Class IV charge models:
A new semiempirical approach in quantum chemistry

Joey W. Storer, David J. Giesen, Christopher J. Cramer* and Donald G. Truhlar*

Department of Chemistry and Supercomputer Institute, University of Minnesota, Minneapolis, MN 55455-0431, U.S.A.

Summary

We propose a new criterion for defining partial charges on atoms in molecules, namely that physical observables calculated from those partial charges should be as accurate as possible. We also propose a method to obtain such charges based on a mapping from approximate electronic wave functions. The method is illustrated by parameterizing two new charge models called AM1-CM1A and PM3-CM1P, based on experimental dipole moments and, respectively, on AM1 and PM3 semiempirical electronic wave functions. These charge models yield rms errors of 0.30 and 0.26 D, respectively, in the dipole moments of a set of 195 neutral molecules consisting of 103 molecules containing H, C, N and O, covering variations of multiple common organic functional groups, 68 fluorides, chlorides, bromides and iodides, 15 compounds containing H, C, Si or S, and 9 compounds containing C-S-O or C-N-O linkages. In addition, partial charges computed with this method agree extremely well with high-level ab initio calculations for both neutral compounds and ions. The CM1 charge models provide a more accurate point charge representation of the dipole moment than provided by most previously available partial charges, and they are far less expensive to compute.

Introduction

Partial charges on the atoms in a molecule are one of the most venerable concepts in chemistry, but such charges are intrinsically nonmeasurable. In other words, partial charges are a theoretical construct which cannot be defined unambiguously in terms of experimental observables. Nevertheless, the concept of partial charges is very powerful for a qualitative understanding of structure and reactivity, and partial charges play a critical role in many force fields used for molecular simulations [1]. The concept of the partial atomic charge is increasingly utilized for quantitative modeling of structure and reactivity [2].

Currently employed methods for defining partial charges may be classified into three general categories: (I) methods that extract them directly from experiment [3], e.g., calculating partial charges for a diatomic molecule by dividing the dipole moment by the bond length; (II) methods that extract them from a quantum-mechanical wave function by analyzing the wave function itself, e.g., Mulliken population analysis [4] or the partitioning of 3D space into regions associated with individual atoms [5];

and (III) methods that extract them from a quantum-mechanical wave function by analyzing a physical observable predicted from the wave function, e.g., fitting of predicted interaction energies [6–8], dipole moments [9], or electrostatic potential fitting [10–18]. The present paper will present a method to improve class II or class III charges.

Wave functions used for electronic structure calculations on molecules are generally classified as ab initio or semiempirical. The former have the advantage of being systematically improvable; the latter have the advantage that – especially for large systems and medium levels of accuracy – they often deliver a given level of accuracy at the lowest computational cost. Semiempirical models have reached a high level of sophistication, with two current pinnacles of the art being the popular AM1 [19] and PM3 [20] models. In both of these the parameters were obtained by fitting to, inter alia, predicted energetic quantities (such as heats of formation). Charges are typically extracted from semiempirical wave functions by the same numerical algorithms, e.g., Mulliken population analysis or electrostatic potential fitting, as used to obtain them

---

*To whom correspondence should be addressed.

from ab initio wave functions. The semiempirical parameters are optimized to reproduce heats of formation, and the resulting approximate wave functions are analyzed by class II or class III methods. Here, we propose adding a qualitative, new type of semiempirical tool to the theoretical tool bag, namely an empirical *charge model* (CM) where a set of class II or class III charges is transformed into a new set of charges by a semiempirical mapping. The mapping parameters are optimized so that physical observables calculated from the new charges are more accurate than those calculated from the original charges. We propose that the charges obtained by such a mapping be called class IV charges.

