Efficient method for the generation and display of electrostatic potential surfaces from *ab-initio* wavefunctions

W. Andrzej Sokalski

Institute of Organic and Physical Chemistry I-4, Wrocław, Poland

Scott F. Sneddon

Department of Chemistry, Carnegie Mellon University, Pittsburgh, PA, USA

A cost effective color graphics representation of molecular electrostatic potential surfaces employing the cumulative atomic or bond multipole moments has been described. A general description of the method used to obtain cumulative multipole moments directly from ab-initio *wavefunctions is given, along with an outline of the algorithm for generating electrostatic potential surfaces in the molecular graphics programs MOL17 (FORTRAN 77, Silicon Graphics 3130 and 4D series workstations) and PCMCAMM (Turbo Pascal, IBM PC and PS/2 computers). Examples are given that illustrate the convergence of the multipole expansion, the degree of basis-set dependence compensated by the use of higher atomic moments, and the effect of placing additional expansion centers along the bonds.*

Keywords: electrostatic potential, electrostatic potential surface, quantum mechanical properties, cumulative atomic multipole moments

INTRODUCTION

Electrostatic Molecular Potentials (EMPs) have become a powerful tool for rationalizing the spatial electronic properties responsible for the specific interactions between, and reactivity of, biomolecules (drugs, inhibitors, receptors, active sites, and so forth).[1] For small molecules the EMP can be obtained directly from the quantum chemical wavefunction as an expectation value. However, for large biomolecules the direct evaluation of EMPs from the wavefunction becomes prohibitively costly, especially when it is performed numerous times for different orientations of the molecule.[2] Therefore, for very large molecular systems permanent storage of the wavefunction and costly recalculation of the EMP seems impractical. (Systems with up to 1 000 basis functions are now achievable with direct self consistent field (SCF) methods.) Consequently, for such large systems the extensive information contained in the wavefunction is most frequently reduced using Mulliken population analysis[3] to atomic charge representations. However, such simple models cannot represent all of the local details of the molecular charge distribution. For example, the anisotropy of the charge distribution around atoms with lone-pair electrons cannot be represented using atom-centered monopoles alone. In addition, Mulliken atomic charges are extremely basis-set dependent, even if very extended basis sets of the Hartree–Fock type are employed.[4] This deficiency can be corrected by supplementing monopole models with higher atomic multipole moments.[5–7] To facilitate their application in modeling electrostatic properties of biomolecules, a database of cumulative atomic multipole moments (CAMMs) has been established, containing atom-centered multipoles for 18 natural, rare and protonated forms of nucleic acid complementary bases,[8] as well as 28 neutral and charged forms of amino acids.[9] This database is constantly being extended due to the availability of the CAMM program,[7] which interfaces with read/write files (RWF) and checkpoint files produced by the popular *ab-initio* program GAUSSIAN 82/86/90.[10] A corresponding CAMM package has been developed[11] to use dumpfiles produced by the ATMOL *ab-initio* program.[12] The programs PCMCAMM and MOL17 enable easy display of the EMP using entries from this database or its extensions as input data.

METHODS

Cumulative atomic multipole moments

Among several multicenter multipole expansions,[13–15] which yield practically equivalent results,[16] one of the simplest can

Address reprint requests to Dr. Sokalski at the Institute of Organic and Physical Chemistry I–4, Wyb. Wyspiańskiego 27, 50–370 Wrocław, Poland.
Received 11 July 1990; accepted 15 October 1990

be obtained[5–7] from the expression

$$M_a^{klm} = Z_a u_a^k v_a^l w_a^m - \sum_{I\in a}^{AO}\sum_{J}^{AO} P_{IJ}\langle I|u^k v^l w^m|J\rangle - \sum_{k'>0}^{k}\sum_{l'>0}^{l}\sum_{m'>0}^{m}\binom{k}{k'}\binom{l}{l'}\binom{m}{m'} \times u_a^{k-k'} v_a^{l-l'} w_a^{m-m'} M_a^{k'l'm'} \quad klm \neq k'l'm' \quad (1)$$

where Z_a denotes nuclear charge, $\langle I|u^k v^l w^m|J\rangle$ the one-electron multipole moment integral, and P_{IJ} the density matrix element ($u,v,w = x,y,z$). The first two terms (M_a^{000}) in Equation 1 are equivalent to the atomic charge definition in Mulliken population analysis.[3]

The higher order CAMMs M_a^{klm} obtained from the recursive formula (Equation 1) represent details of the molecular charge distribution not described by lower moments.[5–7] Further improvement of the atomic multipole expansion may be obtained by assigning additional expansion centers "a," not only at the atomic nuclei, but also on the bonds.[13,14] Hereafter, members of such expansions will be called cumulative multicenter multipole moments (CMMMs).

