Atomic Charges Derived from Semiempirical Methods

Brent H. Besler

Department of Chemistry, University of California, Berkeley, California 94720

Kenneth M. Merz, Jr. and Peter A. Kollman*

Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143

It is demonstrated that semiempirical methods give electrostatic potential (ESP) derived atomic point charges that are in reasonable agreement with ab initio ESP charges. Furthermore, we find that MNDO ESP charges are superior to AM1 ESP charges in correlating with ESP charges derived from the 6-31G* basis set. Thus, it is possible to obtain 6-31G* quality point charges by simply scaling MNDO ESP charges. The charges are scaled in a linear ($y = Mx$) manner to conserve charge. In this way researchers desiring to carry out force field simulations or minimizations can obtain charges by using MNDO, which requires much less computer time than the corresponding 6-31G* calculation.

INTRODUCTION

Many potential functions used in the simulation and the energy minimization of biological molecules (e.g., proteins and nucleic acids) have many characteristics in common. They are pairwise potentials that contain terms of the following form: Bond stretching and bending are modeled with a quadratic harmonic oscillator terms while dihedral torsions are represented with a finite Fourier series. Nonbonded van der Waals are incorporated via a Lennard–Jones 6-12 term and electrostatic interactions are usually represented by an atom-centered monopole expression (Coulomb term). Potentials which share these features include the AMBER,[1] CHARMM,[2] and GROMOS[3] force fields. Similar potentials have been used by Hagler[4] and Scheraga.[5]

The electrostatic term is key to an accurate description of hydrogen bonding in potentials of this type. Along with the 6-12 term, it determines the intermolecular interactions. Thus, it is critically important to model the electrostatic interaction as accurately as possible. It is a difficult task to assign charges based on any type of experimental data,[6] therefore they must be determined by some theoretical technique. The basis set Hartree–Fock method provides a straightforward way of computing molecular properties, but atomic charge is a quantity which is not rigorously defined in quantum mechanics.[7] Two common techniques have been used to extract

atomic charges from approximate HF wave functions: Mulliken population analysis is a computationally simple method,[7] but these charges have proven unsatisfactory.[8] The second method, deriving the charges by fitting the molecular electrostatic potential (ESP fit method), has yielded much more promising results.[9] For example, the dipole and quadrupole moments calculated using these charges compare favorably to the corresponding experimental gas phase quantities.[9] In protein environments and in aqueous solution, polarization effects will increase the effective dipole over the gas phase, thus it may be desirable to employ a quantum mechanical procedure that overestimates the gas-phase dipole moments by ca. 10–20%.[10] The 6-31G* basis set of Pople et al.[11] satisfies this criterion (c.f. Table 2), which makes it useful for determining atomic charges to be used in force fields. Unfortunately, 6-31G* and other large basis set HF calculations can quickly overwhelm any available computational and disk storage resources due to the calculation and storage of $O(N^4)$ two-electron integrals, where N is the number of basis functions. Calculation of a set of atomic charges at a fixed molecular geometry via the ESP fit method with the 6-31G* basis are only practical for molecules with ca. 12 nonhydrogen atoms. It would therefore be desirable to find a method which will produce 6-31G* quality results at a smaller computational cost.

Semiempirical methods such as the CNDO[12] and INDO[13] methods of Pople and the MNDO[14] and AM1[15] methods of Dewar and co-workers scale as $O(N^3)$ and therefore represent attractive

---

*Author to whom correspondence should be addressed.

alternative methods by which the ESP fit charges can be determined. Giessner–Prettre and Pullman[16] showed that CNDO and INDO could yield HF quality molecular ESP surfaces. Later, Culberson and Zerner[17] demonstrated the same for the INDO/S semiempirical method. However, neither of these articles reported fitting the resultant ESP surfaces in order to determine atomic point charges. Thus, it is currently unknown if semiempirically derived ESP surfaces can give reasonable atomic charges. Furthermore, it is logical to assume that MNDO and AM1, would give better ESP surfaces than CNDO and INDO and therefore have a better chance to yield reasonable point charges. This assumption is warranted because AM1 and MNDO do predict properties such as heats or formation, dipole moments, molecular geometries, and ionization potentials which are in much better agreement with experiment than INDO and CNDO.[12-15] Given these facts we therefore decided to explore the use of AM1 and MNDO for the calculation of atomic charges derived from ESP fitting.

COMPUTATIONAL DETAILS

It is possible to compute the expectation value of the electrostatic potential from a linear combination of atomic orbitals (LCAO) type approximate wave function via two different methods: by direct calculation from the wave function[18] or by calculation of the charge density from the wave function and integration of the Poisson equation.[19] For a wave function consisting of cartesian Gaussians, Srebnik et al.[19] showed that the Poisson equation may be integrated analytically to determine the ESP. Klöpper and Volkmer[20] demonstrated that calculation of the expectation value directly yields the same result as integration of the Poisson equation for wave functions of this type. We have chosen to use the former approach, but either approach can be used.

