Atomic Charges Derived from Electrostatic Potentials: A Detailed Study

Lisa Emily Chirlian* and Michelle Miller Francl†

Department of Chemistry, Princeton University, Princeton, NJ 08544

A new algorithm for fitting atomic charges to molecular electrostatic potentials is presented. This method is non-iterative and rapid compared to previous work. Results from a variety of gaussian basis sets, including STO-3G, 3-21G and 6-31G*, are presented. Charges for a representative collection of molecules, comprising both first and second row atoms and anions are tabulated. The effects of using experimental and optimized geometries are explored. Charges derived from these fits are found to adequately reproduce SCF dipole moments. A small split valence representation, 3-21G, appears to yield consistently good results in a reasonable amount of time.

INTRODUCTION

As an aid in understanding the properties of a molecule, the concept of atomic charge is a valuable tool. Unfortunately, atomic charge is not a property which can be directly determined from the Hartree-Fock wavefunction. Some scheme must be adopted to divide the total electronic charge among the atoms in a molecule. The computationally simple technique of Mulliken population analysis[1] is, and has long been, most commonly used. However, in recent years a number of other methods have been developed which allocate charge to a molecule's atoms based on physical criteria rather than simply by the equal partitioning method of Mulliken population analysis.[2-4]

The electrostatic potential is a molecular property directly accessible from an SCF calculation:

$$V(r) = \sum_A \frac{Z_A}{|r - R_A|} - \sum_{\mu,\nu} P\mu\nu \int \frac{\varphi_\mu \varphi_\nu}{|r - r'|} dr'$$

where Z_A is the nuclear charge on atom A centered at R_A, $P\mu\nu$ is the density matrix element determined from the Hartree–Fock procedure, and φ_μ, φ_ν are the basis functions used. Many of the problems of older charge determination methods are avoided by a promising new approach that fits the molecular electrostatic potential to a series of point charges placed at the atomic centers.[5-7] We offer here a new procedure for fitting charges to such potentials, along with a tabulation of data obtained in this manner for a wide assortment of small molecules over a range of basis sets.

A minimum prerequisite for a physically meaningful charge assignment scheme is that it should reproduce other, more precisely defined, molecular properties such as dipole or quadrupole moment. Cox and Williams (CW)[6] have shown that charges fit to the electrostatic potential reproduce these requirements quite successfully, while charges obtained using the Mulliken scheme generally do not reproduce the dipole moment calculated from the Hartree Fock wavefunction. Also, the charges derived from the electrostatic potential compare favorably to charges obtained by the Hirschfeld method[4], and to those obtained by a similar fitting procedure as CW from the charge distributions measured crystallographically.[6] While other charge determination methods have been developed (Hirschfeld,[4] Bader[3]) that also yield charges which reproduce molecular properties, these require a more lengthy computation than the present method.

For any computational method to be useful, it must be both efficient and yield high

---

*Author to whom all correspondence should be addressed.
†Present address: Bryn Mawr College, Dept. of Chemistry, Bryn Mawr, PA 19010.

quality results. If a high quality result may only be obtained with a large basis set, the length of time required for the calculation may be so long as to render the calculation effectively useless. On the other hand, a shorter calculation with a smaller basis set may yield unreliable results. A major aim of this work is to determine the basis set which gives the most reliable results in the shortest possible time. To accomplish this, a larger selection of basis sets has been used for the calculations than previously studied. The selection of molecules includes both first and second row atoms, along with a series of anions. The consequences of adding diffuse functions to basis sets for anions are also discussed.

This work follows that of CW but incorporates some changes in the methodology used to fit the charges to the electrostatic potential. The procedures followed to date have utilized a least-squares iterative method to obtain the best fit, necessitating an initial guess at the charge. Here we apply the method of Lagrange multipliers,[8] allowing the charges to be determined *a priori*. The selection of points for the potential determination have also been modified to yield a set which closely mimics the molecular surface without adding to the computing time required. In addition, a modification in the calculation of the necessary integrals has shortened the computing time significantly without sacrificing the quality of the resulting charges.

Previous studies have focused on atomic charges obtained using experimentally determined geometries. Since the electrostatic potential is dependent upon geometry, changes in structure may make significant differences in the charges obtained. While a molecule's experimentally determined geometry would seem to be the best choice for calculation of atomic charges, it may be unavailable. Therefore, a comparison of the charges determined using experimental and optimized geometries in charge calculations has been performed.

Basis Set Considerations

CW showed that for computation of atomic charges in molecules, a minimal STO-3G basis set gave results, for potential derived charges, comparable to 6-31G**, in about one-seventh the time. A small split valence basis set (3-21G) has been shown to give superior results for molecular geometries and vibrational frequencies compared to STO-3G, without a significant increase in calculation time.[9] A reasonable expectation is that 3-21G charges will be even more closely correlated to 6-31G** charges than are those from the STO-3G basis without sacrificing computation time.

Anions are especially difficult to calculate, due to the diffuseness of the charge distribution. 'Plus' basis sets (diffuse functions included on all heavy atoms) have been shown to give better results for geometries and energies of anions than other gaussian basis sets.[10] As the use of diffuse functions allows charge to be spread out further from the nuclear centers than with other basis sets, it necessitates the calculation of the electrostatic potential over longer distances from the nuclei.