The commonly used Mulliken charge definition is certainly not the only method that can be used to obtain the monopole component of the charge distribution. For example, M_a^{000} can be obtained from the general expression

$$M_a^{000} = Z_a - \sum_{I\in a}^{AO}\sum_{J}^{AO}\sum_{K}^{AO} S_{IK}{}^r P_{JK} S_{JI}{}^{1-r} \quad (2)$$

where S_{IJ} denotes the overlap matrix, and r is an arbitrary constant. For $r = 0$ or $r = 1$ one obtains from Equation 2 the Mulliken charges, whereas for $r = 0.5$ one obtains the Lowdin charges.[17] Fortunately, the ambiguous character of existing atomic charge definitions, represented here by the arbitrary nature of the r parameter, can be corrected by using higher cumulative atomic moments. This can be achieved by using Equation 2 instead of the first two terms in Equation 1. In fact, any other definition of atomic charge can be corrected in the same way.[11] In this way, one can generate many different CAMM representations from the same wavefunction, each summing to give identical molecular moments. From a practical point of view, it is important to select expansions with the best convergence properties. Our preliminary tests, performed for glycine,[11] indicate that the best convergence of the CAMM expansion is obtained for $r = 0$ or $r = 1$; i.e., the Mulliken charge definition used in the original CAMM scheme.[5–7] In addition, the use of higher CAMMs considerably reduces the basis-set dependence of electrostatic properties (e.g., EMPs) calculated from Mulliken charges alone.[5]

Another advantage of CAMMs or CMMMs is the reduction of the extensive information contained in molecular wavefunctions to a compact representation suitable for deposition in a database. This is of particular importance for large biomolecules. CAMMs and CMMMs can also be produced from wavefunctions at higher levels of theory. Recently the CAMM and CMMM approaches have been extended to handle (besides *ab initio* LCAO MO SCF[5]) CI,[6] MBPT,[6] MRD-CI[19,20] and crystal orbital[20] wavefunctions, in the ground state as well as in excited states.[20] The reduction of the wavefunction to a set of CAMMs will allow easy display of quite accurate representations of these higher level wavefunctions.

Alternative representations of the molecular charge distribution, which are superior to the Mulliken monopole approach, have been obtained by fitting atomic charges to reproduce a set of arbitrarily chosen EMP values around a given molecule.[21,22] Although this method partly accounts for some higher multipole moments,[23] such models are still incapable of representing the entire anisotropy of the local charge distribution around atoms or even certain global properties, such as the perpendicular component of the quadrupole moment in planar molecules.

Finally, the use of simple monopole models of the charge distribution in standard molecular dynamics packages, like CHARMM,[24] AMBER,[25] GROMOS[26] and ECEPP[27] may lead to disturbingly different final results.[28–31] Recent analysis performed by Roterman et al.[30,31] indicates that these differences originate from electrostatic interactions that are determined almost invariably from atomic monopole charge distributions, and they conclude that the use of a multicenter multipole expansion can correct this problem.

In this work the values of the EMP on the molecular surface point p, V_p have been calculated from the expression:

$$V_p = \sum_{a\in A}[M_a^{000} R_{ap}^{-1} + (\mu_a \cdot R_{ap}) R_{ap}^{-3} + (R_{ap}\cdot\Theta\cdot R_{ap}) R_{ap}^{-5}] \quad (3)$$

where μ_a denotes the atomic dipole

$$\mu_a = (M_a^{000}, M_a^{010}, M_a^{001}) \quad (4)$$

and Θ_a^{uv} is the traceless atomic quadrupole tensor

$$\Theta_a^{uv} = \frac{1}{2}\left(3 M_a^{uv} - \delta_{uv}\sum_{t}^{x,y,z} M_a^{tt}\right) \quad (u,v = x,y,z) \quad (5)$$

Display of EMP surfaces in the molecular graphics programs MOL17 and PCMCAMM

MOL17 is an interactive molecular graphics program written for the Silicon Graphics 3130 and 4D series workstations. The general facilities of the program include the ability to read coordinates in a variety of formats, a flexible selection and coloring facility, and the ability to visualize a molecule using a number of representations simultaneously. (Vectors, space-filling, ball-and-stick, and ribbon backbones can be displayed in any combination.) The program maintains main and comparison structures, each with its own color and selections parameters. We have used this feature to compare different models of the charge distribution on the same molecule.