For a one-electron operator Θ, such as the electrostatic potential operator, the expectation value from an LCAO wave function will have the form:

$$\Theta(r) = \Theta_{cl} + \sum_{ij} P_{ij}(\psi_i|\Theta|\psi_j) \qquad (1)$$

where Θ_{cl} is the classical ESP calculated from the charges on the nuclei, P_{ij} is the density matrix and i, j are AO indices. For the electrostatic potential this is:

$$V(r) = \sum_A \frac{Z_A}{|r - R_A|} - \sum_{ij} P_{ij} \int \frac{\Psi_i \Psi_j}{|r - r'|} dr' \qquad (2)$$

where the Z_A are the nuclear charges, R_A is the nuclear position, r is the point at which the ESP

is being evaluated, and r' is a dummy integration variable. Calculation of the expectation value at a single point in space will require the computation of $(N^2 + N)/2$ electrostatic potential integrals, where N is the number of contracted Gaussian basis functions. The integrals are invariant with respect to interchange of the indices i and j and they have the same mathematical form as the nuclear attraction integrals involved in the calculation of the electronic energy.[7] An efficient scheme for the computation of various type of integrals involving Cartesian Gaussian functions, including the nuclear attraction type has recently been derived by Obara and Saika.[21] Their method is based on a series of recursion relations which allow the integrals to be expressed as linear combinations of exponential functions and the functions $F_m(T)$ where:

$$F_m(T) = \int_0^1 dt\ t^{2m} \exp(-Tt^2) \qquad (3)$$

This method has the dual advantage of having a lower number of computational steps than methods such as the Rys polynomial method[22] or the McMurchie and Davidson method[23] and also is highly vectorizable. For nuclear attraction type integrals most of the computation time is involved in evaluation of the functions $F_m(T)$. Obara and Saika present a particularly efficient method for the evaluation of $F_m(T)$ based upon a seven-term Taylor series interpolation.[21]

Semiempirical LCAO methods such as CNDO, INDO, MNDO, and AM1 lack an explicit analytic form for their corresponding wave functions, unlike the basis set Hartree–Fock method. Based on comparisons with ESP surfaces computed from *ab initio* Hartree–Fock wave functions, Giessner–Pettre and Pullman[16] demonstrated that 'deorthogonalization' of the semiempirical density matrix led to computed ESP surfaces in reasonable agreement with the *ab initio* ones. This method for circumventing the analytic form problem was first suggested by Schillady et al. for studying charges and dipole moments.[24] It was also chosen by Culberson and Zerner[17] for computation of ESP expectation values from the INDO/S wavefunction.

In this "deorthogonalization" technique, the semiempirical molecular orbital (MO) coefficients are conceived of as being in some orthonormal basis for which there is no analytic form, but which has the dimensionality of a valence minimal Slater orbital basis set. It may be formed from some nonorthonormal basis of the same dimensionality via a technique such a Löwdin, canonical, or Grahm–Schmidt orthogonalization.[7] In Löwdin orthogonalization the orthonormal

basis set Φ is obtained from the nonorthonormal basis Ψ via the transformation:

$$\Phi_i = \sum_j \Psi_j S_{ij}^{-1/2} \qquad (4)$$

$S_{ij}^{-1/2}$ is the inverse square root of the overlap matrix of the nonorthonormal basis Ψ. The transformation may be inverted and the orthonormal eigenvectors may be "deorthogonalized," expressed in the new basis Ψ, and a new density matrix the computed in the basis Ψ:

$$C^{\Psi} = S^{-1/2}C^{\Phi} \qquad (5)$$

$$P_{ij} = 2\sum_k^{occ.} C_{ik}^* C_{kj} \qquad (6)$$

where the Cs are the MO coefficients. The expectation value of the ESP may then be computed using eq. (2).

To provide an adequate sampling of the ESP around the molecule to which the charges will be fit it is desirable to use a procedure which is well defined for all molecules and will produce a uniform surface of points at which the ESP is to be evaluated. Such a method is the surface generation algorithm of Connolly.[25] It computes a spherical surface of points around each atom at a specified multiple of the atoms' van der Waals radius and density. The molecular surface is then constructed by taking the union of all of the atom surfaces and eliminating those points that are within the specified multiple of the van der Waals radius of any of the atoms. Singh and Kollman[26] examined the use of various numbers of these surfaces and densities. They determined that the use of four surfaces with using shells of 1.4, 1.6, 1.8, and 2.0 times the van der Waals radii, with a density of one to five points per square angstrom provides an adequate sampling to which the fitted charges are relatively insensitive. When using the method we have described to compute ESPs with a semiempirical method, it is important to minimize the number of sample points since most of the total computation time involves calculation of the ESP integrals. We therefore use a density of one point per square angstrom in this study. The derived charges are rather insensitive to the position of the electrostatic potential points, provided enough points are used and these are outside van der Waals envelope of the molecule.

