

# An optimal point-charge model for molecular electrostatic potentials

ANDREW C. SIMMONETT†, ANDREW T. B. GILBERT\*‡ and PETER M. W. GILL‡

†School of Chemistry, University of Nottingham, Nottingham, NG7 2RD, UK

‡Research School of Chemistry, Australian National University, ACT 0200, Australia

(Received 29 April 2005; in final form 16 May 2005)

Motivated by the Legendre expansion of the electrostatic potential (ESP), we propose a method for obtaining atomic point-charges for a molecule based on reproducing the low-order multipole moments of the system. The resulting multipole-derived charges (MDCs) are well defined, do not require sampling of the ESP at grid points around the molecule and provide excellent reproduction of the electrostatic potential. No constraints are placed on the magnitude of the atomic charges.

## 1. Introduction

Atomic point-charges are often used to provide simple models for the electrostatic potential (ESP) of a molecule. These models are widely used in applications such as molecular dynamics simulations and rationalising bonding in weakly bound systems. Many prescriptions for obtaining atomic charges have been suggested in the literature [1–12] and the proliferation of schemes reflects the lack of an unambiguous quantum definition of an atomic charge within a molecule. Although the atomic charges are not observable quantities, the electrostatic properties of their aggregate should reflect the observable properties of the original system. The ESP is one such quantity that can help establish the connection between the fictitious atomic charges and the true system. Indeed, many atomic charge models make explicit use of the ESP by evaluating the potential at points around the molecule and then using these values in a least-squares fitting procedure [2, 4, 6–11]. The points are usually chosen to lie on carefully selected shells enveloping the molecule, for example the van der Waals surface, however, this sampling procedure is arbitrary and leads to charges that depend on the orientation of the molecule and the choice of sampling points.

These fitting procedures are usually performed subject to the constraint that the molecular charge is conserved by the sum of the point charges. The dipole moment, and sometimes higher multipole moments, can serve as

additional constraints on the fit [11]. Restraints may also be placed on the values of the charges which introduce penalties if the magnitudes of the charges become too large. This approach yields charges that are more ‘intuitive’ at the expense of the quality of the ESP that is reproduced.

If we seek a point-charge model that is optimal with respect to reproducing the ESP,  $V(\mathbf{r})$ , then ideally it should provide a ‘global’ best-fit to the true potential and should not be dependent on the location of sampling points. Nor should it be degraded by restraints placed on the values of the charges. A naive approach may therefore seek to minimise the least-squares residual between the two potentials:

$$z = \int \left[ V(\mathbf{r}) - \sum_i \frac{q_i}{|\mathbf{r} - \mathbf{r}_i|} \right]^2 d\mathbf{r}, \quad (1)$$

where the  $q_i$  are the point charges. However, a moment’s reflection will reveal that the regions of space that contribute most significantly to  $z$  are those close to the locations,  $\mathbf{r}_i$ , of the point charges where the model potential explodes. These are precisely the regions where we expect the model to perform most poorly. A better approach is to concede that our choice of point charges will always be inadequate for modelling the potential inside the charge distribution of the molecule, and focus on the potential outside this region.

In this paper, we propose a route to well-defined charges that yield the optimal reproduction of the ESP in the chemically important valence regions of a molecule. The charges reproduce the low-order multipole moments of a system and the method avoids the expensive

\*Corresponding author. Email: andrew.gilbert@anu.edu.au



non-vanishing multipole moments. If we define

$$\varepsilon_\ell^m \equiv \sum_{i=1}^N g_i q_i R_\ell^m(\mathbf{A}_i) - \langle R_\ell^m \rangle, \quad (5)$$

where  $\mathbf{A}_i$  is the position of the  $i$ th symmetry-distinct nucleus and  $g_i$  is the number of atoms that are symmetry equivalent to it, then the charges,  $q_i$ , satisfy

$$\varepsilon_\ell^m = 0, \quad \ell = 0, \dots, L, \quad m = -\ell, \dots, \ell. \quad (6)$$

In the  $C_{2v}$  water molecule, for example, there are  $N=2$  symmetry-distinct charges and table 1 shows that these exactly reproduce the charge and dipole moment. Thus,  $L=1$  and the matrix equation that results is

$$\begin{bmatrix} R_0^0(\mathbf{A}_H) & R_0^0(\mathbf{A}_O) \\ R_1^0(\mathbf{A}_H) & R_1^0(\mathbf{A}_O) \end{bmatrix} \begin{bmatrix} 2q_H \\ q_O \end{bmatrix} = \begin{bmatrix} \langle R_0^0 \rangle \\ \langle R_1^0 \rangle \end{bmatrix}. \quad (7)$$

All linear molecules are complete (table 1).

