Point charge representation
of multicenter multipole moments in calculation
# of electrostatic properties*

W. A. Sokalski[1,], M. Shibata[1], R. L. Ornstein[2], and R. Rein[1]**
[1] Department of Biophysics, Roswell Park Cancer Institute, 666 Elm St., Buffalo, NY 14263, USA
[2] Molecular Science Research Center, Pacific Northwest Laboratory***, Richland, WA 99352, USA

Summary. Distributed Point Charge Models (PCM) for CO, $(H_2O)_2$, and HS-SH molecules have been computed from analytical expressions using multi-center multipole moments. The point charges (set of charges including both atomic and non-atomic positions) exactly reproduce both molecular and segmental multipole moments, thus constituting an accurate representation of the local anisotropy of electrostatic properties. In contrast to other known point charge models, PCM can be used to calculate not only intermolecular, but also intramolecular interactions. Comparison of these results with more accurate calculations demonstrated that PCM can correctly represent both weak and strong (intramolecular) interactions, thus indicating the merit of extending PCM to obtain improved potentials for molecular mechanics and molecular dynamics computational methods.

1 Introduction

Molecular force fields form the basis for modeling various properties of biomolecules. Due to the large size of systems involved, the analytical form of the potential functions is only approximate and somewhat arbitrary. This treatment is acceptable in biopolymers for interactions in which the interfragment separation distance is greater than 10 a.u. In this case, the approximation can lead to increased computational economy. However, at smaller separations,

---

approximation of the electrostatic term by point charge interactions is questionable. As a consequence, calculation of the relative stabilities of amino acid complexes can give incorrect estimates of the electrostatic term in a simple point charge approximation (see Fig. 3 in Ref. [15a]). Indeed, recent tests [1–2] indicate that a number of popular force fields can sometimes yield different structures; the major source of error in these tests is the electrostatic term, which uses an atomic point charge model [2]. These tests illustrate a need for a more accurate representation of molecular charge distributions used for calculation of electrostatic interactions for separations close to van der Waals contacts. This work explores refined, yet computationally economic, approaches to calculating electrostatic interactions at small separation distances.

Usually, atomic charges are determined from arbitrary and basis set dependent population analysis [3–4] or, in the best case, from fitting charges to an electrostatic potential grid [5–10] or molecular multipole moments [11–13]. These models do not represent the local anisotropy of charge distributions in large molecules with multiple functional groups such as amino acids; the detailed local description, involving higher atomic multipole moments, is needed to properly describe electrostatic interactions at separation distances close to van der Waals contacts [1–2, 14–15a]. The applications of point charge models, fitting only molecular moments, have previously been used; however, this treatment was limited to small and symmetric molecules [13, 16–19]. In short, these models are not applicable to intra-molecular interactions, such as those in proteins [1–2, 15a]. This deficiency can be overcome with a Point Charge Model (PCM) [15b–c] derived directly from Cumulative Atomic/Multicenter Multipole Moments (CAMM) [15d]. As in other multicenter multipole expansions [20–24], any monopole deficiency can be compensated with higher CAMM moments. This feature allows the accelerated convergence of the multipole expansion by starting it from an appropriate atomic charge definition (Potential Derived (PD) charges [6–7, 9, 10], for example) ensuring minimal contribution from higher moments. Our procedure seems more universal for biological applications since it allows automatic derivation of PCM by analytical formulas from CAMM databases, [15c] without involving an approximate fitting process, which may lead to less accurate results.

Herein we report a study examining the number of expansion centers and the number of terms in the multipole expansion required for obtaining distributed point charges that will yield a convergent representation of electrostatic interactions between molecular subunits at short separations. The method is applied to CO, $(H_2O)_2$ and HS-SH to illustrate the applicability of PCM to represent anisotropic electrostatic properties. These point charges provide an accurate representation of electrostatic interactions, comparable to those obtained from higher moments, with considerable improvement in computational simplicity.