Atom centered monopole charges are least-squares fit to reproduce the quantum mechanically calculated ESPs subject to any desired constraints. It is standard procedure to constrain the charges such that their sum is equal to the total charge on the molecule, since otherwise physically unrealistic, interactions may result

from such a charge distribution. We use the linear least squares fit procedure suggested by Chirlian and Francl[27] in which any constraints are imposed via Lagrange multipliers. It avoids many of the disadvantages of nonlinear least squares fit procedures such as the need for initial guess charges, iterative solution procedures, and possible convergence problems. Nonetheless, we emphasize that such a linear procedure precludes the use of off-center charges whose position could be optimized in the fit. The procedure described in reference 29 does not have this limitation. The derivation of the least squares fit equations is presented in the appendix.

Finally, a basis set has to be chosen to be used in the deorthogonalization procedure and in the evaluation of the nuclear attraction. We decided upon the STO-6G basis set of Pople et al.[28] This basis set was employed because it is a minimal basis set and can be easily fit into the minimal valence only basis used in the semiempirical methods. Furthermore, STO-6G was chosen over STO-3G because it is a larger basis set and should give a better representation of the ESP at a minimal increase in computational cost.

RESULTS AND DISCUSSION

In order to accurately compare the semiempirical and *ab initio* charges we have used the same experimental geometries,[14] surface (*i.e.*, identical number of points and coordinates) and fitting algorithm throughout.

Table I gives the atomic point charges derived using the MNDO and AM1 methods along with STO-3G[28] and 6-31G* charges calculated using Gaussian 80 UCSF[29] for comparison. In the table are also the relative root mean square fits (RRMS), as defined in Table I, of the quantum mechanically calculated ESP values to the classical model. From this table it is readily seen that the semiempirical charges are usually smaller in magnitude to the 6-31G* ones, but agree in sign and relative magnitude. All of the semiempirical RRMS values are of the same magnitude as the corresponding ab inito ones, demonstrating that the semiempirical ESP surface is fit well by a classical model. To determine how well AM1 and MNDO reproduce *ab initio* results, we have least squares fit the semiempirical charges to the corresponding *ab initio* ones using a linear fit without a y-intercept ($y = Mx$) to preserve the total molecular charge if the scaled charges were used. The results are summarized in Table II. A synopsis of the results is as follows: AM1 ESP charges fit the STO-3G and 6-31G* ESP charges with correlation coefficients of 0.863 and 0.886, re-

Table I. Calculated MNDO, AM1, STO-3G, and 6-31G* electrostatic potential charges. (charges in parenthesis are from the "Coulson" method).

