Calculation of Electrostatic Potential Maps and Atomic Charges for Large Molecules

Gyula Tasi,*,† István Pálinkó,‡ Levente Nyerges,§ Pál Fejes,† and Horst Förster‖

Department of Applied Chemistry, József Attila University, Rerrich B. tér 1, H-6720 Szeged, Hungary, Department of Organic Chemistry, József Attila University, Dóm tér 8, H-6720 Szeged, Hungary, Department of Medicinal Chemistry, Szent-Györgyi Albert University, Dóm tér 8, H-6720 Szeged, Hungary, and Institute of Physical Chemistry, University of Hamburg, Bundesstrasse 45, D-2000 Hamburg 13, Germany

Received July 28, 1992

For nucleotide bases and their derivatives, the atomic charges calculated by the extended Mulliken population analysis within the NDDO approximation reproduce excellently the quantum chemical electrostatic potential field outside the molecular van der Waals surface. It follows that in the case of nucleic acids we have a fast and reliable method for the calculation of molecular electrostatic potential maps and atomic charges.

INTRODUCTION

Molecular electrostatic potential (MEP) maps are often used for the qualitative interpretation of electrophilic and nucleophilic reactions,[1] the calculation of atomic charges,[2] and the studies of molecular similarity.[3] Since calculating quantum chemical MEP maps requires considerable CPU time, substantial effort has been devoted for producing MEP maps using simpler methods[4] or reducing the computation time by other means.[5]

By using semiempirical quantum chemical methods, we are able to reduce the computation time, too.[6] For certain molecules, however, MEP maps derived from these calculations are far from reality or differ from those of ab initio calculations.

The traditional Mulliken (MP) atomic charges[7] are not suitable for calculating MEP maps, and they are not able to reproduce the SCF dipole moments of molecules,[8] too. In spite of these facts, most frequently these charges are used in similarity studies because they can be calculated simply and fast.

Recently, we described a new method for calculating atomic charges.[9,10] It was pointed out that the atomic charges that originated from extended Mulliken population analysis[11] give dipole moments very close to the SCF dipole moments. Furthermore, these atomic charges differ only slightly within the NDDO approximation (MNDO,[12] AM1[13]) from those calculated from exact quantum chemical MEP maps, thus suggesting that outside the van der Waals surface of the molecule, the electrostatic field generated by the point charge system, obtained by extended Mulliken population analysis, reflects the properties of the quantum chemical electrostatic field. This paper provides a quantitative evaluation of this suggestion using examples of nucleotide bases of vital importance.

CALCULATION OF ATOMIC CHARGES

Within the LCAO approximation, the exact MEP value at an r point in the surroundings of a molecule is given as follows (eq 1):

$$V(r) = \sum_{A} \frac{Z_A}{\|r - \mathbf{R}_A\|} - \sum_{\mu}\sum_{\nu} P_{\mu\nu} \int \frac{\varphi_\mu(r')\varphi_\nu(r')\,dr'}{\|r - r'\|} \quad (1)$$

where Z_A is the charge of core A, $\mathbf{R}_A$ is the position vector of atom A, $P_{\mu\nu}$ is an element of the first-order density matrix $\mathbf{P}$, and ϕ_μ is a real atomic basis function. Within the NDDO approximation, the three-center integrals in the second part of the expression need not to be calculated. If we maintain the NDDO approximation during the MEP calculation with MNDO and AM1 methods, the CPU time can be reduced by about 1 order of magnitude.

Luque et al. thoroughly studied the applicability of MNDO and AM1 methods for calculating MEP maps and electrostatic potential (EP) atomic charges.[14] With the help of the inverse Löwdin's transformation, they calculated a quasi ab initio density matrix and determined the three-center integrals as well. Orozco and Luque suggested a practical method for determining the EP atomic charges of large molecules.[15] The molecule is divided into several parts, and these parts are treated independently. The influence exerted by the other units is taken into account with the help of small groups of atoms.

Ferenczy et al., however, calculated the MEP maps and EP atomic charges within the *NDDO approximation*.[16] It was pointed out that the quality of AM1 and STO-3G EP atomic charges were close to each other, with either the optimized or the experimental molecular geometries being used. Good correlation was found between the AM1 and STO 6-31G** EP charges, too.