In the incomplete-shell case, we constrain the charges so that they exactly reproduce all the moments up to order  $L-1$ , and perform a least-squares fit to the non-zero  $L$ th-order moments. This is effected by making the objective function

$$z = \frac{1}{2} \sum_{m=-L}^L [\varepsilon_L^m]^2 + \sum_{\ell=0}^{L-1} \sum_{m=-\ell}^{\ell} \lambda_\ell^m \varepsilon_\ell^m \quad (8)$$

stationary with respect to the unknown charges,  $q_i$ , and Lagrange multipliers,  $\lambda_\ell^m$ . This yields the simultaneous equations

$$\begin{aligned} \sum_{m=-L}^L \varepsilon_L^m R_L^m(\mathbf{A}_k) + \sum_{\ell=0}^{L-1} \sum_{m=-\ell}^{\ell} \lambda_\ell^m R_\ell^m(\mathbf{A}_k) &= 0, \\ k &= 1, \dots, N, \\ \varepsilon_\ell^m &= 0, \quad \ell = 0, \dots, L-1, \quad m = -\ell, \dots, \ell. \end{aligned} \quad (9)$$

The resulting coefficient matrix is symmetric and can be treated using specialized symmetric-matrix techniques, if desired.

The formaldehyde molecule is an example of this second case. Table 1 shows that, because the molecule has  $C_{2v}$  symmetry and  $N=3$  symmetry-distinct atoms, its MDCs will exactly reproduce the charge and  $\langle R_1^0 \rangle$  dipole moment, and least-squares fit the  $\langle R_0^0 \rangle$  and  $\langle R_2^0 \rangle$

quadrupole moments. The resulting equation is

$$\begin{bmatrix} F_{C,C} & F_{C,O} & F_{C,H} & R_0^0(\mathbf{A}_C) & R_1^0(\mathbf{A}_C) \\ F_{C,O} & F_{O,O} & F_{O,H} & R_0^0(\mathbf{A}_O) & R_1^0(\mathbf{A}_O) \\ F_{C,H} & F_{O,H} & F_{H,H} & R_0^0(\mathbf{A}_H) & R_1^0(\mathbf{A}_H) \\ R_0^0(\mathbf{A}_C) & R_0^0(\mathbf{A}_O) & R_0^0(\mathbf{A}_H) & 0 & 0 \\ R_1^0(\mathbf{A}_C) & R_1^0(\mathbf{A}_O) & R_1^0(\mathbf{A}_H) & 0 & 0 \end{bmatrix} \times \begin{bmatrix} q_C \\ q_O \\ 2q_H \\ \lambda_0^0 \\ \lambda_1^0 \end{bmatrix} = \begin{bmatrix} G_C \\ G_O \\ G_H \\ \langle R_0^0 \rangle \\ \langle R_1^0 \rangle \end{bmatrix}, \quad (10)$$

where

$$\begin{aligned} F_{x,y} &\equiv R_2^0(\mathbf{A}_x) R_2^0(\mathbf{A}_y) + R_2^2(\mathbf{A}_x) R_2^2(\mathbf{A}_y), \\ G_x &\equiv R_2^0(\mathbf{A}_x) \langle R_2^0 \rangle + R_2^2(\mathbf{A}_x) \langle R_2^2 \rangle. \end{aligned} \quad (11)$$

This second case occurs often in nonlinear molecules, as a partially filled shell is more likely to be found than a completely filled one.

The algorithm embodied in equations (6) and (9) has been implemented within a development version of the QCHEM software package [14].

### 3. Results

The MDCs, and a selection of charges obtained by other methods, are shown in table 2 for methanol at the B3LYP/6-31G(d) level of theory. The MDCs and CHELMO charges both reproduce the two non-zero dipole moments in the molecule, but take very different values. The absolute error in the ESP produced by these point charges, relative to the exact B3LYP/6-31G(d) potential, is shown in figure 1. The population analyses (NPA, AIM) tend to have poorer ESP reproduction, as expected, although even the CHELP and CHELP-BOW methods, which aim to fit the ESP, perform relatively poorly in this case. The ESP produced by the MDCs is of a similar quality to that produced by the better performing ESP-derived charges.