2 Point Charge Model (PCM) derived from multicenter multipole moments

The expectation value of any molecular multipole moment $M^{klm} = \langle x^k y^l z^m \rangle$ can be expressed as a sum of additive segmental contributions. Transformation

of these additive moments to a local coordinate system provides the expression for the Cumulative Multicenter Multipole Moments (CMMM) [15c,e]:

$$m_c^{klm} = \delta_{ac} Z_a x_a^k y_a^l z_a^m - \sum_{r \in c}^{AO} \sum_s^{AO} D_{rs} \langle r | x^k y^l z^m | s \rangle$$
$$- \sum_{\substack{k' > 0 \\ k'l'm' \neq klm}}^{k} \sum_{l' > 0}^{l} \sum_{m' > 0}^{m} \binom{k}{k'} \binom{l}{l'} \binom{m}{m'} x_c^{k-k'} y_c^{l-l'} z_c^{m-m'} m_c^{k'l'm'} \qquad (1)$$

where Z_a stands for atomic core charge; D_{rs}, the generalized density matrix [25]; and $\langle r | x^k y^l z^m | s \rangle$, the multipole integral in atomic orbital basis with Cartesian operators defined relative to a segmental origin [20]. In contrast with an atomic multipole expansion CAMM [15d], where only atomic centers are included, Eq. (1) permits the use of any number of additional expansion centers outside atomic nuclei (on bonds, etc.) allowing increased accuracy and convergence of multipole expansion.

Each Cumulative Multicenter Multipole Moment m_c^{klm} can be expressed [15b,c] as a set of point charges q_p (PCM) located at position $p(x_p, y_p, z_p)$ at a distance R from expansion center c:

$$m_c^{klm} = \sum_{p=1}^{s} q_p (x_p - x_c)^k (y_p - y_c)^l (z_p - z_c)^m \qquad (2)$$

Charges can be generated at any level of moments such as monopole ($s = 1$), dipole ($s = 3$), quadrupole ($s = 9$). We have chosen in this work the quadrupole level ($s = 9$). The detailed expressions for generating PCM are given elsewhere [15b,c].

3 Results and discussion

3.1 PCM for CO molecule

One of the most difficult systems to generate a reliable point charge model is the CO molecule. This is reflected by the large root mean square deviation of the electrostatic molecular potential calculated from potential derived (PD) charges (either correlated or uncorrelated) [26]. Previous work [9] shows that systems containing a carbonyl group often exhibit large differences in interaction energies due to PD charges versus segmental multipole moments. In addition, correlation effects can reverse the polarity of the CO molecule.

A set of PCM derived from atomic multipoles (up to quadrupole) is given in Table 1 for CO molecule. In this case, the extended 6-311G(3d) basis set has been used within Configuration Interaction – Single and Double excitations (CISD) approach to generate the generalized density matrix [25], so that the Hellman–Feynman theorem is satisfied. As the corresponding SCF value of molecular dipole moment (-0.088 a.u.) is close to the corresponding Hartree–Fock limit value (-0.100 a.u. [27]) one may consider that a significant portion of the basis set extension error has been eliminated. Correlated CAMM obtained in well-saturated basis set can be regarded as an accurate representation of a molecular charge distribution [15a], thus serving as a solid basis for systematic development of more approximate models.

The set of 14 point charges presented in Table 1 exactly reproduces molecular and atomic charges, dipoles and quadrupoles of the CO molecule. The magnitudes

Table 1. Point charge model derived from correlated CAMM m^{klm} for CO in 6-311G(3d) basis set and generalized CISD density matrix ($R = 0.1$ a.u.)

Point charge	CAMM	X_p [a.u.]	Y_p [a.u.]	Z_p [a.u.]	q_p [a.u.]
1	C	0	0	-1.157	1184.262
2	m_c^{001}, m_c^{002}	0	0	-1.257	-227.273
3	m_c^{001}, m_c^{002}	0	0	-1.057	-222.767
4	0	0	0	0.868	1112.594
5	m_0^{001}, m_0^{002}	0	0	0.768	-181.179
6	m_0^{001}, m_0^{002}	0	0	0.968	-180.859
7–8	m_c^{200}	$-/+0.1$	0	-1.157	-183.508
9–10	m_c^{020}	0	$-/+0.1$	-1.157	-183.508
11–12	m_c^{200}	$-/+0.1$	0	0.868	-187.685
13–14	m_0^{020}	0	$+/-0.1$	0.868	-187.685

of these charges are directly related to the arbitrarily chosen radius R on which all these charges (except the atom-centered one) are located. The first six nonzero longitudinal components (located on the z axis) can, in principle, be reduced to two atomic centers (using the relation $M^{001} = q_a R_{CO}$) while reproducing the molecular dipole. However, such a two-charge model cannot reproduce the local anisotropy due to atomic dipoles. Moreover, the effect of the remaining eight components cannot be described accurately by charges located on nuclei.