Geometry Considerations

The electrostatic potential is a property dependent on the total molecular charge distribution, and thus on the nuclear positions. Virtually all work to date concerning atomic charges fit to the electrostatic potential has employed the experimentally determined geometry of the molecule. Using this method good agreement is obtained among the results for the various basis sets studied (STO-3G, 6-31G, 6-31G**). However, experimentally determined geometries are not always available. Optimized geometries may be obtained, but the agreement of the optimized and experimental geometries will vary with the basis set used. Especially poor results may be expected with minimal basis sets. The effects of these different geometries on the calculated atomic charges are discussed below.

METHOD

A program, CHELP[11] (CHarges from ELectrostatic Potentials) has been written to both calculate the electrostatic potential and fit it to point charges. The geometry optimizations and SCF wavefunctions were obtained from the GAUSSIAN 82[12] program using the basis sets STO-3G[13], 3-21G[14], 6-31G[15], 6-31G*[15], 6-31G**[15]. In addition, 6-31+G[16] was used for calculations on anions.

The electrostatic potential is determined at a selected number of points around the molecule, chosen in spherical shells, 1Å apart, of fourteen nearly symmetrically placed points around each atom. Points which fall within the van der Waals radius of any of the atoms are discarded due to the large distortions caused by close proximity to the nuclei. The total number of points is selected so that the region extending to 3Å from the van der Waals surface of the molecule is considered, typically 100–300 points for the molecules considered here.[17]

The best least square fit of the charges to the electrostatic potential is obtained by finding the minimum of y

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

$$m = \text{total number of points to fit}$$

where V_i is given for LCAO wavefunctions as:

$$V_i = \sum_A \frac{Z_A}{|r_1 - R_A|} - \sum_{\mu,\nu} P\mu\nu \int \frac{\varphi_\mu \varphi_\nu}{|r_1 - r'|} dr'$$

and E_i, the electrostatic potential in the monopole approximation, is given by:

$$E_i = \sum_{j=1}^{n} (q_j/r_{ij})$$

$$n = \text{number of atoms in molecule}$$

The minimum of y can be obtained by finding the stationary points of the Lagrangian function z:

$$z(q_1, q_2, \ldots q_n) = y(q_1, q_2, \ldots q_n)$$
$$+ \lambda g(q_1, \ldots q_n)$$

where g is the constraint imposed on the fit and λ is the Lagrangian multiplier associated with it. We have required the fit to reproduce the total molecular charge. The corresponding function is:

$$g(q_1, q_2, \ldots q_n) = \left(\sum q_i\right) - q_{tot} = 0$$

Extremas of z are then found by solving for $\partial z/\partial \lambda = 0$ and $\partial z/\partial q_k = 0$. This yields a set of $n + 1$ equations in $n + 1$ unknowns, where n is the number of atoms. The solution of the corresponding matrix equation yields the charges.

This method may be modified to impose additional constraints on the fit. For example, the dipole (or higher) moment may be used. However, we find in general that the charges calculated without additional constraints re-

produce the quantum mechanically computed dipole very well (see below).[18]

RESULTS AND DISCUSSION

Table I shows charges for a series of molecules calculated at their experimental geometries. The rms deviations in (kcal/mol) of the electrostatic potentials are also included. This number indicates how well the calculated charges reproduce the electrostatic potential determined from the wavefunction. A selection of CW's results is also presented. In each case, the CHELP charges show good agreement with those of CW. Since a larger basis set will increase the computation time, the ability of the smaller basis sets to reproduce the results of the largest is of interest. The correlation coefficients (r) for a comparison of each basis set relative to 6-31G** are .985 for STO-3G, 0.997 for 3-21G, 0.995 for 6-31G and 0.997 for 6-31G*. A value of one indicates a perfect correlation.

Reasonable correlation is obtained for all basis sets. However, STO-3G shows a much lower correlation coefficient than any of the other basis sets, as expected since minimal basis sets generally yield poorer values for molecular properties than split valence basis sets. The difference among the split valence basis sets is slight, except when calculation time is considered (Table II). The three basis sets considered have almost the same correlation coefficient, yet a 6-31G* calculation takes about two and a half times as long as 3-21G. The 3-21G calculations take 25% longer than those with STO-3G but the resulting improvement in the quality of the results certainly justifies the slightly larger investment in time.

While considering the differences in the results calculated with each basis set, it is interesting to note that the 6-31G results show a slightly poorer correlation with the 6-31G** values than the smaller 3-21G basis set, while taking almost twice as long to calculate. Certainly nothing is gained by using a larger split-valence basis set without the addition of polarization functions.

To illustrate the effect of an increase in the number of points used for fitting, the atomic charges for SH_2, PH_3 and formamide computed using several different numbers of points are presented in Table III. The value of the charges do not change once the rms

Table I. Atomic charges: wavefunction and electrostatic potential calculated at each basis set using experimental geometry.[a] RMS deviation — kcal/mol. Data in parenthesis is from Cox and Williams.