A class IV charge model can be developed for any level of ab initio or semiempirical wave function. For example, one can write an improved partial charge for atom k as

$$q_k^{CM} = f(q_k^{(0)}, A, B, ...) \qquad (1)$$

where f is a semiempirical functional, $q_k^{(0)}$ are the 'zero-order' charges obtained from an ab initio or semiempirical wave function by any well-defined class II or class III method, and A, B, ... are parameters adjusted to achieve improved agreement with a set of physical observables. One possible physical observable to be used is the experimental dipole moment. Alternatively, the electrostatic potential computed from a high-level wave function (e.g., Møller–Plesset second order perturbation theory (MP2) [21] with a polarized basis set, e.g., 6-31G* [22], cc-pVDZ [23] or cc-pVTZ [23]) may serve as the physical observable. A critical element in the definition of both class III and class IV charges is that they are designed to lead to accurate physical observables as calculated directly from the partial charges. This should be distinguished from using a class II charge model; with such models the physical observable calculated as an expectation value using the full wave function may be far more accurate than the physical observable calculated from the partial charges. The distinction between class III and class IV charges, on the other hand, is that obtaining the latter involves a semiempirical mapping such that using the resulting partial charges to predict the physical observables is more accurate than using the underlying wave function from which they were mapped. Thus, the specification of a class IV charge model requires several underlying choices: (i) the level of wave function; (ii) the method used to obtain the set of $q_k^{(0)}$ from the wave function; (iii) the functional form of the mapping; and (iv) the data set used to optimize the parameters in the mapping. Here we will illustrate the new approach with two examples of new charge models.

In Charge Model 1A (CM1A), the $q_k^{(0)}$ are obtained by Mulliken analysis of AM1 wave functions. The mapping is a multilinear form in which each q_k^{CM} depends on all $q_k^{(0)}$ and on some of the bond orders of the molecule under

consideration. The data set used to obtain the parameters consists of 186 experimental dipole moments for diverse molecules containing H, C, N, O, F, Si, S, Cl, Br and I plus nine high-level theoretical dipole moments for compounds made of H, C, N, O and S. In addition, partial charges in 25 ions were compared to high-level ab initio partial charges. Charge Model 1P (CM1P) is identical with respect to points (ii)–(iv), except that it uses a smaller number of bond orders and the parameters for all molecules are determined using only the 186 experimental dipole moments. With regard to choice (i), the CM1P model is based on using PM3 wave functions instead of AM1 wave functions.

We decided to base the class IV charges presented here on the dipole moments of small, primarily monofunctional organic molecules plus NH_3, NH_2OH, H_2O and H_2S. To motivate this, we note first that fitting to dipole moments is the same as fitting to electrostatic potentials on an infinite-radius surface where only the effect of the leading multipole moment survives. Moreover, it has two advantages over fitting to the electrostatic potential on surfaces: (i) no arbitrary choices have to be made as to the location of the surfaces; and (ii) experimental values are available for a wide variety of molecules. We note that, although the reproduction of specific multipole moments can be included as a constraint in class III fitting procedures, the moments themselves are typically derived not from experiment but rather from the wave function (thereby defining the methodology as class III). By using experimental dipole moments, class IV charges are designed to correct for systematic errors that occur even in the continuous density function corresponding to a given level of electronic structure. As such, we want a data set that isolates these systematic errors – monofunctional compounds do this efficiently. By forcing the charge model to predict a large number of such dipole moments accurately, one builds in more accurate local charge balances. To the extent that one can view higher order multipoles as a superposition of local monopoles and dipoles from individual functional groups, higher order terms in the multipole expansion should also be improved.

Although in principle it is straightforward to optimize partial charges to best reproduce dipole moments for neutral molecules, this is not the case for ions. Unlike the neutral case, the dipole moment of a charged molecule depends on the point at which it is evaluated. Common choices for evaluation of the dipole moment are the center of charge or the center of mass, and both of these can shift dramatically with only small changes in the point charges or the molecular geometry – thus, calculation of the ionic dipole moment is particularly sensitive to both of these quantities. Finally, experimental data are rarely available because of obvious technical difficulties associated with measurement.