In order to illustrate the role of off-atom point charges, in Table 2, the exact expectation values of molecular electrostatic potential have been compared with results obtained from various simple CAMM and PCM models. The CAMM (R^{-3}) and CAMM (R^{-5}) represent atomic multipole expansions terminated at quadrupole and hexadecapole moments, respectively. PCM ($c = 14$) denotes full Point Charge Model generated from CAMM up to quadrupole,

Table 2. Molecular electrostatic potentials calculated from generalized correlated CISD wavefunction in 6-311G(3d) basis set form various PCM and CAMM models versus accurate expectation values for CO molecule. R distance (in [a.u.]) is measured from the center of the mass in the direction perpendicular to molecular axis

R [a.u.]	exact	CAMM (R^{-5})	CAMM (R^{-3})	PCM ($c = 14$)	PCM ($c = 6$)
3	0.0245	0.0159	0.0190	0.0188	0.1000
4	0.0098	0.0091	0.0096	0.0095	0.0528
5	0.0053	0.0052	0.0054	0.0053	0.0300
6	0.0032	0.0032	0.0033	0.0032	0.0183
7	0.0021	0.0021	0.0021	0.0021	0.0119
8	0.0014	0.0014	0.0014	0.0014	0.0082
9	0.0010	0.0010	0.0010	0.0010	0.0058
10	0.0007	0.0007	0.0007	0.0007	0.0043

whereas PCM ($c = 6$) denotes its subset, where the off-axis charges have been collapsed to atomic centers. There is a good agreement between CAMM (R^{-3}) and PCM ($c = 14$) values while PCM ($c = 6$) fails to reproduce the CAMM (R^{-3}) results at all the distances examined. These results demonstrate the need to include the off-atom point charges by PCM.

In order to test the convergence of the PCM model, based on CMMM in calculating electrostatic interactions, we have chosen to investigate two cases:

– intermolecular interactions in a hydrogen bonded water dimer, and
– intramolecular (torsional) interactions in hydrogen disulfide.

3.2 Intermolecular interactions in $(H_2O)_2$

Figure 1 illustrates relative errors of several CMMM variants of electrostatic multipole interaction energies for a water dimer at different R_{00} distances ($R_{00} = 5.57$ a.u. – equilibrium). It seems that the use of an atomic ($c = 3$) multipole expansion, at any R^{-n} level, is convergent with results from $c = 5$ (one additional center per bond) and $c = 7$ (two additional centers per bond) expansions. As the reference values, we used the electrostatic interaction energy obtained by perturbational treatment using a ($11s7p2d/6s1p$) basis set [28]. Because the perturbational values also include a penetration term not covered within the multipole expansion, the error at equilibrium R_{00} distance does not converge at zero, even at the R^{-5} level. On the same figure, the values from PCM obtained from three sets of atomic multipoles ($c = 3$), up to quadrupole level ($s = 9$) using radius of $R = 0.1$ have also been plotted. The PCM values coincide almost exactly with atomic multipole estimates terminated at the quadrupole-quadrupole term, so these were not plotted here. For example, for $R = 5.67$ a.u. the PCM value is -6.272 kcal/mol, whereas the corresponding CAMM estimate (expansion truncated at quadrupoles) yields -6.270 kcal/mol. Since this PCM model accurately represents the molecular charge distribution, it can be useful in benchmark calculations testing more approximate charge distribution models. The number of charges per expansion center ($c = 9$) can be gradually reduced, while monitoring the level of convergence and its possible influence on accuracy.

3.3 Intramolecular interactions in HS-SH

Figure 2 illustrates torsional barrier estimates obtained in supermolecular LCAO MO SCF, CMMM and PCM approaches ($R = 0.1$ a.u., quadrupole level, 6-31G* basis set). In this case, besides atomic expansion centers, one additional center has been located in the middle of each S–H bond ($c = 6$). The close correspondence of CMMM and PCM results indicates that PCM model can also reasonably represent strong intramolecular interactions, although at such short distances one cannot expect impressive accuracy. However, these preliminary results suggest that it may be possible to estimate torsional potentials in a completely nonempirical way.