Atom[b]	STO-3G	3-21G	6-31G	6-31G*	6-31G**
HF					
F	−0.298 (−0.294)	−0.495	−0.527 (−0.524)	−0.464	−0.462 (−0.442)
H	0.298 (0.294)	0.495	0.527 (0.524)	0.464	0.462 (0.442)
rms	0.009	0.012	0.011	0.027	0.027
H_2O					
O	−0.615 (−0.614)	−0.875	−0.948 (−0.940)	−0.812	−0.796 (−0.786)
H	0.308 (0.307)	0.438	0.474 (0.470)	0.406	0.398 (0.343)
rms	0.009	0.017	0.021	0.042	0.041
NH_3					
N	−0.953 (−0.972)	−1.165	−1.230 (−1.278)	−1.043	−1.014 (−1.047)
H	0.318 (0.121)	0.388	0.410 (0.426)	0.348	0.338 (0.349)
rms	0.027	0.040	0.056	0.056	0.056
CH_4					
C	−0.463 (−0.484)	−0.564	−0.473 (−0.528)	−0.490	−0.496 (−0.556)
H	0.116 (0.121)	0.141	−0.118 (0.132)	0.123	0.124 (0.139)
rms	0.012	0.015	0.020	0.020	0.021
C_2H_2					
C	−0.180	−0.289	−0.278	−0.295	−0.292
H	0.180	0.289	0.278	0.295	0.292
rms	0.008	0.010	0.012	0.014	0.013
C_2H_4					
C	−0.159	−0.360	−0.339	−0.343	−0.339
H	0.080	0.180	0.169	0.172	0.169
rms	0.021	0.027	0.030	0.027	0.027
CO_2					
C	0.808	1.116	1.165	0.908	0.908
O	−0.404	−0.558	−0.583	−0.454	−0.454
rms	0.003	0.008	0.015	0.003	0.003
H_2CO					
C	0.549	0.671	0.780	0.578	0.578
O	−0.344	−0.516	−0.593	−0.503	−0.502
H	−0.102	−0.078	−0.094	−0.037	−0.037
rms	0.059	0.056	0.064	0.040	0.042
CH_3OH					
C	0.141	0.278	0.420	0.118	0.212
O	−0.470	−0.705	−0.787	−0.638	−0.627
H_a	0.022	0.032	0.003	0.052	0.035
H_b	0.002	−0.018	−0.047	−0.002	−0.005
H_c	0.304	0.431	0.458	0.401	0.390
rms	0.054	0.058	0.071	0.053	0.049
H_2NCHO					
O	−0.404	−0.610	−0.674	−0.580	−0.580
N	−0.697	−0.992	−1.028	−0.931	−0.921
C	0.523	0.789	0.863	0.664	0.660
H_a	−0.026	−0.032	−0.032	0.016	0.017
H_b	0.295	0.433	0.423	0.405	0.401
H_c	0.307	0.412	0.448	0.426	0.423
rms	0.048	0.067	0.081	0.056	0.066
HCOOH[c]					
O_1	−0.531	−0.744	−0.804	−0.646	−0.671
O_2	−0.403	−0.587	−0.643	−0.554	−0.598
C	0.597	0.774	0.851	0.630	0.781
H_a	−0.021	0.041	0.047	0.092	0.027
H_b	0.358	0.515	0.550	0.478	0.461
rms	0.056	0.052	0.054	0.041	0.068
CH_3CN					
N	−0.448	−0.504	−0.542	−0.503	−0.430
C_1	−0.497	−0.515	−0.422	−0.390	−0.390

Table I. *(Continued)*

Atom[b]	STO-3G	-21G	6-31G	6-31G*	6-31G*
C_2	0.475	0.477	0.495	0.433	0.430
H	0.156	0.181	0.186	0.153	0.130
rms	0.034	0.033	0.027	0.022	0.023
$H_2C=C=O$					
C_1	-0.768	-1.193	-1.143	-1.079	-1.076
C_2	0.652	0.883	0.909	0.775	0.774
O	-0.315	-0.457	-0.507	-0.412	-0.411
H	0.216	0.383	0.371	0.358	0.357
rms	0.027	0.062	0.063	0.062	0.061
PH_3^d					
P	-0.149	-0.303	-0.316	-0.110	-0.193
H	0.050	0.101	0.105	0.037	0.064
rms	0.077	0.146	0.144	0.133	0.130
SH_2^d					
S	-0.226	-0.395	-0.404	-0.296	-0.291
H	0.113	0.197	0.202	0.148	0.145
rms	0.040	0.104	0.106	0.134	0.132
HCl^d					
Cl	-0.288	-0.307	-0.309	-0.258	-0.246
H	0.288	0.307	0.309	0.258	0.246
rms	0.039	0.065	0.066	0.092	0.091

[a]molecular geometries from *J. Am. Chem. Soc.*, **92**, 4796 (1970), except as noted.
[b]Atomic labels refer to the following diagrams:
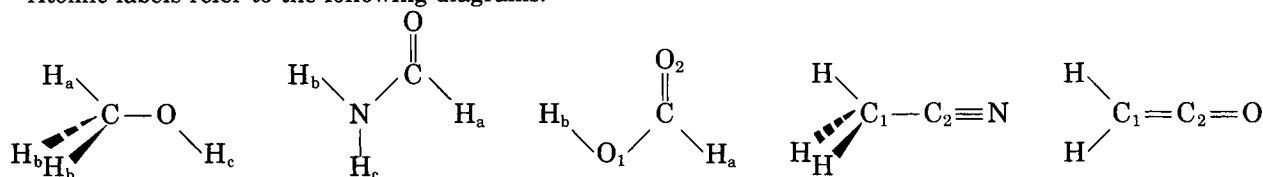
[c]*Acta. Chem. Scand.*, **23**, 2848, (1969).
[d]*J. Am. Chem. Soc.*, **104**, 5039 (1982).