Huzinaga et al. discussed the theoretical background of Mulliken population analysis in detail.[11] For the ab initio LCAO-MO SCF electron density function, a molecule can be represented by a point charge system containing $n + n(n + 1)/2$ charges according to the extended Mulliken population analysis. The point charge system consists of nuclei (or cores) as well as atomic and overlap electron populations. Within the NDDO approximation, the number of charges is $2n$ since the overlap populations disappear. The centers of electron populations are calculated exactly by first-momentum integrals.

However, for large molecules, the treatment of the point charge system becomes complicated due to the large number of point charges. For reducing the number of point charges or calculating atomic charges, the same method can be used as for calculating EP atomic charges: the exact quantum chemical MEP map should only be changed to the classical electrostatic potential map generated by the extended Mulliken point charge system. The atomic charges obtained are called *population electrostatic potential* (PEP) atomic charges.[9,10]

The numerical method applied for the calculation of EP and PEP atomic charges is a constrained optimization

† Department of Applied Chemistry, József Attila University.
‡ Department of Organic Chemistry, József Attila University.
§ Department of Medicinal Chemistry.
‖ Institute of Physical Chemistry.

0095-2338/93/1633-0296$04.00/0

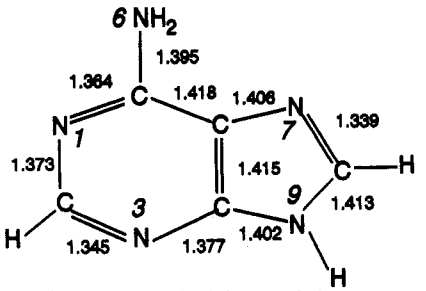

Figure 1. Some PM3 geometrical data and the atom numbering of adenine molecule.

Table I. SCF Dipole Moments of Nucleotide Bases Calculated by Semiempirical Quantum Chemical Methods (Debye)

molecule	CNDO/2	MNDO	AM1	PM3
adenine	2.68	2.51	2.18	2.49
guanine	6.88	5.26	5.91	5.45
cytosine	7.03	5.58	6.26	5.68
thymine	4.09	4.09	4.22	3.88
uracil	4.33	4.13	4.29	3.90

procedure. Generally, one constraint only is taken into account: the sum of atomic charges ought to equal the charge of the molecule. This can be fulfilled in the easiest way by using the Lagrange multiplier method. Then the object function is as follows (eq 2):

$$S(q_1, q_2, ..., q_n, \lambda) = \sum_{i=1}^{N}\left[V_i - \sum_{j=1}^{n}\frac{q_j}{r_{ij}}\right]^2 + \lambda\left[\sum_{j=1}^{n} q_j - q_{mol}\right] \quad (2)$$

where N is the number of points generated outside the van der Waals surface of the molecule; n is the number of atoms; V_i is the electrostatic potential at point i generated by the extended Mulliken point charge system (or for EP atomic charges the quantum chemical electrostatic potential calculated at point i); r_{ij} is the distance between point i and atom j; q_j is the charge of the atom j, q_{mol} is the charge of the molecule; and λ is the Lagrange multiplier. After elementary mathematical operations, a matrix equation is obtained. Finally, for calculating the atomic charges, only one matrix inversion must be performed. The remaining task is just checking the regularity of the matrix.[17] For the generation of points outside the van der Waals surface of the molecule, the same algorithm was used as detailed earlier.[9] Four shells were considered, taking the 1.4, 1.6, 1.8, and 2.0 times of atomic van der Waals radii,[18] respectively.[19] Gaussian orbitals (STO-5G[20]) were used for calculating the potential as well as the first-momentum integrals.

RESULTS AND DISCUSSION

From the results of MNDO-type methods (MNDO[12], AM1,[13] and PM3[21]), those of the PM3 calculations are discussed here in detail. The NDDO approximation was maintained during the calculation of various atomic charges (MP, EP, and PEP) because we found the SCF dipole moments obtained within this approximation to be closer to the experimental values than those calculated with deorthogonalized basis functions.[9,10]

For both the generation and the display of the initial geometry and for the semiempirical quantum chemical calculations, the PcMol program package was used.[22]

Full geometry optimization was performed for the nucleotide bases (adenine, guanine, thymine, cytosine, and uracil), in order to determine the ground-state equilibrium molecular geometries. Some data of the molecular geometry optimized by the PM3 method and the atomic numbering for the adenine molecule are seen in Figure 1. The SCF dipole moments calculated by various methods are shown in Table I.