Molecule[a]	MNDO	AM1	STO-3G	6-31G*
NH$_3$				
N	−0.711 (−0.246)	−0.743 (−0.362)	−0.974	−1.091
H	+0.237 (+0.082)	+0.248 (+0.121)	+0.325	+0.363
RRMS[b]	+0.062	+0.056	+0.077	+0.135
CH$_4$				
C	−0.147 (−0.018)	−0.477 (−0.263)	−0.458	−0.490
H	+0.037 (+0.072)	+0.119 (+0.066)	+0.114	+0.123
RRMS	+0.469	+0.20 +0.202	+0.264	
Acetylene				
C	−0.263 (−0.157)	−0.341 (−0.219)	−0.180	−0.298
H	+0.263 (+0.157)	+0.341 (+0.219)	+0.180	+0.298
RRMS	+0.023	+0.020	+0.043	+0.053
Ethylene				
C	−0.212 (−0.088)	−0.344 (−0.224)	−0.170	−0.360
H	+0.106 (+0.044)	+0.172 (+0.112)	+0.085	+0.180
RRMS	+0.204	+0.150	+0.298	+0.159
CH$_3$OH				
C	+0.109 (+0.191)	−0.244 (−0.062)	−0.019	+0.103
O	−0.462 (−0.340)	−0.332 (−0.343)	−0.458	−0.657
H(anti)	+0.059 (+0.022)	+0.162 (+0.102)	+0.070	+0.088
H(gauche)	+0.003 (−0.022)	+0.085 (+0.053)	+0.049	+0.022
H(hydroxyl)	+0.289 (+0.171)	+0.244 (+0.196)	+0.309	+0.422
RRMS	+0.141	+0.157	+0.175	+0.154
H$_2$NCOH				
C	+0.533 (+0.370)	+0.231 (+0.267)	+0.589	+0.736
N	−0.781 (−0.437)	−0.716 (−0.475)	−0.748	−1.016
O	−0.445 (−0.368)	−0.314 (−0.355)	−0.419	−0.594
H(trans)	+0.341 (+0.199)	+0.339 (+0.231)	+0.327	+0.405
H(carbon)	+0.312 (+0.186)	+0.322 (+0.218)	+0.286	+0.457
H(cis)	+0.040 (+0.049)	+0.138 (+0.115)	−0.036	+0.013
RRMS	+0.064	+0.086	+0.101	+0.071
Formic acid				
H	+0.060 (+0.081)	+0.184 (+0.149)	−0.050	+0.057
C	+0.533 (+0.349)	+0.162 (+0.263)	+0.664	+0.682
O(carbonyl)	−0.450 (−0.354)	−0.304 (−0.341)	−0.426	−0.569
O	−0.476 (−0.272)	−0.316 (−0.295)	−0.565	−0.642
H(hydroxyl)	+0.332 (+0.196)	+0.273 (+0.225)	+0.377	+0.472
RRMS	+0.108	+0.129	+0.157	+0.093
Acetonitrile				
C	−0.158 (−0.118)	−0.376 (−0.127)	−0.529	−0.467
H	+0.086 (+0.021)	+0.167 (+0.108)	+0.162	+0.171
C(nitrile)	+0.129 (+0.136)	+0.028 (−0.148)	+0.489	+0.464
N	−0.231 (−0.081)	−0.153 (−0.050)	−0.447	−0.512
RRMS	+0.041	+0.058	+0.056	+0.037
Formaldehyde				
H	+0.006 (+0.015)	+0.081 (+0.078)	−0.003	+0.040
C	+0.339 (+0.273)	+0.081 (+0.130)	+0.266	+0.364
O	−0.352 (−0.303)	−0.244 (−0.286)	−0.260	−0.443
RRMS	+0.106	+0.113	+0.231	+0.087
CO$_2$				
C	+0.642 (+0.447)	+0.280 (+0.410)	+0.807	+0.884
O	−0.321 (−0.223)	−0.140 (−0.205)	−0.403	−0.442
RRMS	+0.056	+0.306	+0.015	+0.087
CO				
C	+0.087 (+0.185)	+0.025 (+0.167)	+0.019	+0.074
O	−0.087 (−0.185)	−0.025 (−0.167)	−0.019	−0.074
RRMS	+0.795	+0.960	+0.990	+0.864
H$_2$O				
O	−0.551 (−0.318)	−0.455 (−0.385)	−0.616	−0.808
H	+0.275 (+0.159)	+0.227 (+0.193)	+0.308	+0.404
RRMS	+0.050	+0.154	+0.041	+0.115

Table I. *(Continued)*

Molecule[a]	MNDO	Method AM1	STO-3G	6-31G*
Benzene				
C	−0.093 (−0.059)	−0.156 (−0.129)	−0.056	−0.124
H	+0.093 (+0.059)	+0.156 (+0.129)	+0.056	+0.124
RRMS	+0.124	+0.102	+0.184	+0.172
Trimethylammonium[c]				
C	−0.194 (+0.077)	−0.534 (−0.198)	−0.393	−0.304
N	+0.176 (−0.061)	+0.388 (+0.008)	+0.541	+0.148
H(nitrogen)	+0.264 (+0.206)	+0.247 (+0.246)	+0.209	+0.296
H(anti)	+0.125 (+0.070)	+0.207 (+0.148)	+0.147	+0.142
H(gauche)	+0.135 (+0.069)	+0.224 (+0.149)	+0.165	+0.166
RRMS	+0.005	+0.066	+0.006	+0.011
CH_3NO_2				
C	−0.197 (+0.087)	−0.373 (−0.202)	−0.456	−0.439
N	+0.655 (+0.432)	+0.491 (+0.515)	+0.864	+0.857
O	−0.405 (−0.337)	−0.328 (−0.359)	−0.435	−0.479
H	+0.117 (+0.052)	+0.180 (+0.135)	+0.154	+0.180
RRMS	+0.045	+0.085	+0.050	+0.069
Malononitrile[c]				
C	−0.036 (−0.120)	−0.111 (+0.025)	−0.629	−0.304
C(nitrogen)	+0.102 (+0.220)	−0.015 (−0.151)	+0.525	+0.415
N	−0.188 (−0.044)	−0.102 (−0.007)	−0.422	−0.445
H	+0.104 (+0.055)	+0.173 (+0.145)	+0.211	+0.182
RRMS	+0.054	+0.078	+0.062	+0.045
SH_2[c]				
S	−0.112 (+0.064)[e]		−0.261	−0.372
H	+0.056 (−0.032)		+0.130	+0.186
RRMS	+0.602		+0.213	+0.406
CH_3SH[c]				
C	−0.061 (−0.030)		−0.058	−0.155
S	−0.159 (+0.015)		−0.211	−0.365
H(anti)	+0.079 (+0.029)		+0.060	+0.076
H(sulfur)	+0.084 (−0.027)		+0.134	+0.217
H(gauche)	+0.028 (+0.006)		+0.038	+0.152
RRMS	+0.378		+0.323	+0.322
$(CH_3)_2S$[d]				
C	−0.163 (−0.012)		−0.214	−0.321
S	−0.127 (−0.040)		−0.089	−0.236
H(anti)	+0.088 (+0.026)		+0.084	+0.163
H(gauche)	+0.069 (+0.003)		+0.087	+0.138
RRMS	+0.283		+0.380	+0.306
$(CH_3)_2S_2$[d]				
C	−0.100 (−0.025)		−0.179	−0.242
S	−0.080 (−0.027)		−0.054	−0.130
H(anti)	+0.083 (+0.029)		+0.078	+0.149
H(gauche)	+0.049 (+0.011)		+0.077	+0.112
RRMS	+0.253		+0.385	+0.285