Table II. Time (sec) for each calculation[a]

Molecule	STO-3G	3-21G	6-31G	6-31G*	6-31G**
HF	52	63	118	155	178
H_2O	88	106	186	224	287
NH_3	143	177	299	352	480
CH_4	197	243	480	484	670
C_2H_2	194	244	464	613	720
C_2H_4	384	495	879	1107	1443
CO_2	225	284	586	849	846
H_2CO	201	255	480	614	761
CH_3OH	397	509	902	1114	1446
H_2NCHO	518	681	1230	1629	1958
HCOOH	329	450	835	1126	1340
CH_3CN	494	643	1191	1643	1859
average ratio	1.0	1.27	2.35	2.98	3.36

[a]Calculations run on a VAX 11/780. Calculations not included here were run on a VAX 11/750, which runs at approximately 75% the speed of the 11/780, times for those calculations are not comparable with the data presented here.

Table III. Atomic charges: wave function and electrostatic potential calculated with 3-21G basis and experimental geometry. RMS deviation — kcal/mol, time — seconds.

| | Number of Points Used for Fitting | | | | |
	50	150	250	350	450
H_2NCHO					
C	0.847	0.789	0.789	0.790	0.790
O	−0.623	−0.609	−0.610	−0.611	−0.611
N	−1.007	−0.991	−0.992	−0.993	−0.993
H_a	−0.060	−0.032	−0.032	−0.032	−0.032
H_b	0.430	0.433	0.433	0.434	0.434
H_c	0.414	0.411	0.412	0.412	0.412
rms	0.261	0.091	0.067	0.053	0.044
time	76	450	681	763	920
PH_3					
P	−0.301	−0.303	−0.304	−0.305	−0.305
H	0.100	0.101	0.101	0.102	0.101
rms	0.370	0.241	0.141	0.099	0.766
time	86	130	224	350	414
SH_2					
S	−0.389	−0.394	−0.395	−0.395	−0.395
H	0.195	0.197	0.197	0.197	0.197
rms	0.375	0.189	0.104	0.080	0.065
time	49	96	174	221	274

deviation falls below 0.15. Usually, fitting at enough points to cover 3Å beyond the van der Waals radii of the atoms is adequate to insure that the values calculated for the charges do not change significantly, regardless of how many points are chosen.

The potential derived charges on second row atoms have not previously been investigated. The most significant differences between the results for first and second row atoms is in the rms deviation. Fitting at the same number of points results in a much higher rms deviation for the molecules containing second row atoms when compared to the analogous first row molecules. Presumably, this is due to the larger size and diffuseness of the atoms, which makes a fit at a greater number of points necessary. However, as noted above, increasing the number of points would probably have little effect on anything but the rms deviation, the values of the charges would not be expected to differ significantly.

Table IV shows the same information as above, except that geometries optimized at each basis set were used. When compared to the charges obtained with the geometry optimized at the 6-31G** basis set, similar results to the experimental geometry cases are obtained. The correlations coefficients obtained here are: STO-3G — 0.986, 3-21G — 0.996, 6-31G — 0.994, 6-31G* — 0.999.

A comparison may also be made between the charges calculated using the experimental geometry and those calculated with the geometry optimized at a particular basis set level. In each case, the charges were calculated with the same basis set used for the geometry optimization. Again, the split valence basis sets show superior correlation (3-12G: $r = 0.997$) when compared to the minimal basis set (STO-3G: $r = 0.980$), with the basis sets with polarization (6-31G* and 6-31G**) yielding even better agreement, probably due to the higher quality of the wavefunction. The charges for molecules optimized at either the 6-31G* ($r = 0.998$) or the 6-31G** ($r = 0.998$) level are almost identical to those calculated using the experimental geometry and the 6-31G** basis set, indicating that optimized geometries may be used to yield charges equivalent to those calculated with experimental geometries. For those cases where atomic charges are needed by the experimentally determined geometry is not available, the optimized geometry may be used with a high degree of confidence.

In some cases, the use of a polarized basis set may require too much time for the geometry optimization to be practical. In these cases, it would be desirable to be able to use the 3-21G basis set. Figure 1 shows the 3-21G charges calculated with the 3-21G optimized geometry falling on a line that, while

Table IV. Atomic charges: wave function and electrostatic potential calculated with each basis set using the geometry optimized at that basis set level.

Atom	STO-3G// STO-3G	3-21G// 3-21G	6-31G// 6-31G	6-31G*// 6-31G*	6-31G**// 6-31G**
HF					
F	−0.278	−0.489	−0.526	−0.464	−0.463
H	0.278	0.489	0.526	0.464	0.463
H_2O					
O	−0.567	−0.887	−0.996	−0.824	−0.811
H	0.283	0.443	0.498	0.412	0.406
NH_3					
N	−0.916	−1.272	−1.431	−1.060	−1.037
H	0.305	0.424	0.477	0.353	0.346
CH_4					
C	−0.464	−0.563	−0.471	−0.490	−0.496
H	0.116	0.141	0.118	0.123	0.124
C_2H_2					
C	−0.180	−0.286	−0.275	−0.291	−0.288
H	0.180	0.286	0.275	0.291	0.288
C_2H_4					
C	−0.157	−0.355	−0.336	−0.338	−0.334
H	0.078	0.177	0.168	0.169	0.167
CO_2					
C	0.772	1.122	1.164	0.924	0.924
O	−0.386	−0.561	−0.582	−0.462	−0.462
H_2CO					
C	0.539	0.679	0.779	0.606	0.606
O	−0.343	−0.521	−0.595	−0.508	−0.507
H	−0.098	−0.079	−0.092	−0.049	−0.050
CH_3OH					
C	0.178	0.305	0.471	0.255	0.242
O	−0.442	−0.708	−0.812	−0.648	−0.639
H_a	0.001	0.016	0.022	0.026	0.027
H_b	−0.006	−0.018	−0.051	−0.018	−0.016
H_c	0.276	0.425	0.464	0.402	0.398
H_2NCHO					
O	−0.428	−0.632	−0.713	−0.590	−0.615
N	−0.788	−1.062	−1.120	−0.996	−1.020
C	0.658	0.881	0.985	0.752	0.780
H_a	−0.064	−0.054	−0.058	−0.014	−0.010
H_b	0.301	0.420	0.433	0.412	0.414
H_c	0.322	0.447	0.473	0.436	0.453
HCOOH					
O_1	−0.522	−0.784	−0.863	−0.666	−0.661
O_2	−0.395	−0.604	−0.662	−0.575	−0.575
C	0.586	0.843	0.929	0.694	0.688
H_a	0.018	0.016	0.020	0.060	0.062
H_b	0.039	0.530	0.576	0.487	0.487
CH_3CN					
N	−0.445	−0.512	−0.543	−0.509	−0.522
C_1	−0.478	−0.507	−0.399	−0.364	−0.276
C_2	0.462	0.485	0.490	0.433	0.432
H	−0.316	−0.470	−0.518	−0.414	−0.413
$H_2C=C=O$					
C_1	−0.763	−1.235	−1.172	−1.120	−1.114
C_2	0.645	0.918	0.932	0.804	0.800
O	−0.316	−0.470	−0.518	−0.414	−0.413
H	0.217	0.393	0.379	0.365	0.393