Nevertheless, prediction of useful (accurate) atomic

partial charges in ions is an important goal, especially since it represents a challenging test for any class IV charge model. In particular, ions often include atoms with total bond orders distinct from those found in neutral molecules (e.g., quaternary ammonium species) and have initial class II or class III charges which are similarly unique by comparison to neutral analogs. Thus, we have considered a number of ions in our parameterization test sets; however, given the ambiguity of comparing calculated dipole moments, we measure the goodness of the class IV charge model by how well it reproduces very high quality class III charges, in this case atomic partial charges calculated from ChelpG [15] fitting to MP2/cc-pVTZ [21–24] densities.

The next section presents the mapping strategy employed in CM1A and CM1P. The parameterization was accomplished in stages, and the Database section presents the databases for the various stages. The Parameterization section presents the new parameters, and the subsequent section compares the new class IV partial charges to those obtained by several class II and class III methods; it also compares root-mean-square (rms) errors in the dipole moments computed from the various class II, class III and class IV partial charges, relative to experiment. The final section contains a brief summary and concluding remarks.

Theory

Both AM1 and PM3 are based on the assumption [25] of neglect of diatomic differential overlap (NDDO). The class II atomic partial charges used in this study are given by the Mulliken definition of net atomic charge, which – when applied to an NDDO wave function – yields:

$$q_k^{(0)} = Z_k - \sum_{\mu \in k} P_{\mu\mu} \qquad (2)$$

where Z_k is the nuclear charge, μ is a basis function index and $P_{\mu\mu}$ is the diagonal element of the density matrix constructed from the orthonormal eigenvectors of the wave function. (In valence-electron-only methods, such as the AM1 or PM3 models, Z_k is the nuclear charge minus the number of core electrons.)

The mapping consists of two steps. In the first step, the initial partial charge is adjusted using the relation

$$q_k^{(1)} = q_k^{(0)} + B_k \Delta q_k \qquad (3)$$

where

$$\Delta q_k = c_k q_k^{(0)} + d_k \qquad (4)$$

and c_k is a scale factor, d_k an offset and B_k the sum of the bond orders of atom k to all other atoms. Because of the bond order factor in Eq. 3, the change in partial charge decreases smoothly to zero along a dissociation coordinate where B_k goes to zero. The parameters c_k and d_k take the general forms

$$c_k = \hat{c}_k + \sum_{k' \neq k} f^{(c)}(B_{kk'}) c_{kk'} \qquad (5)$$

and

$$d_k = \hat{d}_k + \sum_{k' \neq k} f^{(d)}(B_{kk'}) d_{kk'} \qquad (6)$$

In Eqs. 5 and 6, $\hat{c}_k$ and $\hat{d}_k$ are constants that depend on the atomic number of k (e.g., C, N, O, ...), $B_{kk'}$ is the bond order between atoms k and k', $f^{(c)}(B_{kk'})$ and $f^{(d)}(B_{kk'})$ are functions of $B_{kk'}$, and $c_{kk'}$ and $d_{kk'}$ depend on the atomic number of atoms k and k'. In the CM1A and CM1P models, $c_{kk'}$ and $d_{kk'}$ are made non-zero only in specific cases, as explained below.

In the second step of the mapping, the partial charges are readjusted to force the total charge on the molecule or ion to be the proper integral value. This is done by shifting charge locally between each atom whose charge has been adjusted and the atoms to which it has non-zero bond order. The final partial charge is then

$$q_k = q_k^{(0)} + B_k \Delta q_k - \sum_{k' \neq k} B_{kk'} \Delta q_{k'} \qquad (7)$$

where $B_{kk'}$ is the bond order between atoms k and k'. The sum of the bond orders from atom k to all other atoms is

$$B_k = \sum_{k' \neq k} B_{kk'} \qquad (8)$$

Thus, Eq. 7 always ensures that molecular charge is conserved. Focusing for the moment on the primary change at a single atom k, note that the renormalization (the last term in Eq. 7) of the other partial charges, to provide the charge needed for the primary change, affects only those atoms with a non-zero bond order to the atom undergoing the primary change and is further proportional to that bond order. This is a critical element of the method. If the renormalization were spread out globally over the whole molecule, then one would predict a different C-N bond polarity in ethylamine and n-octylamine, for example, just because the renormalizations would be more spread out in the latter case. Such behavior would be unphysical.