[a]All structures taken from reference 13, except where noted.
[b]RRMS (Relative Root Mean Square Fit):

$$RRMS = \left(\frac{\sum_{i=1}^{m} (V_i^{QM} - V_i^{CL})^2}{\sum_{i=1}^{m} (V_i^{QM})^2} \right)^{1/2}$$

[c]The AM1 optimized geometry was used.
[d]Sulfur compound structures from M. J. S. Dewar and M. L. McKee *J. Compt. Chem.*, **4**, 84 (1983).
[e]AM1 parameters do not exist for Sulfur.

Table II.   Linear Fit Crossreference for Charges ($y = Mx$).

Fits including all molecules:

	MNDO ESP	AM1 ESP	MNDO "Coulson"	AM1 "Coulson"
dep. variable				
6-31G*				
Corr. Coeff.	0.96975	0.88572	0.86726	0.88437
Scaling Con.	1.42048	1.38593	1.95457	1.78807
STO-3G				
Corr. Coeff.	0.91962	0.86347	0.80857	0.80431
Scaling Con.	1.18217	1.20505	1.55903	1.45063

6-31G* vs STO-3G with 6-31G* as dep. variable:

Corr. Coeff.	0.96463
Scaling Con.	1.09946

Fits including only heteroatoms (O, N, S) and hydrogens bonded to them:

	MNDO ESP	AM1 ESP
dep. variable		
6-31G*		
Corr. Coeff.	0.98347	0.95280
Scaling Con.	1.39731	1.55001
STO-3G		
Corr. Coeff.	0.96363	0.96054
Scaling Con.	1.15158	1.32819

Fits including only carbons and hydrogens bonded to carbon:

	MNDO ESP	AM1 ESP
dep. variable		
6-31G*		
Corr. Coeff.	0.93809	0.7426
Scaling Con.	1.47584	1.13141
STO-3G		
Corr. Coeff.	0.84570	0.70198
Scaling Con.	1.25027	1.01403

spectively. MNDO on the other hand, has corresponding correlation coefficients of 0.929 and 0.969. Furthermore, we have fit the charges determined directly from the MNDO and AM1 wavefunction using the method described by Pople and Beveridge[30] (referred to by some as the "Coulson" method) with 6-31G* ESP charges and found correlation coefficients of 0.867 for MNDO and 0.884 for AM1. For comparison, STO-3G charges fit the 6-31G* ones with a correlation coefficient of 0.965. Thus, from our data set we find that the MNDO is superior to AM1 and to a lesser extent STO-3G in its ability to reproduce 6-31G* charges. The scaling equation for MNDO vs. 6-31G* is 6-31G* = 1.422*MNDO. If the MNDO and AM1 ESP charges are fit separately for the set of heteroatoms and hydrogens bonded to heteroatoms (those capable of forming hydrogen bonds) and the set of carbons and nonheteroatom bonded hydrogens the results are: For the former, both the MNDO and AM1 charges correlate well with the 6-31G* ones (correlation coefficients 0.983 and 0.952 respectively). For the later set, the MNDO charges correlate well

(0.936), while AM1 does very poorly (0.742). The fits of the semiempirical charges with the STO-3G ESP charges have the same behavior as the 6-31G* ones, but the correlation coefficients are somewhat lower, and the scaling constant is markedly smaller. Thus this set of atoms drastically alters the fit for AM1 when it is done for all atoms. While the AM1 method generally reproduces experimental quantities such as dipole moments better than MNDO,[14] the "deorthogonalization" technique, which has been used to extract ESP's from the semiempirical wavefunctions, requires an "*ab initio* like" density matrix to produce ESP's in qualitative agreement with those from *ab initio* calculations. Thus it is our expectation that MNDO gives a density matrix that is similar to that from a minimal basis Slater orbital wavefunction whereas AM1 does not.

It is worth examining the scaled MNDO charges in more detail, to see where they reproduce the 6-31G* charges well and where not so well. For example, the charges on H_2O and NH_3 are reproduced to better than 10%, as are the charges on the hydroxyl hydrogen and oxygen of formic acid.

Table III. Calculated MNDO, AM1, STO-3G, and 6-31G* dipole moments (Debye).