Table IV. *(Continued)*

Atom	STO-3G// STO-3G	3-21G// 3-21G	6-31G// 6-31G	6-31G*// 6-31G*	6-31G**// 6-31G**
PH₃					
P	−0.172	−0.344	−0.360	−0.293	−0.235
H	0.057	0.115	0.120	0.098	0.079
H₂S					
S	−0.232	−0.417	−0.427	−0.316	−0.310
H	0.116	0.209	0.213	0.158	0.155
HCl					
Cl	−0.284	−0.303	−0.305	−0.253	−0.247
H	0.284	0.303	0.305	0.253	0.247
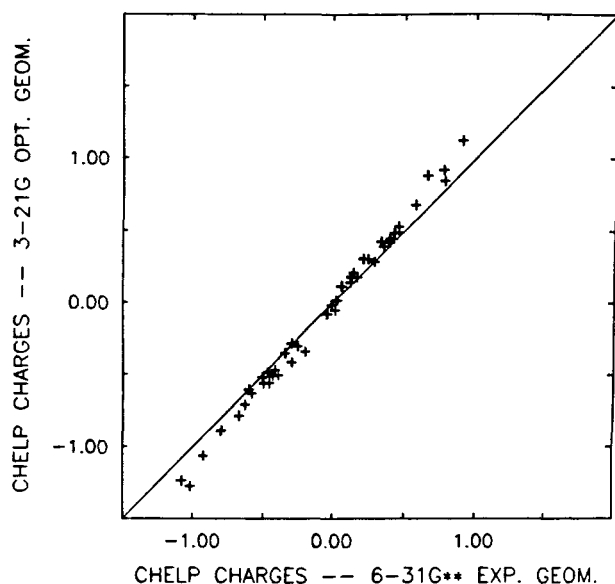
Figure 1. Plot of CHELP charges calculated from 3-21G *wavefunctions using the 3-21G optimized geometry* and those obtained with the 6-31G** wavefunction and experimental geometries. Solid line corresponds to $q_{3\text{-}21G} = q_{6\text{-}31G**}$.

Figure 2. Plot of 3-21G corrected CHELP charges (3-21G optimized geometry) with 6-31G** CHELP charges (experimental geometry). Actual 3-21G CHELP charges have been corrected to fit the 6-31G** charges (see text). Solid line corresponds to $q_{3\text{-}21G} = q_{6\text{-}31G**}$.

straight, is skewed from the optimum fit with the 6-31G** values for the experimental geometry. This suggests that some correction factor may be applied to bring the 3-21G values in line with the 6-31G** values. Plotted against the 6-31G** charges obtained with experimental geometries, the 3-21G values fall on a line with slope 1.161 and intercept −0.00195. Using these values a correction factor can be obtained.[19] Using the corrected charges a new plot (Figure 2) shows that the 3-21G values do fall on the line representing perfect correlation to the 6-31G** values. Using 3-21G optimized geometry and these corrected charges should yield results consistent with those obtained with the larger basis set and optimized geometry.

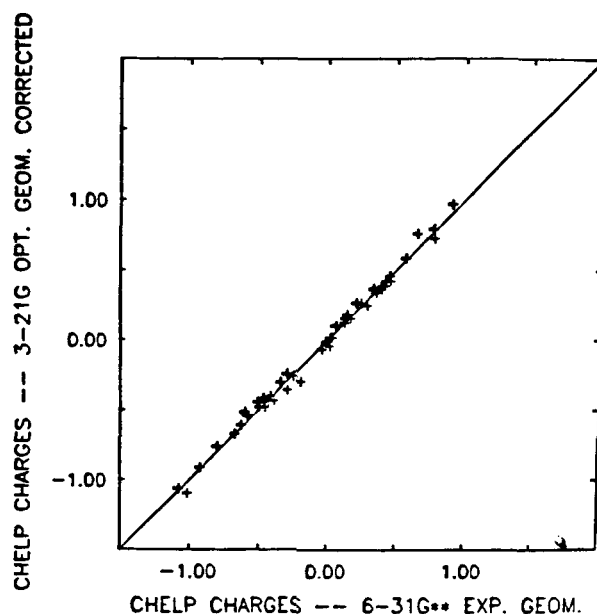One measure of the utility of an atomic charge definition is how well it reproduces