Since our primary aim was studying how well the PEP atomic charges reproduce the exact quantum chemical electrostatic potential field outside the van der Waals surface of the molecules, the correlation between EP vs PEP atomic charges was investigated by linear regression (eq 3):

$$q(\mathrm{EP}) = \beta q(\mathrm{PEP}) \quad (3)$$

For the sake of comparison, linear regression analysis was also performed for the MP atomic charges. The values of the regression coefficients (β) with their estimated standard errors [SE(β)] and the correlation coefficients (r) are listed in Table II. The relations of q(EP) vs q(MP) and the q(EP) vs q(PEP) in the case of PM3 atomic charges are displayed in Figures 2 and 3.

For illustration, the results obtained by the CNDO/2[23] method are shown as well. For these calculations, molecular geometries determined by the AM1 method were used. The same approximation was applied with this method for calculating the atomic charges as with the MNDO-type methods.

The statistical data show clearly that the correlation is very good between EP and PEP atomic charges, while the MP charges rather poorly reproduce the quantum chemical electrostatic potential field outside the van der Waals surface of the molecule. MP charges calculated by the PM3 method result in the worst correlation. Of course, the good correlation between EP and PEP charges does not necessarily mean that each method describes the reality equally well. One thing seems to be certain; however, if EP charges give a good description of the molecules in question, then PEP charges

Table II. Statistical Data Obtained by Linear Regression for the q(EP) vs q(PEP) and q(EP) vs q(MP) Relations for Nucleotide Bases (I), Nucleotide Base Pairs (II), and Nucleosides (III)

	EP vs PEP charges			EP vs MP charges		
method	β	SE(β)	r	β	SE(β)	r
I						
MNDO	1.0112	0.0071	0.9983	2.2694	0.0899	0.9493
AM1	0.9925	0.0103	0.9962	1.8580	0.1037	0.9060
PM3	0.9778	0.0108	0.9957	1.6295	0.3031	0.5406
CNDO/2	1.0373	0.0100	0.9967	3.3540	0.1193	0.9585
II						
PM3	0.9213	0.0127	0.9920	1.6114	0.2459	0.5793
AM1	0.9678	0.0125	0.9930	1.7488	0.0920	0.8998
III						
PM3	0.9150	0.0182	0.9707	1.6581	0.1303	0.7147
AM1	0.9378	0.0150	0.9808	1.7217	0.0615	0.9138

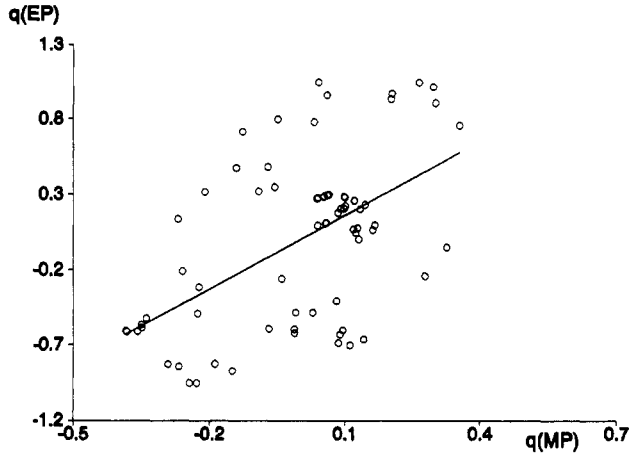

Figure 2. Graphical representation of the relation q(EP)-PM3 vs q(MP)-PM3 for nucleotide bases.

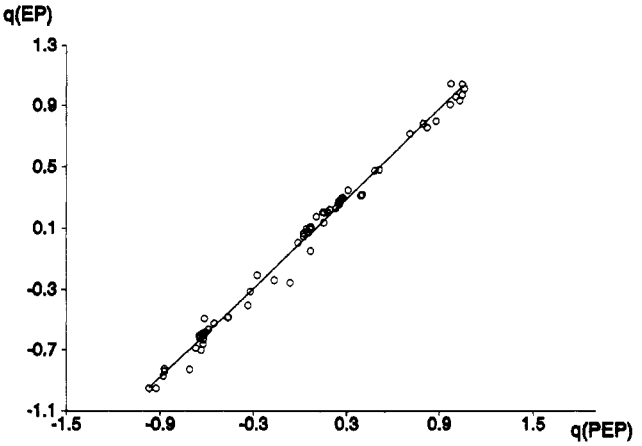

Figure 3. Graphical representation of the relation q(EP)-PM3 vs q(PEP)-PM3 for nucleotide bases.