Molecule	METHOD				
	MNDO	AM1	STO-3G	6-31G*	Exp.[a]
NH_3	1.740	1.927	1.786	1.949	1.47
CH_3OH	1.619	1.666	1.503	1.940	1.70
CO	-0.020[b]	-0.180	-0.169	-0.331	0.11
H_2O	1.790	1.854	1.726	2.225	1.85
CH_3NO_2	4.019	4.166	3.230	4.273	—[c]
Malononitrile	2.696	2.906	3.135	4.078	—
H_2NCOH	3.207	3.313	2.579	3.887	3.73
Formic Acid	1.290	1.304	0.777	1.570	1.41
Acetonitrile	2.651	2.881	3.063	4.064	3.92
Formaldehyde	2.237	2.352	1.510	2.746	2.33
SH_2	1.439	—[c]	1.076	1.427	0.97
CH_3SH	1.627	—	1.062	1.832	1.52
$(CH_3)_2S$	1.755	—	1.016	1.944	1.55
$(CH_3)_2S_2$	2.172	—	1.231	2.385	—
Corr. Coeff.	0.93	0.94	0.93	0.98	
Corr. Coeff. excluding S compounds	0.95	—	0.94	0.98	

[a]Dipole moments taken from reference 13, except for sulfur compounds which were from M. J. S. Dewar and M. L. McKee *J. Comp. Chem.*, **4**, 84 (1983).

[b]The dipole moments for CO are listed as negative since the calculated direction is opposite to that found experimentally.

[c]Experimental values do not exist for entries marked with a —.

[d]AM1 parameters do not exist for Sulfur.

The scaled MNDO charges of the formic acid carbon and oxygen are ca. 10% larger than those of 6-31G*. On the other hand, the scaled MNDO charges of acetonitrite are significantly smaller than the 6-31G* charges. Thus, not all charges are well reproduced by this scaling procedure with a single, average scale factor. Nonetheless, one should not lose sight of the fact that the main use of such charges will be in molecular mechanics and dynamics methods,[1-5] where it is most critical that they reproduce the properties of intermolecular interactions of the various functional groups.

Even though the scaled MNDO acetonitrite N is significantly less negative than the 6-31G* atom, the neighboring carbon is also less positive. Thus, for example, the H-bonding properties of these models are likely to be more similar than suggested by the nitrogen charge alone. It is important that these issues be addressed before such charges are actually used in molecular mechanical models.

Table III gives the dipole moment determined from MNDO, AM1, STO-3G*, and 6-31G* wavefunctions, while Table IV reports the dipole moments determined from the derived point charges for MNDO and AM1. From Table III one finds that all of these methods give dipole moments that are in reasonable agreement with the experimental value. MNDO, AM1, and STO-3G

Table IV. Calculated MNDO and AM1 dipole moments derived from charges (Debye).

Molecule	METHOD		
	MNDO	AM1	Exp.[a]
NH_3	1.301	1.360	1.47
CH_3OH	1.295	1.009	1.70
CO	-0.470[b]	-0.138	0.11
H_2O	1.550	1.281	1.85
CH_3NO_2	3.948	3.924	—[c]
Malononitrile	2.489	2.646	—
H_2NCOH	2.945	2.753	3.73
Formic Acid	1.324	1.386	1.41
Acetonitrile	2.450	2.610	3.92
Formaldehyde	2.070	1.930	2.33
SH_2	0.476	—[d]	0.97
CH_3SH	0.839	—	1.52
$(CH_3)_2S$	0.999	—	1.55
$(CH_3)_2S_2$	1.329	—	—
Corr. Coeff.	0.91	0.97	—
Corr. Coeff. excluding S compounds	0.94	—	—

[a]Dipole moments taken from reference 13, except for sulfur compounds which were from M. J. S. Dewar and M. L. McKee *J. Compt. Chem.*, **4**, 84 (1983).

[b]The dipole moments for CO are listed as negative since the calculated direction is opposite to that found experimentally.

[c]Experimental values do not exist for entries marked with a —.

[d]AM1 parameters do not exist for Sulfur.

give dipole moments that can be either larger or smaller than the experimental value, while 6-31G* gives, as expected, dipole moments that are uniformly larger than the experimental value. The data in Table IV shows that the calculated semiempirical ESP charges give dipole moments that are in reasonable agreement with experiment, as was the case for *ab initio* methods.[9,26] From Table III we also observe that in general the dipole moments determined from the ESP charges are smaller than the experimental value (the only exception to this CO). In Tables III, IV we have also included the correlation coefficients between the experimental dipole moments and the calculated ones. We find that the 6-31G* calculated dipoles correlate well with the experimental dipoles (0.98), while dipole moments determined from semiempirical and STO-3G wavefunctions do not correlate quite as well (0.95 to 0.93). Interestingly, the AM1 dipole moments determined form ESP charges correlate better with the experimental values (0.97 vs. 0.94). MNDO dipole moments determined from ESP charges, on the other hand, do not correlate quite as well (0.94). It is possible to improve the agreement between the point charge derived dipole moment and the experimental one by forcing the charges to also fit the observed dipole moment. This, however, requires the experimental knowledge of the X, Y, Z components of the dipole moment for the molecule in question. Alternatively, one could use the results from semiempirical or *ab initio* methods.