The shading of polygons and depth cueing of vectors is performed, using the standard technique of interpolation, on a color map containing shades of white, red, green, blue and yellow. The color map also contains a region used for shading the electrostatic potential surfaces. The colors in this region of the map are arranged in the order yellow, bright red, pink, white, light blue, bright blue and cyan. The shading of surfaces according to electrostatic potential is a matter of determining the potential at each polygon vertex and indexing through the color map (starting in the center of the map, which is white) in the "red" direction for negative values of the potential, and in the "blue"

direction for positive values of the potential. The surface polygons are smoothly shaded between the colors at their vertices. MOL17 prompts the user for maximum and minimum values of the potential; all values greater than the maximum are made cyan, all values less than the minimum are made yellow. A "net" surface connecting the polygon vertices can be added to aid in visualizing the surface of more complicated molecules (e.g., polypeptide fragments).

PCMCAMM is a program written for the IBM PC and PS/2 personal computers, and shares many of the features of the MOL17 program. PCMCAMM does not currently generate solid surfaces, but rather a dot surface of the EMP. (Dot-surface EMPs are also available in MOL17 and in this program can be interactively rotated.) PCMCAMM does not support a color map, so discrete colors are used and the user is prompted for two lower and two upper cutoff values for the potential. The calculation of the EMP is identical to that used in MOL17.

RESULTS AND DISCUSSION

The importance of including higher moments of the charge distribution when representing the EMP from molecular orbital calculations is demonstrated by the qualitative differences between the potential surfaces plotted in Color Plates 1a and 1b. The Color Plate shows the EMP surface for a benzoic acid dimer (minimal $6s3p/3s$ basis set[32]), calculated with monopoles only (Color Plate 1a) and with monopoles, dipoles and quadrupoles (Color Plate 1b). (All EMPs shown in this work were computed on the van der Waals surface. The EMP can be displayed on all surfaces currently supported by MOL17 and PCMCAMM: These include the van der Waals surface, solvent-accessible surfaces with an adjustable probe radius, and radii read from the fourth element of coordinate files where supported, that is, in the CHARMM and Brookhaven Protein Databank formats.) The center of the benzyl ring has a potential of -32 kcal/(mol charge) in the monopole representation, while the potential at the ring center is 50 kcal/(mol charge) when dipole and quadrupole contributions to the charge distribution are included. Presumably this sort of qualitative difference in electrostatic potential would influence the orientation of polar groups with respect to the ring face. Other features that are present in the quadrupole representation but absent in the monopole representation are the alternation of charge on the carboxyl groups, and the detailed structure of the oxygen lone-pair electrons.

Another potential problem with the use of monopole charge distributions derived from Mulliken population analysis is demonstrated in Color Plates 2a–d. Color Plates 2a and b show the monopole potential surfaces for the amino acid glycine for two similar basis sets, 6-31g* and 6-31g**; Color Plates 2c and 2d show the quadrupole surface for these two basis sets. At the quadrupole level the two EMP surfaces look quite similar, while at the monopole level there are qualitative differences. These differences (which are an artifact of the partitioning of charge in Mulliken population analysis) are largest around the lone-pair electrons, and are largely compensated for when one incorporates higher moments of the charge density. The question of the convergence of the multipole expansion is addressed by comparing the EMP surface obtained from atom-centered multipoles with the surface generated by including multipole expansion centers on the bonds as well. In the case of glycine the additional expansion centers located on the bonds resulted in an EMP surface practically identical to that shown in Color Plate 2d. In previous work the use of bond multipoles was advocated,[14] but at least in the case described above no significant differences have been observed. The bond multipoles can, however, be included in the calculation as a check of convergence.

The size of the basis set required for converged potential surfaces is indicated by comparing Color Plates 2c, 2d and 2e, which show the potential surface (through quadrupoles) for glycine charge densities from the 6-31g*, 6-31g** and 6-31g basis sets respectively. These results indicate that to accurately reproduce the lone-pair structures a basis set of at least 6-31g* quality is required.