Although any definition of bond order could be used in the above equations, we use the covalent bond index defined by Armstrong et al. [26], namely

$$B_{kk'} = \sum_{\mu \in k} \sum_{v \in k'} P_{\mu v}^2 \qquad (9)$$

The scalar dipole moment, μ, is calculated from the charges, q_k, by

$$\mu = \pm \left[\left(\sum_k q_k x_k \right)^2 + \left(\sum_k q_k y_k \right)^2 + \left(\sum_k q_k z_k \right)^2 \right]^{1/2} \qquad (10)$$

where x_k, y_k and z_k are the Cartesian coordinates of atom k. The sign in Eq. 10 is considered positive if the calculated dipole moment vector has a positive scalar product with the correct dipole moment vector and negative if

TABLE 1
DIPOLE MOMENTS (D) OF MOLECULES CONTAINING H, C AND O

Compound	Experiment	AM1		PM3	
		Mulliken	CM1A	Mulliken	CM1P
Alcohols					
Water	1.854	1.09	2.02	0.97	1.92
Methanol	1.70	1.14	1.63	0.93	1.59
Ethanol	1.441	1.12	1.56	0.94	1.56
1,2-Ethanediol	2.28	1.60	2.13	1.20	1.97
Cyclopropanol	1.46	0.84	1.28	0.60	1.22
anti-1-Propanol	1.555	1.12	1.53	0.90	1.52
gauche-1-Propanol	1.58	1.23	1.68	1.03	1.64
2-Propanol	1.224	1.22	1.65	1.03	1.66
1,2-Propanediol (CH$_3$ *anti*)	2.568	1.77	2.27	1.29	2.03
1,2-Propanediol (CH$_3$ *gauche*)	2.32	1.56	1.98	1.19	1.87
Cyclobutanol	1.62	1.04	1.48	0.88	1.52
Phenol	1.45	0.74	1.27	0.53	1.25
Esters, lactones					
Methyl formate	1.77	1.42	1.56	1.49	1.70
3-Oxetanone	0.887	0.44	0.55	0.44	0.60
β-Propiolactone	4.18	3.81	4.32	3.34	4.13
Methyl acetate	1.72	1.84	1.96	1.84	2.05
anti-Ethyl formate	1.98	1.71	1.82	1.70	1.89
gauche-Ethyl formate	1.81	1.47	1.59	1.53	1.74
γ-Butyrolactone	4.27	4.14	4.65	3.71	4.51
2(5*H*)-Furanone	4.905	4.30	4.83	3.88	4.68
Aldehydes, ketones					
Formaldehyde	2.332	2.02	2.42	1.83	2.41
Ketene	1.42	1.37	1.81	1.04	1.70
Acetaldehyde	2.75	2.35	2.74	2.12	2.70
Cyclopropanone	2.67	1.97	2.37	1.67	2.28
Propenal	3.12	2.71	3.10	2.41	2.99
Methylketene	1.79	1.55	1.98	1.17	1.82
Propanal	2.52	2.23	2.62	2.06	2.64
Acetone	2.88	2.55	2.95	2.33	2.92
Cyclobutane-1,2-dione	3.831	3.21	3.77	2.71	3.56
Cyclobutanone	2.89	2.38	2.76	2.14	2.73
E-2-Butenal	3.67	3.17	3.55	2.82	3.40
2-Butanone	2.78	2.42	2.81	2.24	2.83
Cyclopentadienone	3.132	2.59	2.98	2.32	2.93
Cyclopentanone	3.30	2.55	2.92	2.38	2.94
4-Cyclopentene-1,3-dione	1.68	1.61	1.86	1.64	2.04
3-Cyclopentenone	2.79	2.33	2.69	2.23	2.79
Acids					
s-cis-Formic acid	1.41	1.76	1.52	1.74	1.62
s-trans-Formic acid	3.787	3.15	4.23	3.02	4.32
Acetic acid	1.70	2.30	1.98	2.16	1.96
Glyoxylic acid	1.86	1.74	2.37	1.65	2.37
(C-C)*s-cis*-Propenoic acid	1.46	2.36	1.97	2.17	1.92
(C-C)*s-trans*-Propenoic acid	2.02	2.84	2.54	2.62	2.43
Acetoacetic acid	2.30	2.08	2.59	1.78	2.43
Propanoic acid	1.46	2.29	1.94	2.16	1.94
gauche-2-Methoxyacetic acid	4.72	4.07	5.08	3.75	5.03
Ethers					
Dimethyl ether	1.30	1.31	1.53	1.11	1.44
s-cis-Vinyl methyl ether	0.96	0.58	0.77	0.39	0.66
Oxetane	1.94	1.73	2.01	1.50	1.93
Furan	0.661	0.36	0.66	0.07	0.54
3-Methyleneoxetane	1.63	1.65	1.93	1.39	1.83
Tetrahydrofuran	1.63	1.54	1.76	1.32	1.67
1,3-Dioxane	2.06	1.66	2.05	1.46	2.01
3,4-Dihydro-2,4-pyran	1.283	0.86	1.07	0.68	1.01
Tetrahydropyran	1.53	1.31	1.56	1.13	1.51
Anisole	1.38	0.89	1.09	0.73	1.01

