has been put in constructing appropriate *fit sets*, and the currently 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, which was not altered at all. With this new set, the differences

**TABLE I.**  
**Comparison of Standard and New Fit Set.**

	Standart		New	
	Average	Maximum	Average	Maximum
$Q_{tot}$	0.000	0.000	0.000	0.000
$\mu_x$	0.004	0.039	0.004	0.060
$\mu_y$	0.002	0.023	0.001	0.014
$\mu_z$	0.038	0.303	0.026	0.258
$\Theta_{xx}$	0.490	3.640	0.105	0.267
$\Theta_{xy}$	0.047	0.807	0.035	0.413
$\Theta_{xz}$	0.000	0.000	0.000	0.000
$\Theta_{yy}$	0.373	1.788	0.102	0.251
$\Theta_{yz}$	0.000	0.000	0.000	0.000
$\Theta_{zz}$	0.685	3.577	0.139	0.502
Fit test	$5.05 \cdot 10^{-5}$		$3.95 \cdot 10^{-5}$	
Energy difference	0.0289 kcal/mol			

Absolute differences between *exact* and *fitted* molecular multipole moments, fit test difference, and absolute difference in total energy, all averaged over the set of 31 molecules<sup>14</sup>(a.u.).

between the *exact* and *fitted* molecular multipole moments were much smaller (see Table I for values averaged over the set of 31 molecules<sup>14</sup>). 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.

## 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

**TABLE II.** Atomic Charges and Molecular Multipole Moments (a.u.).

Name	Charges	Multipole Moment	DFT Exact	DFT Fitted	Experimental
Benzene	C <sub>6</sub> H <sub>6</sub>				
$q_C$	-0.123	$\Theta_{xx} = \Theta_{yy}$	2.78	2.79	3.23
$q_H$	0.123	$\Theta_{zz}$	-5.55	-5.59	$-6.47 \pm 0.37^{25}$
Carbonmonoxide	CO				
$q_C$	-0.537	$\mu_z$	-0.07	-0.09	-0.04 <sup>26</sup>
$q_O$	-0.852	$\Theta_{xx} = \Theta_{yy}$	0.76	0.74	
$q_{XX}$	1.389	$\Theta_{zz}$	-1.52	-1.48	$-1.86 \pm 0.22^{26}$
Carbondisulfide	CS <sub>2</sub>				
$q_C$	-0.358	$\Theta_{xx} = \Theta_{yy}$	-1.35	-1.58	
$q_S$	0.179	$\Theta_{zz}$	2.71	3.15	$3.17 \pm 0.22^{26}$
Fluorine	F <sub>2</sub>				
$q_F$	0.186	$\Theta_{xx} = \Theta_{yy}$	-0.38	-0.33	
$q_{XX}$	-0.371	$\Theta_{zz}$	0.76	0.66	0.56 <sup>26</sup>
Hydrogen	H <sub>2</sub>				
$q_H$	0.610	$\Theta_{xx} = \Theta_{yy}$	-0.25	-0.31	-0.24
$q_{XX}$	-1.220	$\Theta_{zz}$	0.51	0.61	$0.47 \pm 0.03^{26}$
Hexafluorobenzene	C <sub>6</sub> F <sub>6</sub>				
$q_C$	0.094	$\Theta_{xx} = \Theta_{yy}$	-2.99	-2.78	-3.53
$q_F$	-0.094	$\Theta_{zz}$	5.99	5.56	$7.07 \pm 0.37^{25}$
Hydrogen fluoride	HF				
$q_H$	0.915	$\mu_z$	-0.69	-0.69	-0.72 <sup>26</sup>
$q_F$	0.128	$\Theta_{xx} = \Theta_{yy}$	-0.87	-0.95	
$q_{XX}$	-1.043	$\Theta_{zz}$	1.74	1.91	$-1.75 \pm 0.02^{26}$
Water	H <sub>2</sub> O				
$q_O$	0.458	$\mu_z$	-0.71	-0.73	-0.73 <sup>26</sup>
$q_H$	0.892	$\Theta_{xx}$	-1.82	-1.93	$-1.86 \pm 0.01^{26}$
$q_{XX}$	-1.121	$\Theta_{yy}$	1.91	2.01	$1.96 \pm 0.01^{26}$
		$\Theta_{zz}$	-0.09	-0.08	$-0.10 \pm 0.02^{26}$
Water dimer	(H <sub>2</sub> O) <sub>2</sub>				
$q_H$	0.902	$\mu_x$	-1.38	-1.41	
$q_{XX}$	-1.161	$\mu_y$	-0.58	-0.59	
$q_O$	0.264	$\Theta_{xx}$	2.68	2.84	
$q_{XX}$	-0.748	$\Theta_{xy}$	-3.24	-3.35	
$q_H$	0.699	$\Theta_{yy}$	-2.08	-2.19	
$q_{O'}$	-0.261	$\Theta_{zz}$	-0.60	-0.65	
$q_{H'}$	0.930				
$q_{XX'}$	-1.188				