Finally, in Color Plate 2f we present the EMP surface from our database,[8–9] which was obtained from *ab-initio* LCAO-MO-SCF calculations with a well balanced minimal valence basis set with effective MODPOT core potentials.[32] The differences between EMPs derived from these wavefunctions and those obtained from more extended 6-31g* and 6-31g** basis set can be seen by comparing Color Plates 2c, 2d and 2f. Although some minor qualitative differences are observed between these charge distributions, our previous extensive studies of intermolecular interactions of 12 hydrogen-bonded dimers indicate that results obtained in quite modest MODPOT basis sets quite closely matched the 6-31g* and 6-31g** results.[33] However, for the most accurate estimates of molecular charge distributions the use of correlated wavefunctions with multiple polarization functions seems to be unavoidable.[6] The need to use multicenter expansions has been recently supported[23,34] within the alternative and equivalent distributed multipole analysis.[13]

CONCLUSION

The display of molecular electrostatic potential surfaces has become an important tool in aiding our understanding of intermolecular interactions. An accurate representation of the charge distribution is essential to retain important features of the interaction between groups. We have presented a method for displaying accurate representations of the EMP from *ab-initio* wavefunctions that requires only slightly more computational effort than for point-charge potentials, and is vastly more efficient than obtaining the potential as a quantum mechanical expectation value. This method is currently in use in our laboratories to study intermolecular interactions, the charge distribution of correlated wavefunctions, and differences between ground state and excited state charge distributions. The use of CAMMs compensates the basis-set dependency of EMPs generated from arbitrarily defined atomic charges.[35]

REFERENCES

1 Tomasi, J. in *Chemical Applications of Atomic and Molecular Electrostatic Potentials* (P. Politzer and D. Truhlar, Eds.), Plenum, New York (1981), 257–94

2 Quarendon, P., Naylor, C.B. and Richards, W.G. Display of quantum mechanical properties on van der Waals surfaces. *J. Mol. Graphics* 1984, **2**, 47