CONCLUSIONS

Herein, we have demonstrated that semiempirical methods can be expected to give reasonable ESP derived charges, which, when scaled, should in many cases yield 6-31G* quality ESP charges at a small fraction of the cost. MNDO is found to be superior to AM1 and to a lesser extent to STO-3G in this regard. We have carried out some test calculations where we compared the speed of our MNDO ESP calculations to the corresponding calculations with STO-3G and 6-31G* ones. Of course, the SCF calculations themselves are much faster for MNDO and AM1 than for the *ab initio* calculations, but for our current algorithm for the ESP determination we find that STO-3G is faster than MNDO and 6-31G* is slower than both MNDO and STO-3G. For large molecules, the semiempirical methods will always be faster than the corresponding than an *ab inito* calculations.[51] The vast majority (>90%) of the total time in the semiempirical ESP calculation is spent evaluating the ESP integrals. Improvements in the ESP integral code should

result in a substantial speed increase. We also plan to try using STO-3G instead of STO-6G, which if the results are comparable should again decrease our computational cost by a significant amount. The main savings involved in using MNDO currently is the reduced disk capacity needed for large molecules (molecules with greater than 20 heavy atoms). The modified version of MOPAC 5.0 has been submitted to QCPE. Further development of these charge models for incorporation into molecular mechanics/dynamics approaches is in progress.

APPENDIX A

Derivation of linear least squares fit equations for n monopole charges (q) to be fit to m ESP values:

The least square fit criterion is given by finding the minimum of the function:

$$\gamma(q_1, q_2, \ldots, q_n) = \sum_{i=1}^{m} (V_i - E_i)^2$$

V_i is the quantum mechanically calculated ESP at point i
E_i is the ESP at point i calculated classically from the charges q_j.

$$E_i \approx \sum_{j=1}^{n} \frac{q_j}{r_{ij}}$$

Any constraints (g) on the charges may be imposed via Lagrange multipliers (λ). The constraints must be a function of the charge values only to keep the fit equations linear. The function to be minimized with w different constraints imposed is:

$$z = \gamma + \lambda_1 g_1 + \lambda_2 g_2 + \cdots + \lambda_w g_w$$

The minimum of z and the corresponding qs may be found by solving the $(n + w)$ by $(n + w)$ system of equations:

$$\sum_{k=1}^{n} \frac{\partial z}{\partial q_k} = 0 \quad \text{and} \quad \sum_{l=1}^{w} \frac{\partial z}{\partial \lambda_1}$$

The single constraint used in this study is the sum of the charges must equal the total charge on the molecule. Expressed in the notation above the corresponding constraint function is:

$$g_1 = \sum_{j=1}^{n} q_j - q_{\text{tot}} = 0$$

We proceed to derive the system of equations and its solution explicitly in matrix form:

$$z = \sum_{i=1}^{m} \left(V_i - \sum_{j=1}^{n} \frac{q_j}{r_{ij}} \right)^2 + \lambda \left(\sum_{j=1}^{n} q_j - q_{\text{tot}} \right)$$

$$\frac{\partial z}{\partial \lambda} = 0 = \sum_{j=1}^{n} q_j - q_{\text{tot}}$$

$$\frac{\partial z}{\partial q_k} = 0 = \sum_{i=1}^{m} \frac{2}{r_{ik}} \left(V_i - \sum_{j=1}^{n} \frac{q_j}{r_{ij}} \right) + \lambda$$

(Since λ is an essentially arbitrary constant it is left as $+\lambda$ in the algebra) Simplifying,

$$\sum_{j=1}^{n} q_j = q_{\text{tot}}$$

$$\sum_{j=1}^{n} \sum_{i=1}^{m} \frac{1}{r_{ij} r_{ik}} = \sum_{i=1}^{m} \frac{V_i}{r_{ik}} + \lambda$$

Now defining,

$$A_{jk} = \sum_{i=1}^{m} \frac{1}{r_{ij} r_{ik}} \quad \text{and} \quad B_k = \sum_{i=1}^{m} \frac{V_i}{r_{ik}}$$

yields a matrix equation:

$$\begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} & 1 \\ A_{21} & A_{22} & \cdots & A_{2n} & 1 \\ \vdots & \vdots & \ddots & \vdots & 1 \\ A_{n1} & A_{n2} & & A_{nn} & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \\ \lambda \end{bmatrix} = \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_n \\ q_{\text{tot}} \end{bmatrix}$$

or

$$Aq = B$$

Solving for q:

$$q = A^{-1}B$$

References

1. S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, Jr. and P. Weiner, *J. Am. Chem. Soc.*, **106**, 765 (1984). S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case, *J. Compt. Chem.*, **27**, 230 (1986).