MDC-Q charges for some molecules out of the set of 31.<sup>14, 24</sup>

Molecular multipole moments from *exact* and *fitted* density from DFT calculations.

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, to check the represented quadrupole moments from both the *exact* and *fitted* densities.

In Table I, we give the absolute differences between the *molecular* multipole moments from the *exact* density on one hand, the *fitted* density on the

**TABLE III.**   
 **MDC-Q Atomic Charges and Molecular Multipole Moments (a.u.).**

Name	Charges	Multipole Moment	<i>DFT Exact</i>	<i>DFT Fitted</i>
Cysteine (Cys, no H)				
N	-1.028	Q	-1.00	-1.00
H	0.299	$\mu_x$	-1.85	-1.93
H	0.350	$\mu_y$	2.23	2.17
C	0.461	$\mu_z$	2.73	2.76
H	-0.139	$\Theta_{xx}$	-17.78	-18.86
C	0.319	$\Theta_{xy}$	5.23	5.34
O	-0.536	$\Theta_{xz}$	5.09	4.64
H	-0.014	$\Theta_{yy}$	14.92	15.60
C	0.398	$\Theta_{yz}$	-8.94	-9.05
H	-0.106	$\Theta_{zz}$	2.86	3.26
H	-0.117			
S	-0.886			
Glycine (Gly)				
N	-0.798	$\mu_x$	-0.37	-0.39
H	0.237	$\mu_y$	-0.56	-0.58
H	0.232	$\mu_z$	-0.05	-0.04
C	1.076	$\Theta_{xx}$	2.46	2.50
H	-0.249	$\Theta_{xy}$	-1.94	-2.39
C	0.316	$H_{xz}$	4.91	4.96
O	-0.517	$H_{yy}$	0.80	0.70
H	-0.051	$H_{yz}$	7.07	7.27
H	-0.247	$H_{zz}$	-3.26	-3.23
Histidine (His-A)				
N	-0.882	$\mu_x$	2.20	2.15
H	0.287	$\mu_y$	-0.15	-0.16
H	0.291	$\mu_z$	0.63	0.63
C	0.505	$H_{xx}$	15.07	14.86
H	-0.087	$H_{xy}$	-1.79	-1.41
C	0.390	$H_{xz}$	6.49	6.36
O	-0.471	$H_{yy}$	-8.51	-8.45
H	-0.039	$H_{yz}$	0.47	0.91
C	0.440	$H_{zz}$	-6.56	-6.41
H	-0.109			
H	-0.178			
C	0.062			
N	-0.613			
C	0.382			
H	0.081			
N	-0.537			
H	0.381			
C	-0.035			
H	0.132			



TABLE III.  
(Continued)

Name	Charges	Multipole Moment	DFT Exact	DFT Fitted
Methionine (Met)				
N	-0.882	$\mu_x$	0.58	0.57
H	0.332	$\mu_y$	-0.99	-1.10
H	0.280	$\mu_z$	0.10	0.18
C	0.361	$\Theta_{xx}$	0.86	1.01
H	-0.103	$\Theta_{xy}$	2.26	1.97
C	0.367	$\Theta_{xz}$	1.27	1.62
O	-0.434	$\Theta_{yy}$	2.65	3.16
H	-0.021	$\Theta_{yz}$	-0.50	-1.24
C	0.496	$\Theta_{zz}$	-3.52	-4.17
H	-0.113			
H	-0.203			
C	0.176			
H	-0.033			
H	-0.057			
S	-0.298			
C	-0.142			
H	0.079			
H	0.104			
H	0.089			

MDC-Q charges for some amino acid residues.<sup>24</sup>Molecular multipole moments from *exact* and *fitted* density from DFT calculations.

other, averaged over the set of 31 molecules (see Computational Details). 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 (available as supplementary data). 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 a.u. for the quadrupole moments.