The nitrobenzene molecule (figure 2) has HF/6-31G(d) MDCs that are far removed from the more 'reasonable' MPA values. The MDCs on the ring carbons are large and oscillatory, and the oxygen atoms carry a positive charge, against electronegativity arguments. Figure 3 shows that, despite their counter-intuitive values, the MDCs yield a more accurate ESP

Table 2. B3LYP/6-31G(d) atomic charges in methanol, according to various methods.<sup>a</sup>

Method	C	O	H <sub>in-plane</sub>	H <sub>out-of-plane</sub>	H <sub>hydroxyl</sub>
MDC	1.00	-0.76	-0.16	-0.23	0.38
MPA	-0.21	-0.60	0.16	0.13	0.39
NPA	-0.31	-0.74	0.21	0.18	0.47
AIM	0.49	-1.07	0.03	0.00	0.54
CHELPG	0.26	-0.55	-0.02	-0.02	0.33
CHELPG	0.18	-0.55	0.01	0.01	0.35
RESP	0.04	-0.52	0.04	0.04	0.35
MK	0.07	-0.53	0.04	0.04	0.35
CHELPG-BOW	0.14	-0.51	0.02	0.02	0.33
CHELMO	-0.02	-0.48	0.06	0.06	0.33

<sup>a</sup>Values and acronyms are from [11].

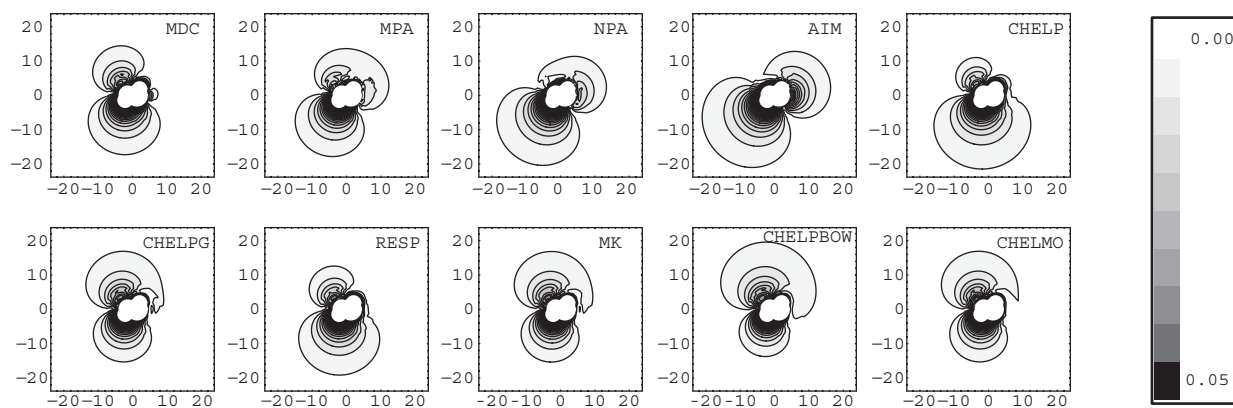


Figure 1. The absolute error in the ESP produced by B3LYP/6-31G(d) charges in methanol (table 2), relative to the exact B3LYP/6-31G(d) potential. The slice through the symmetry plane of the molecule is shown and the white area in the centre of the plot represents the van der Waals surface of the molecule. All charges, except MDCs and MPA, were obtained from [11].

than either the CHELP or Mulliken charges. There is no inherent problem with this conflict, as long as we concede that the different charges have different applications. The MDCs will be useful as part of a force-field for a simulation, but cannot be considered indicative of the charge distribution within the molecule.

#### 4. Conclusions

A simple and efficient route to multipole-derived charges has been proposed. Like the MPA, the charges can be quickly obtained from an *ab initio* charge density. The charges exactly reproduce the lower-order electrostatic moments by construction and provide an excellent description of the ESP about the molecule. Unlike the ESP fitting methods, the potential does not have to be sampled at prescribed points around the molecule. This results in a much simpler algorithm for

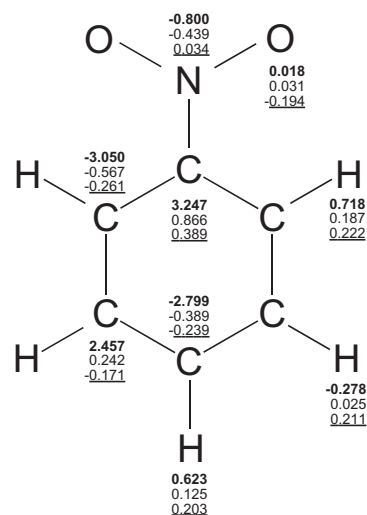


Figure 2. The nitrobenzene molecule showing charges obtained from the MDC (shown in bold), CHELP and MPA (underlined) methods.