It should be noted that use of lower PCM levels ($s = 1, 3$) yields no potential barrier, in agreement with our preliminary study [29].
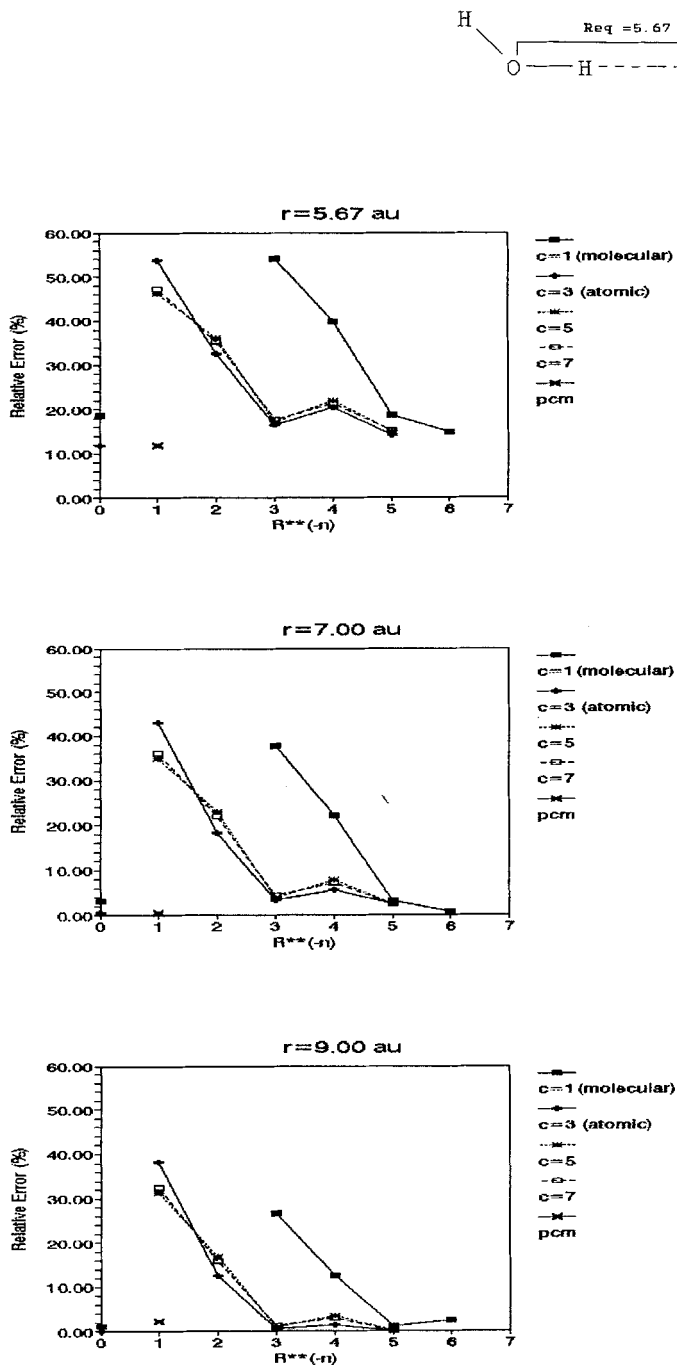
Fig. 1. Relative error in the CMMM estimates of electrostatic multipole interaction energy E_{MTP} $[R^{-n}]$ in water dimer. Accurate perturbational results EL_{exact} [28] obtained in ($11s7p2d/6s1p$) basis set was taken as the reference value in computing the relative errors. PCM results correspond to 3 expansion centers ($c = 3$, quadrupole level ($s = 9$), $R = 0.1$ a.u.
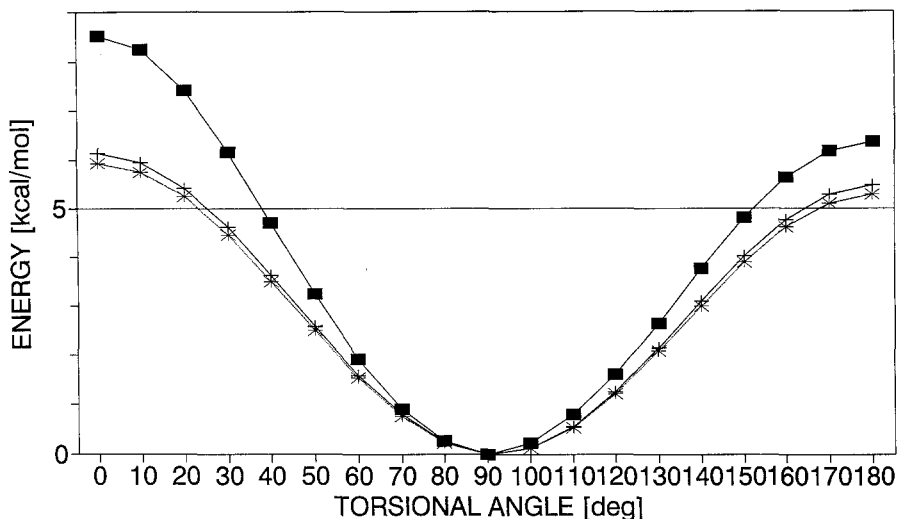
Fig. 2. Torsional potentials for HS–SH molecule calculated in 6-31G* basis set: a) *ab initio* SCF results, —■— *ab initio* SCF; b) CMMM estimates ($c = 6$) up to quadrupole-quadrupole term, —+— CMMM (Q-Q); c) PCM results ($c = 6, s = 9, R = 0.1$ a.u.), —*— PCM ($R = 0.1$ a.u.)