2. B. B. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. J. Karplus, *J. Compt. Chem.*, **4**, 187 (1983).

3. J. Hermans, H. J. C. Berendsen, W. F. van Gunsteren, and J. P. M. Postma, *Biopolymers*, **23**, 1513 (1984).

4. P. Dauber and A. Hagler, *Acc. Chem. Res.*, **13**, 105 (1980).

5. F. Momany, R. McGuire, A. Burgess, and H. Scheraga, *J. Phys. Chem.*, **79**, 2361 (1975).

6. D. P. Pearlman and S. H. Kim, *Biopolymers*, **24**, 327 (1985).

7. A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, Macmillan, New York, 1989.

8. D. E. Williams and J. M. Yan, *Adv. Atomic Mol. Phys.*, **23**, 87 (1988).

9. S. R. Cox and D. E. Williams, *J. Compt. Chem.*, **2**, 304 (1981).

10. K. Watanabe and M. L. Klein, *Chem. Phys.*, **131**, 157 (1989).

11. P. C. Hariharan and J. A. Pople, *Chem. Phys. Lett.*, **66**, 217 (1972).

12. J. A. Pople, D. P. Santry, and G. A. Segal, *J. Chem. Phys.*, **43**, S129 (1965). J. A. Pople and G. A. Segal, *J. Chem. Phys.*, **43**, S136 (1965). J. A. Pople and G. A. Segal, *J. Chem. Phys.*, **44**, 3289 (1966).

13. J. A. Pople, D. L. Beveridge and P. A. Dobosh, *J. Chem. Phys.*, **47**, 2026 (1967).

14. M. J. S. Dewar and W. Thiel, *J. Am. Chem. Soc.*, **99**, 4899 (1977). M. J. S. Dewar and W. Thiel, *J. Am. Chem. Soc.*, **99**, 4907 (1977).

15. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. P. P. Stewart, *J. Am. Chem. Soc.*, **107**, 3902 (1985).

16. C. Giessner-Prettre and A. Pullman, *Theoret. Chim. Acta. (Berl)*, **25**, 83 (1972). C. Giessner-Prettre and A. Pullman, *Theoret. Chim. Acta. (Berl)*, **23**, 91 (1974)

17. J. C. Culberson and M. C. Zerner, *Chem. Phys. Lett.*, **122**, 436 (1985).

18. E. Scrocco, *Adv. Quantum Chem.*, **11**, 115 (1978). see reference 8.

19. S. Srebrenik, H. Weinstein, and R. Pauncz, *Chem. Phys. Lett.*, **20**, 419 (1972).

20. D. Klöpper and P. Volkmer, *Chem. Phys. Lett.*, **32**, 186 (1975).

21. S. Obara and A. Saika, *J. Chem. Phys.*, **84**, 3963 (1986).

22. H. F. King and M. Dupuis, *J. Comput. Phys.*, **21**, 144 (1976). M. Dupuis, J. Rys and H. F. King, *J. Chem. Phys.*, **65**, 111 (1976). J. Rys, M. Dupuis and H. F. King, *J. Comp. Chem.*, **4**, 154 (1983).

23. L. E. McMurchie and E. R. Davidson, *J. Comput. Phys.*, **26**, 218 (1978).

24. D. D. Schillady, F. P. Billingssley and J. E. Bloor, *Theoret. Chim. Acta. (Berl)*, **21**, 1 (1961).

25. M. L. Connolly, *J. Appl. Cryst.*, **16**, 548 (1983).

26. U. C. Singh and P. A. Kollman, *J. Comp. Chem.*, **5**, 129 (1984).

27. L. E. Chirlian and M. M. Francl, *J. Comp. Chem.*, **8**, 894 (1987).

28. W. J. Hehre, L. Radom, P. v. R. Schleyer, and J. A. Pople, *Ab Initio Molecular Orbital Theory*, John Wiley & Sons, New York, 1986 p. 68.

29. U. C. Singh and P. A. Kollman, *QCPE Bulletin*, **2**, 17 (1982).

30. J. A. Pople and D. L. Beveridge, *Approximate Molecular Orbital Theory*, McGraw-Hill, New York, 1970 p. 67.

31. For example, for the amino acid arginine terminated with methyl groups, the 6-31G* ESP calculation takes approximately 17 h, while the MNDO calculation takes 40 minutes. These timing are for a Multiflow Trace 14/300. For the 6-31G* calculation the direct SCF code of Gaussian 88 was used, since over 8 Gigabytes of scratch disk space would have been required to perform the calculation in the standard fashion.

32. J. J. P. Stewart, MOPAC 5.0, QCPE, Indiana University, Bloomington, Indiana 47405.