molecular properties. The most common quantity analyzed is the dipole moment. Table V shows the dipole moment calculated from the wavefunction (SCF-dipole), the CHELP charges and the Mulliken charges for several basis sets. For comparison, the experimentally determined dipole moment is also listed. Figure 3 shows plots of CHELP dipoles, Mulliken dipoles and SCF dipoles computed with the experimental geometry plotted versus experimentally determined dipole moments for the STO-3G, 3-21G and 6-31G** basis sets. In each case, the CHELP dipole correlates extremely well with the SCF dipole while the Mulliken dipoles are notably erratic. The 6-31G** values show the most consistent relationship with the experimen-

Table Va. Dipole moments calculated from a) SCF wave function, b) CHELP charges, c) Mulliken charges. Experimental geometries used for wave function, Mulliken charges and CHELP charges.

STO-3G	Experiment[a]	SCF	CHELP	Mulliken
HF	1.820	1.289	1.311	0.492
H_2O	1.850	1.728	1.746	0.547
NH_3	1.470	1.783	1.739	0.854
H_2CO	2.330	1.513	1.421	0.753
CH_3OH	1.700	1.492	1.440	0.583
H_2NCHO	3.730[b]	2.937	2.885	1.232
HCOOH	1.410	0.745	0.745	1.127
CH_3CN	3.920	3.066	3.137	1.306
H_2CCO	1.420[b]	0.710	0.684	0.515
PH_3	0.580[c]	0.537	0.555	0.687
SH_2	0.970[c]	1.015	1.008	0.172
HCl	1.110[b]	1.734	1.766	0.557
3-21G	Experiment	SCF	CHELP	Mulliken
HF	1.820	2.155	2.182	1.057
H_2O	1.850	2.445	2.485	1.090
NH_3	1.470	2.170	2.127	0.809
H_2CO	2.330	2.629	2.545	1.981
CH_3OH	1.700	2.127	2.064	1.533
H_2NCHO	3.730	4.331	4.271	2.464
HCOOH	1.410	1.290	1.287	1.162
CH_3CN	3.920	3.897	3.963	3.385
H_2CCO	1.420	1.736	1.777	2.336
PH_3	0.580	1.176	1.130	0.169
SH_2	0.970	1.810	1.758	0.373
HCl	1.110	1.849	1.882	0.669
6-31G**	Experiment	SCF	CHELP	Mulliken
HF	1.820	1.972	2.033	0.921
H_2O	1.850	2.196	2.258	1.011
NH_3	1.470	1.886	1.850	0.762
H_2CO	2.330	2.749	2.695	1.644
CH_3OH	1.700	1.914	1.864	1.336
H_2NCHO	3.730	4.486	4.458	2.412
HCOOH	1.410	1.592	1.505	0.940
CH_3CN	3.920	4.088	4.137	2.472
H_2CCO	1.420	1.753	1.790	1.493
PH_3	0.580	0.800	0.718	0.318
SH_2	0.970	1.373	1.294	0.310
HCl	1.110	1.480	1.509	0.626

[a]Experimental values from: R. D. Nelson, Jr., et al., *Selected Values of Electric Dipole Moments for Molecules in the Gas Phase*, 1967, NBRDS-NBS 10. Values accurate to at least ±1%, except as noted.

[b]values accurate to ±2%.

[c]values accurate to ±.01D.

tally determined values. This shows the value of obtaining charges that correlate well with 6-31G** values. A similar comparison for dipole moments from charges and wavefunctions obtained with optimized geometries yields qualitatively similar results.

Wavefunctions for anions, as a class of molecules, have proven to be more difficult to calculate. Basis sets with diffuse functions have been developed and have been shown to treat anions better than split valence basis sets. The charges calculated on anions with different types of basis sets are shown in

Table VI. It is more difficult to comment on the accuracy of these charges since experimental dipole moments do not exist for comparison. The validity of the computed dipole moments is also questionable. In general, the most significant difference between the three basis sets studied is in the rms deviation. In each case studied, the 6-31+G rms deviation is higher than that for the other basis sets. The diffuse functions used in this basis can place considerable electron density farther from the nuclear centers than the functions used in the other basis sets, necessitating the

Table Vb. Same as Table Va except geometries optimized at each basis set were used.

STO-3G	Experiment[a]	SCF	CHELP	Mulliken
HF	1.820	1.252	1.279	0.467
H_2O	1.850	1.709	1.732	0.534
NH_3	1.470	1.876	1.873	0.477
H_2CO	2.330	1.537	1.449	0.752
CH_3OH	1.700	1.509	1.484	0.590
H_2NCHO	3.730	2.642	2.574	1.127
HCOOH	1.410	0.627	0.622	0.398
CH_3CN	3.920	3.097	3.171	1.333
H_2CCO	1.420	0.684	0.863	0.557
PH_3	0.580	0.600	0.617	0.659
SH_2	0.970	1.023	1.022	2.167
HCl	1.110	1.765	1.791	0.580