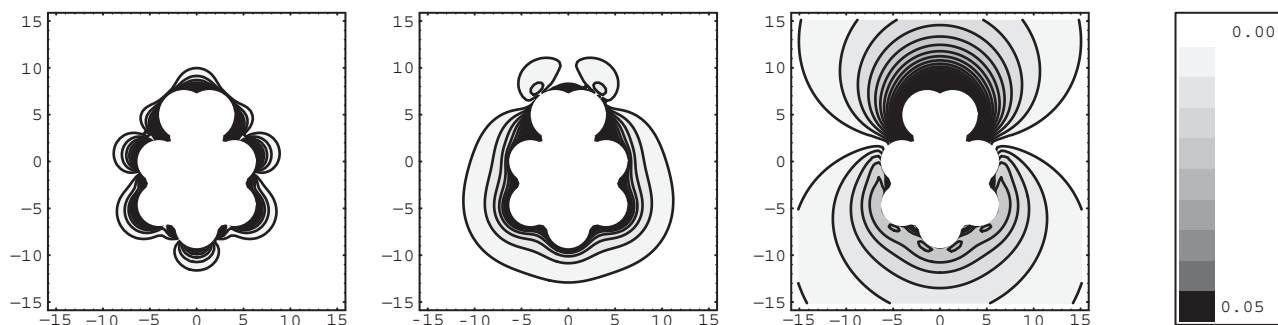


Figure 3. The absolute difference in the ESP due to MDCs (left), CHELP (centre) and Mulliken charges (right), relative to the HF/6-31G(d) ESP, in nitrobenzene. The white area in the centre of the plot represents the van der Waals surface of the molecule.

determining the charges and does not introduce any bias arising from the particular choice of sampling points. The resulting charges have been shown to yield a potential that is as good, if not better, than that obtained from other point-charge models. The molecular multipole moments are observable quantities so the charges can, in principle, be obtained experimentally.

#### Acknowledgments

This work was supported by EPSRC (JREI grant GR/R62052). ATBG is grateful for funding and support from QCHEM Inc. ACS thanks S. Wheeler and M. Schuurman for helpful comments.

#### References

- [1] R.S. Mulliken. *J. chem. Phys.*, **23**, 1833 (1955).
- [2] U.C. Singh, P.A. Kollman. *J. comput. Chem.*, **5**, 129 (1984).
- [3] A.E. Reed, R.B. Weinstock, F.A. Weinhold. *J. chem. Phys.*, **83**, 735 (1985).
- [4] L.E. Chirlian, M.M. Francl. *J. comput. Chem.*, **6**, 894 (1987).
- [5] R.F.W. Bader. *Atoms in Molecules, A Quantum Theory*, Clarendon Press, Oxford (1990).
- [6] B.H. Besler, K.M. Merz, P.A. Kollman. *J. comput. Chem.*, **11**, 431 (1990).
- [7] C.M. Breneman, K.B. Wiberg. *J. comput. Chem.*, **11**, 361 (1990).
- [8] C.I. Bayly, P. Cieplak, W.D. Cornell, P.A. Kollman. *J. phys. Chem.*, **97**, 10269 (1993).
- [9] J.G. Angyán, C. Chipot. *Int. J. quantum Chem.*, **52**, 17 (1994).
- [10] K. Hinsen, B. Roux. *J. comput. Chem.*, **18**, 368 (1997).
- [11] E. Sigfridsson, U. Ryde. *J. comput. Chem.*, **19**, 377 (1998).
- [12] M. Swart, P.Th. van Duijnen, J.G. Snijders. *J. comput. Chem.*, **22**, 79 (2001).
- [13] A. Gelessus, W. Thiel, W. Weber. *J. chem. Educ.*, **72**, 505 (1995).
- [14] J. Kong, C.A. White, A.I. Krylov, C.D. Sherrill, R.D. Adamson, T.R. Furlani, M.S. Lee, A.M. Lee, S.R. Gwaltney, T.R. Adams, C. Ochsenfeld, A.T.B. Gilbert, G.S. Kedziora, V.A. Rassolov, D.R. Maurice, N. Nair, Y. Shao, N.A. Besley, P.E. Maslen, J.P. Dombroski, H. Dachsel, W. Zhang, P.P. Korambath, J. Baker, E.F.C. Byrd, T. Van Voorhis, M. Oumi, S. Hirata, C. Hsu, N. Ishikawa, J. Florian, A. Warshel, B.G. Johnson, P.M.W. Gill, M. Head-Gordon, J.A. Pople. *J. comput. Chem.*, **21**, 1532 (2000).