### MDC CHARGES

The MDC-Q charges as well as the calculated *exact* multipole moments, fitted multipole moments, and experimental values of some of the 31 small molecules<sup>14, 24</sup> and the amino acid residues are collected in, respectively, Tables II and III.

The other data are available as supplementary data.<sup>24</sup>

In these tables, we give the charges on all atoms, and added points (indicated as dummy atoms XX),

as needed to give an accurate description of the charge distribution. These dummy points were centered on the bond midway point between the atoms where necessary. It proved necessary not only for the diatomic molecules, but also for some other small molecules like, for instance, water. There we put two equivalent dummy atoms in the middle between the oxygen and each hydrogen atom. We included them at the same positions in the water dimer. Comparing the atomic charges, it is interesting to see the charge shift away from the hydrogen involved in the hydrogen bond (H3), as was to be expected. It can be seen immediately from the dummy atom next to it, which shifts from -1.12 to -0.75.

A classical case is the carbonmonoxide molecule, with an experimental dipole moment of -0.04 a.u. leading to a small negative charge on the carbon and an equally small but positive charge on the oxygen atom if only reconstructing up to dipole; MDC-D charges -0.0402 (C), +0.0402 (O). However, they do not give a good representation for the molecular quadrupole moment ( $\Theta_{zz}$ ): -1.522, MDC-D -0.026. This is corrected by adding a third point in the

bond middle point, which again reproduces the *fitted* value: MDC-Q  $-1.475$ .

## 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 article. 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, for example, 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. Because 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.

## References

1. Stone, A. J.; Alderton, M. *Mol Phys* 1985, 56, 1047.
2. Stone, A. J. *Chem Phys Lett* 1981, 83, 233.
3. Jensen, F. *Introduction to Computational Chemistry*; Wiley & Sons: Chichester, UK, 1999.
4. Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*; Academic Press: New York, 1996.
5. Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press: Oxford, 1987.
6. Bendall, D. S. *Protein Electron Transfer*; BIOS Scientific: Oxford, 1996.
7. Bachrach, S. In *Reviews in Computational Chemistry*, Vol. 5; Lipkowitz, K. B.; Boyd, D. B., Eds.; VCH: Weinheim, 1994.
8. Wiberg, K. B.; Rablen, P. R. *J Comput Chem* 1993, 14, 1504.
9. Sigfridsson, E.; Ryde, U. *J Comput Chem* 1998, 19, 377.
10. Chirlian, I. E.; Francl, M. M. *J Comput Chem* 1987, 8, 894.
11. Breneman, C. M.; Wiberg, K. B. *J Comput Chem* 1990, 11, 361.
12. Thole, B. T.; van Duijnen, P. Th. *Theor Chim Acta* 1983, 63, 209.
13. Stone, A. J. *The Theory of Intermolecular Forces*, Appendix B; Clarendon Press: Oxford, 1996.
14. 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.
15. Clementi, E.; Corongiu, G. *Methods and Techniques in Computational Chemistry METECC-95*; STEF: Cagliari, 1995.
16. ADF 2.3.3, *Theoretical Chemistry*; Vrije Universiteit: Amsterdam, 1997.
17. ADF 99, *Theoretical Chemistry*; Vrije Universiteit: Amsterdam, 1999.
18. Fonseca Guerra, C.; Snijders, J. G.; te Velde, G.; Baerends, E. J. *Theor Chem Acc* 1998, 99, 391.
19. Becke, A. D. *Phys Rev A* 1988, 38, 3098.
20. Perdew, J. P. *Phys Rev B* 1986, 33, 8822; Erratum: *Ibid.* 1986, 34, 7406.
21. Buckingham, A. D. In *Intermolecular Interactions: From Diatomics to Biopolymers*; Pullman, B., Ed.; John Wiley & Sons: Chichester, 1978, p. 1.
22. Buckingham, A. D. *Q Rev Chem Soc Lond* 1959, 13, 183.
23. Buckingham, A. D.; Disch, R. L.; Dunmur, D. A. *J Am Chem Soc* 1968, 90, 3104.
24. Supplementary data available on: <http://www.interscience.wiley.com/jpages/0192-8651/suppmat>.
25. Battaglia, M. R.; Buckingham, A. D.; Williams, J. H. *Chem Phys Lett* 1981, 78, 421.
26. Gray, C. G.; Gubbins, K. E. *Theory of Molecular Fluids. Volume 1: Fundamentals*, Appendix D; Clarendon Press: Oxford, 1984.