3-21G	Experiment[a]	SCF	CHELP	Mulliken
HF	1.820	2.174	2.202	1.073
H_2O	1.850	2.387	2.431	1.062
NH_3	1.470	1.751	1.725	0.629
H_2CO	2.330	2.658	2.578	1.997
CH_3OH	1.700	2.121	2.090	1.509
H_2NCHO	3.730	4.113	4.017	2.256
HCOOH	1.410	1.397	1.408	1.125
CH_3CN	3.920	3.884	3.950	2.564
H_2CCO	1.420	1.817	1.861	2.396
PH_3	0.580	1.242	1.207	0.165
SH_2	0.970	1.830	1.814	0.391
HCl	1.110	1.856	1.885	0.689

6-31G**	Experiment[a]	SCF	CHELP	Mulliken
HF	1.820	1.944	2.002	0.885
H_2O	1.850	2.148	2.212	0.968
NH_3	1.470	1.839	1.811	0.732
H_2CO	2.330	2.663	2.608	1.575
CH_3OH	1.700	1.834	1.790	1.269
H_2NCHO	3.730	4.094	4.080	2.188
HCOOH	1.410	1.625	1.623	0.883
CH_3CN	3.920	4.066	4.114	2.443
H_2CCO	1.420	1.631	1.669	1.360
PH_3	0.580	0.891	0.828	0.297
SH_2	0.970	1.392	1.345	0.307
HCl	1.110	1.472	1.503	0.340

Table VI. Atomic charges on anions. Geometries optimized at each basis set. rms deviation — kcal/mol.

	Basis Set — Number of Points				
	3-21G 250	6-31$^+$G 250	6-31$^+$G 500	6-31$^+$G 1000	6-31G** 250
CN^-					
C	−0.537	−0.489	−0.490	−0.490	−0.504
N	−0.463	−0.511	−0.510	−0.510	−0.496
rms	0.1455	0.1769	0.0811	0.0440	0.1079
OH^-					
O	−1.095	−1.305	−1.306	−1.306	−1.171
H	0.096	0.305	0.306	0.306	0.171
rms	0.0238	0.1221	0.0611	0.0305	0.0796
$HCOO^-$					
C	0.929	1.184	1.183	1.183	0.929
H	−0.240	−0.221	−0.220	−0.220	−0.226
O	−0.845	−0.982	−0.982	−0.982	−0.852
rms	0.0221	0.0451	0.0220	0.0115	0.0267
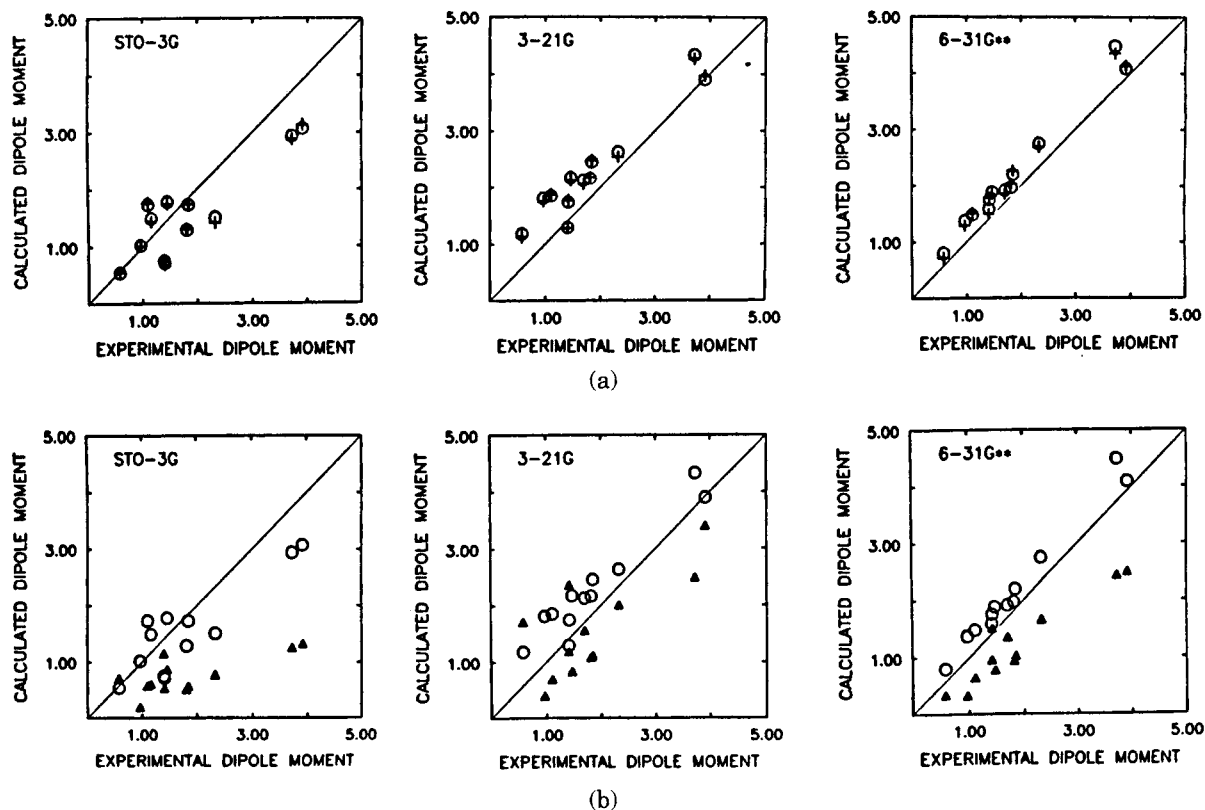
Figure 3. a) Plot of calculated versus experimental dipole moments for CHELP (+) and SCF-wave-function (○), using the STO-3G, 3-21G and 6-31G** basis sets. Solid line corresponds to $\mu_{calc} = \mu_{exp}$. b) Same as a, except dipole moments for Mulliken (▲) and SCF-wavefunction (○) are plotted. The correlation coefficients are:

	r_{SCF}	r_{CHELP}	$r_{Mull.}$
STO-3G	0.872	0.868	0.722
3-21G	0.958	0.961	0.752
6-31G**	0.989	0.992	0.940

use of more points for an equally good fit. However, the charges do not change significantly with the increase in points.