Table III. Charge Distribution of Adenine Molecule Calculated by PM3 Method

atom	EP	PEP	MP
N1	-0.954	-0.971	-0.228
N3	-0.873	-0.880	-0.149
N6	-0.661	-0.624	0.142
N7	-0.589	-0.627	-0.068
N9	-0.237	-0.166	0.279

Table IV. EP Charges of Adenine Molecule Calculated by Various Methods

atom	CNDO/2	MNDO	AM1
N1	-1.089	-0.951	-0.795
N3	-1.116	-0.866	-0.795
N6	-0.617	-0.897	-0.611
N7	-0.851	-0.615	-0.516
N9	-0.668	-0.625	-0.536

with the regression coefficient as scaling factor can also be used.

Let us examine the charge distribution of the adenine molecule. The charges on the nitrogen atoms calculated by the PM3 method are listed in Table III. For comparison, EP charges calculated by CNDO/2, MNDO, and AM1 methods are displayed in Table IV.

It can be seen that, according to the PM3 calculations, the atoms N1 and N3 are the most negative; thus, in a first approximation, they should be the primary targets of an electrophilic attack. Experiments show that adenine is protonated on N1.[24] In alkylation reactions, however, N3 is more active. It is generally accepted that alkylation starts with an electrophilic attack.[25]

Table V. CPU Times (s) for Generating MEP Maps for Calculation of Atomic Charges

molecule	no. of points	PEP	EP
adenine	645	1.2	110.6
adenine-thymine	1027	3.5	329.3
adenosine	1117	4.1	365.2

The CNDO/2 and AM1 calculations also show that N1 and N3 are the most negative atoms in the molecule, although these methods practically do not discriminate between the nitrogen atoms as far as their negative charges are concerned. The MNDO method overestimates the negative charge on the nitrogen atom of the amino group. The nitrogen atoms N6 and N7 are less negative than N1 or N3 according to PM3 EP charges, and finally, N9 is the least negative of all. For the CNDO/2, MNDO, and AM1 methods the sequence is different.

Full geometry optimization and atomic charge calculation were performed for the nucleotide base pairs (adenine–thymine, guanine–cytosine, and adenine–uracil), too. The statistical data obtained for the PM3 and AM1 atomic charges are shown in Table II. It can be seen that the correlations between EP and PEP atomic charges are excellent. The MNDO method was not found suitable for describing the hydrogen bonded dimers: the geometries obtained through optimization are far from reality. The results of calculations performed by the PM3 and AM1 methods for the nucleoside molecules are also presented in Table II. The results reveal that PEP charges reproduce well the quantum chemical potential field in the surroundings of molecules in this case as well.

Finally, it seems worthwhile to consider CPU times of the calculations. CPU times for some molecules are displayed in Table V. It is to be seen that PEP charges can be calculated faster than EP charges. Consequently, the number of points taken outside the van der Waals surface can be increased considerably, and thus, the incidental symmetry restrictions for the atomic charges will automatically be satisfied.

ACKNOWLEDGMENT

Grants from the Hungarian Academy of Sciences (OTKA No. 616/91, F4297/92) and the Deutsche Forschungsgemeinschaft are gratefully acknowledged.