3 Mulliken, R.S. Electronic population analysis on

LCAO-MO molecular wavefunctions. *J. Chem. Phys.* 1955, **23**, 1833–40
4 Kern, C.W. and Karplus, M. Analysis of charge distributions: Hydrogen fluoride. *J. Chem. Phys.* 1964, **40**, 1374–89
5 Sokalski, W.A. and Poirier, R.A. Cumulative atomic multipole representation of the molecular charge distribution and its basis set dependence. *Chem. Phys. Lett.* 1983, **98**, 86–92
6 Sokalski, W.A. and Sawaryn, A. Correlated molecular and cumulative atomic multipole moments. *J. Chem. Phys.* 1987, **87**, 526–34
7 Sawaryn, A. and Sokalski, W.A. Cumulative atomic multipole moments and point charge models describing molecular charge distributions. *Comp. Phys. Commun.* 1989, **52**, 397–408
8 Sokalski, W.A., Hariharan, P.C., and Kaufman, J.J. Library of cumulative atomic multipole moments. I. Nucleic acid bases. *Intl. J. Quantum Chem.* 1987, **14**, 111–26
9 Sokalski, W.A., Maruszewski, K., Hariharan, P.C. and Kaufman, J.J. Library of cumulative atomic multipole moments. II. Neutral and charged amino acids. *Intl. J. Quantum Chem.* 1989, **16**, 119–64
10 Gaussian 90, development version (revision D), Frisch, M.J., Head-Gordon, M., Trucks, G.W., Foresman, J.B., Schlegel, H.B., Raghavachari, K., Robb, M., Binkley, J.S., Gonzalez, C., Defrees, D.J., Fox, D.J., Whiteside, R.A., Seeger, R., Melius, C.F., Baker, J., Martin, R.L., Kahn, L.R., Stewart, J.J.P., Topiol, S. and Pople, J.A. Gaussian Inc., Pittsburgh, PA, USA
11 Sokalski, W.A., unpublished
12 Moncrieff, D. and Saunders, V.R. ATMOL. University of Manchester
13 Stone, A.J. Distributed multipole analysis, or how to describe a molecular charge distribution. *Chem. Phys. Lett.* 1981, **83**, 233–9
14 Vigne-Maeder, F. and Claverie, P. The exact multicenter multipolar part of a molecular charge distribution and its simplified representations *J. Chem. Phys.* 1988, **88**, 4934–48
15 Mezei, M. and Campbell, E.S. Efficient multipole expansion: Choice of order and partitioning techniques. *Theor. Chim. Acta* 1977, **43**, 227–37
16 Spackman, M.A. A simple quantitative model of hydrogen bonding. *J. Chem. Phys.* 1986, **85**, 6587–601
17 Löwdin, P.O. On the nonorthogonality problem connected with the use of atomic wavefunctions in the theory of molecules and crystals. *J. Chem. Phys.* 1950, **8**, 365–75
18 Kaufman, J. and Sokalski, W.A. Library of atomic multipole moments for biopolymer building blocks. In *Theoretical Biochemistry and Molecular Biophysics* (D. Beveridge and R. Lavery, Eds.), Adenine Press (1991), 305–316
19 Roszak, S., Sokalski, W.A. and Kaufman, J.J. in preparation
20 Sokalski, W.A., Keegstra, P.B., Roszak, S. and Kaufman, J.J. *Int. J. Quantum Chem.* 1990, **24**, 51–63
21 Singh, U.C. and Kollman, P. An approach to computing electrostatic charges for molecules. *J. Comp. Chem.* 1984, **5**, 129–45
22 Williams, D.E. Representation of the molecular electrostatic potential by atomic and bond dipole models. *J. Comp. Chem.* 1988, **9**, 745–63, and references quoted therein
23 Price, S.L., Harrison, R.J. and Guest, M.F. An *ab initio* distributed multipole study of the electrostatic potential around an undecapeptide cyclosporin derivative and comparison with point charge electrostatic models. *J. Comp. Chem.* 1989, **10**, 552–67
24 Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M. CHARMM: A program for macromolecular energy, minimization and dynamics calculations. *J. Comp. Chem.* 1983, **4**, 187–217
25 Weiner, S.J., Kollman, P.A., Nguen, D.T. and Case, D.A. An all-atom force field for simulations of proteins and nucleic acids. *J. Comp. Chem.* 1986, **7**, 230–52
26 Hermans, J., Berendsen, H.J.C., van Gunsteren, W. and Postma, J.P.M. A consistent empirical potential for water–protein interactions. *Biopolymers* 1984, **23**, 1513–8
27 Momany, F.A., McGuire, R.F., Burgess, A.W. and Scheraga, H.A. Energy parameters in polypeptides VII: Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, intrinsic torsional potential for the naturally occurring amino acids. *J. Phys. Chem.* 1975, **79**, 2361–81
28 Gilson, M. and Honig, B. Energetics of charge–charge interactions in proteins. *Proteins* 1988, **3**, 32–52
29 Bellido, M.N. and Rullman, J.A.C. Atomic charge models for polypeptides derived from *ab initio* calculations. *J. Comp. Chem.* 1989, **10**, 479–87
30 Roterman, I.K., Gobson, K.D. and Scheraga, H.A. A comparison of the CHARMM, AMBER and ECEPP potentials for peptides. I. Conformational predictions for the tandemly repeated peptide (Asn-Ala-Asn-Pro)9. *J. Biomol. Struct. Dyn.* 1989, **7**, 391–420
31 Roterman, I.K., Lambert, M.H., Gibson, K.D. and Scheraga, H.A. A comparison of the CHARMM, AMBER and ECEPP potentials for peptides. II. Ψ–Φ map for *N*-acetyl alanine *N'*-methyl amide; Comparisons, contrasts and simple experimental tests. *J. Biomol. Struct. Dyn.* 1989, **7**, 420–53
32 Popkie, H.E. and Kaufman, J.J. Molecular calculations with MODPOT, VRDDO and MODPOT/VRDDO procedures I: HF, F_2, HCl and Cl_2. *Int. J. Quantum Chem.* 1976, **10**, 47–57
33 Sokalski, W.A., Hariharan, P.C. and Kaufman, J.J. A self-consistent field interaction energy decomposition study of 12 hydrogen-bonded dimers. *J. Phys. Chem.* 1983, **87**, 2803–10
34 Richards, N.G.J. and Price, S.L. The representation of molecular electrostatics using molecular graphics. *Int. J. Quantum Chem.* 1989, **16**, 3773–85
35 Hermsmeier, M.H. and Gund, T.M. A graphical representation of the electrostatic potentials and electric fields on molecular surfaces. *J. Mol. Graph.* 1989, **7**, 150–2