CONCLUSIONS

Useful atomic charges have been shown to result from the fitting of point charges to the ab initio electrostatic potential. A new program, CHELP, calculates these charges in considerably shorter time than older methods, but yields similar results. CHELP calculates high quality atomic charges with a small split-valence (3-21G) basis set, in a reasonably short time. Fitting point charges centered on the nuclei to the electrostatic potential calculated from the wavefunction obtained with the experimental geometry gives atomic charges which reproduce other molecular properties.

If the experimental geometry for a molecule is not available, good charges may still be obtained by using an optimized geometry.

For this case, calculations using the 6-31G* or 6-31G** basis sets and geometries optimized at the same level give results which reproduce the charges obtained with the experimental geometry and the 6-31G** basis set. However, this represents a considerable investment in computation time. The charges obtained using the 3-21G basis set for both charge and geometry calculations may be scaled to reproduce the 6-31G** experimental geometry values quite well.

The rms deviation gives an indication of the quality of the fit of the calculated atomic charges to the molecular electrostatic potential. While it is advisable to do a greater number of points than necessary to insure adequate representation of the electrostatic potential, the charges calculated are relatively insensitive to the change in number of points and actually level off quite quickly. By calculating the electrostatic potential at enough points so that the region extending 3Å beyond the van der Waals radii of the

atoms is considered, adequate coverage is insured. Fitting to the electrostatic potential calculated beyond this point will lower the rms deviation but will only lengthen the calculation with no effect on the charges.

The charges produced by CHELP reproduce the dipole moments of the molecules calculated at the same level of theory. The dipoles calcuated at 6-31G** have the best correlation with the experimentally measured values. While a calculation performed at 6-31G** may take a great deal of time, charges calculated at the 3-21G level correlate well after scaling with those calculated at 6-31G** and take about one-third the time.

The program CHELP efficiently extracts charges from the ab initio-SCF wavefunction which reproduce molecular quantities well in a reasonably short time. More detailed comparison of these charges with other charge methods over a wider range of molecules than presently is available should give an even better description of the utility of these charges. Also, an analysis of the relationship between these charges and other molecular properties such as NMR chemical shift, and reactivity should yield interesting results.

References

1. R. S. Mulliken, *J. Chem. Phys.*, **23**, 1833 (1955).
2. P. Politzer, R. Harris, *J. Am. Chem. Soc.*, **92**, 6451 (1970).
3. R. F. W. Bader, P. M. Beddall, P. E. Cade, *J. Am. Chem. Soc.*, **93**, 3095 (1971).
4. F. L. Hirshfeld, *Theor. Chem. Acta*, **44**, 129 (1977).
5. F. A. Momany, *J. Phys. Chem.*, **82**, 592 (1978).
6. S. R. Cox, D. E. Williams, *J. Comp. Chem.*, **2**, 304 (1981).
7. U. C. Singh, P. A. Kollman, *J. Comp. Chem.*, **5**, 129 (1984).
8. See: B. S. Gottfried and J. Weisman, *Introduction to Optimization Theory*, Prentice Hall, New Jersey, 1973, p. 43 for an example of this method.
9. J. S. Binkley, J. A. Pople, W. J. Hehre, *J. Am. Chem. Soc.*, **102**, 939 (1980).
10. J. Chandrasekhar, J. G. Andrade, P. v.R. Schleyer, *J. Am. Chem. Soc.*, **103**, 5609 (1981).
11. L. E. Chirlian, M. M. Francl, *QCPE*, **7**, 39 (1987).
12. J. S. Binkley, M. Frisch, R. Krishnan, D. J. De Frees, H. B. Schlegel, R. Whiteside, E. Fluder, R. Seeger, J. A. Pople, GAUSSIAN 82, release Carnegie–Mellon University, 1982.
13. W. J. Hehre, R. F. Stewart, J. A. Pople, *J. Chem. Phys.*, **51**, 2657 (1969).
14. J. S. Binkley, J. A. Pople, W. J. Hehre, *J. Am. Chem. Soc.*, **102**, 939 (1980).
15. P. C. Harihan, J. A. Pople, *Theor. Chem. Acta*, **28**, 213 (1973).
16. G. W. Spitznagel, T. Clark, J. Chandrasekhar, P. v.R. Schleyer, *J. Comp. Chem.*, **3**, 363 (1982).
17. Results presented here were obtained by fitting to 250 points, representing enough points to adequately treat the largest molecule studied. The excess of points for the smaller molecules will not alter the charges obtained (see text).
18. If the dipole moment is included as a constraint, it may not be used to assess the quality of the calculated charges.
19. To correct the 3-21G charges to the 6-31G** results, the difference between the equations of the two lines must be determined. Equation of the ideal line: $y = x$. The 3-21G data falls along a line $y = 1.161x - 0.00195$. Putting this equation in the form $y = x$ gives: $(y + 0.00195)/1.161 = x$. Thus, the corrected 3-21G charges may be obtained by adding .00195 to each and then dividing by 1.161.