4 Conclusions

Inter- and intramolecular electrostatic multipole interactions can be estimated with reasonable accuracy using the PCM and CMMM approach terminated at the quadrupole level. PCM derived from multicenter multipole moments can represent the local and global molecular charge distribution with any desirable accuracy, including the correlated level, without involving any fitting procedure.

References

1. Roterman IK, Gibson KD, Scheraga HA (1989) J Biomol Struct Dyn 7:391
 2. Roterman IK, Lambert MH, Gibson KD, Scheraga HA (1989) J Biomol Struct Dyn 7:421
 3. Kern CW, Karplus M (1964) J Chem Phys 40:1374
 4. Perrin CL (1991) J Am Chem Soc 113:2865
 5. Momany F (1978) J Phys Chem 82:592
 6. Cox SR, Williams DE (1981) J Comp Chem 2:304
 7. Woods RJ, Khalil M, Pell W, Moffat SH, Smith VH (1990) J Comp Chem 11:297
 8. Lavery R, Zakrzewska K, Pullman A (1984) J Comp Chem 2:263
 9. Ray NK, Shibata M, Bolis G, Rein R (1985) Int J Quantum Chem 27:427
10. Chirlian LE, Francl MM (1987) J Comp Chem 8:894

11. Brobjer TJ, Murrell JN (1981) Chem Phys Lett 77:601
12. Brobjer TJ, Murrell JN (1982) J Chem Soc Faraday Trans II 78:1853
13. Bauer GEW, Huiszoon C (1982) Mol Phys 47:565
14. Mitchell JBO, Price SL (1990) J Comp Chem 11:1217
15. a) Sokalski WA, (1989) in: Beveridge DL, Lavery R (eds) Theoretical chemistry and molecular biophys. Adenine Press, p 239–256
    b) Sawaryn A, Sokalski WA (1989) Comp Phys Comm 52:397
    c) Sokalski WA, Sawaryn A (1992) J Mol Struct 256:91 and references quoted therein
    d) Sokalski WA, Poirier RA (1983) Chem Phys Lett 98:86
    e) Sokalski WA, Sneddon SF (1991) J Mol Graphics 9:74
16. Berns GEW, van der Avoird A (1980) J Chem Phys 72:610
17. Votava C, Ahlrichs R (1981) in: Pullman B (ed) Intermolecular forces. Reidel, 133
18. Hoinkins J, Ahlrichs R, Bohm HJ (1983) Int J Quantum Chem 23:821
19. Bohm HJ, Ahlrichs R, Scharf P, Schiffer H (1983) J Chem Phys 81:1389
20. Rein R (1973) Adv Quantum Chem 7:335
21. Port GNJ, Pullman A (1973) FEBS Lett 31:70
22. Stone AJ (1981) Chem Phys Lett 83:233
23. Vigne-Maêder F, Claverie P (1988) J Chem Phys 88:4934
24. Rico JF, Alvarez-Collado JR, Paniagua M (1985) Mol Phys 56:1145
25. Foresman JB, Head-Gordon M, Pople JA, Frisch MJ (1992) J Phys Chem 96:135
26. Luque FJ, Orozco M, Illas F, Rubio J (1991) J Am Chem Soc 113:5203
27. Laaksonen L, Pykko P, Sundholm D (1986) Comp Phys Rep 4:313
28. Jeziorski B, van Hemert M (1976) Mol Physic 31:713
29. Sokalski WA, Lai J, Luo N, Sun S, Shibata M, Rein R, Ornstein R (1991) Int J Quantum Chem, Quantum Biol Symp 18:61