REFERENCES AND NOTES

(1) Politzer, P.; Daiker, K. C. *Models for Chemical Reactivity.* In *The Force Concept in Chemistry*; Deb, R. C., Ed.; Van Nostrand Reinhold Co.: New York, 1981; pp 294–387.
(2) Cox, S. R.; Williams, D. E. Representation of the Molecular Electrostatic Potential by a Net Atomic Charge Model. *J. Comput. Chem.* **1981**, *2*, 304–323.
(3) Carbó, R.; Calabuig, B. MOLSIMIL-88: Molecular Similarity Calculations Using a CNDO-like Approximation. *Comput. Phys. Commun.* **1989**, *55*, 117–126.
(4) Kikuchi, O.; Horikoshi, K.; Takahashi, O. Rapid Evaluation of Molecular Electrostatic Potential Maps by Simple Analytical Functions. *J. Mol. Struct. (Theochem)* **1992**, *256*, 47–60.
(5) Gadre, S. R.; Bapat, S. V.; Sundararajan, K.; Shrivastava, I. A. General Parallel Algorithm for the Generation of Molecular Electrostatic Potential Maps. *Chem. Phys. Lett.* **1990**, *175*, 307–312.
(6) Giessner-Prettre, G.; Pullman, A. On the Molecular Electrostatic Potentials Obtained with CNDO and INDO Wave Functions. *Theor. Chim. Acta* **1974**, *33*, 91–94.
(7) Mulliken, R. S. Electronic Population Analysis on LCAO-MO Molecular Wave Functions. *J. Chem. Phys.* **1955**, *23*, 1833–1846.
(8) Momany, F. A. Determination of Partial Atomic Charges from Ab Initio Molecular Electrostatic Potentials. Applications to Formamide, Methanol, and Formic Acid. *J. Phys. Chem.* **1978**, *82*, 592–601.
(9) Tasi, G.; Kiricsi, I.; Förster, H. Representation of Molecules by Atomic Charges: a New Population Analysis. *J. Comput. Chem.* **1992**, *13*, 371–379.

(10) Tasi, G.; Kiricsi, I.; Förster, H. Atomi töltések számítása MNDO módszerrel. *Magy. Kem. Foly.* **1991**, *97*, 441–449.
(11) Huzinaga, S.; Sakai, Y.; Miyoshi, E.; Narita, S. Extended Mulliken Electron Population Analysis. *J. Chem. Phys.* **1990**, *93*, 3319–3325.
(12) Dewar, M. J. S.; Thiel, W. Ground States of Molecules. 38. The MNDO Method. Approximations and Parameters. *J. Am. Chem. Soc.* **1977**, *99*, 4899–4907.
(13) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: a New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
(14) (a) Luque, F. J.; Illas, F.; Orozco, M. Comparative Study of the Molecular Electrostatic Potential Obtained from Different Wavefunctions. Reliability of the Semiempirical MNDO Wavefunction. *J. Comput. Chem.* **1990**, *11*, 416–430. (b) Orozco, M.; Luque, F. J. On the Use of AM1 and MNDO Wave Functions To Compute Accurate Electrostatic Charges. *J. Comput. Chem.* **1990**, *11*, 909–923.
(15) Orozco, M.; Luque, F. J. A Practical Procedure for the Determination of Electrostatic Charges of Large Molecules. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 411–426.
(16) Ferenczy, G. G.; Reynolds, C. A.; Richards, W. G. Semiempirical AM1 Electrostatic Potentials and AM1 Electrostatic Potential Derived Charges: a Comparison with ab Initio Values. *J. Comput. Chem.* **1990**, *11*, 159–169.
(17) Williams, D. E. Net Atomic Charge and Multipole Models for the ab Initio Molecular Electric Potential. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, **1991**; Vol. 2, pp 219–271.
(18) The values of van der Waals radii (nm) used in this work: H, 0.1185; C, 0.175; N, 0.1525; O, 0.14.
(19) Singh, U. C.; Kollman, P. A. An Approach to Computing Electrostatic Charges for Molecules. *J. Comput. Chem.* **1984**, *5*, 129–145.
(20) Poirier, R.; Kari, R.; Csizmadia, I. G. *Handbook of Gaussian Basis Sets*; Elsevier: Amsterdam, 1985; p 120.
(21) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods. I. Method. *J. Comput. Chem.* **1989**, *10*, 209–220.
(22) Tasi, G.; Pálinkó, I.; Halász, J.; Náray-Szabó, G. *Semiempirical Quantum Chemical Calculations on Microcomputers*; CheMicro Limited: Budapest, 1992.
(23) Pople, J. A.; Segal, G. A. Approximate Self-Consistent Molecular Orbital Theory. III. CNDO Results for AB, and AB_2 Systems. *J. Chem. Phys.* **1966**, *44*, 3289–3296.
(24) Kistenmacher, T. J.; Shigematsu, T. The Crystal Structure of Adenine Dihidrochloride. *Acta Crystallogr.* **1974**, *B30*, 1528–1533.
(25) Pal, B. C. Studies on the Alkylation of Purines and Pyrimidines. *Biochemistry* **1962**, *1*, 558–563.