that scalar product is negative, i.e., it indicates whether the calculated dipole moment vector is approximately parallel or antiparallel to the correct one. (This sign is discussed further in the Parameterization section.) Because of the monofunctional nature of the compounds in the database, these two limiting cases prevail.

Database

The primary database, used for stages 1 and 2 of the parameterization, consists of the experimental gas-phase dipole moments of 186 neutral organic molecules and theoretical atomic partial charges for 22 ions. The experimental dipole moments were taken from four standard compilations [27] and several additional sources [28]. The experimental dipole moments vary in their relative precision but are typically precise to within ±0.02 D [27]. Atomic partial charges in the ions were calculated using the ChelpG [15] method at the MP2 [21] level of theory with the cc-pVTZ [23] basis set.

The molecules in the primary database are mainly monofunctional organic molecules that have been chosen to isolate possible systematic deficiencies in the AM1 and PM3 wave functions. We chose small molecules with well-defined geometries and, for neutral molecules, accurate experimental dipole moments to eliminate ambiguities related to thermal conformational averaging, thereby permitting easy identification and correction of systematic errors in the relevant class II charge distributions.

In 24 cases, we allowed flexible molecules to enter the primary dipole database. These molecules have internal rotations that do not preserve axial symmetry. However, experimental dipole moments are known for at least one unique conformer in each of the 24 cases, and the conformer is specified explicitly in Tables 1–8. For example, experimental dipole moments for both the s-cis and s-trans isomers of formic acid are known [28] and both are included in the data set; see Table 1.

The number of experimental dipole moments used to parameterize c_k and d_k for a given halogen varies considerably, from 31 fluorides to only five iodides; this variation is dictated in part by the available data. Similarly, compounds in the sulfur database include thiols, sulfides and thioaldehydes while the silicon database includes only silanes. The total number of compounds containing halogens, Si or S in the primary database is 83.

The primary data set, with the exception of hydroxylamine, contains no formal heteroatom–heteroatom bonds, e.g., N-O, S-N and S-O. These were considered in stage 3. The database used for stage 3, which we will refer to as the secondary database, includes nine molecules for which we use an MP2/cc-pVDZ calculated dipole moment (evaluated from the MP2 density at the MP2-optimized geometry) and partial charges for three ions calculated at the MP2/cc-pVTZ level using the ChelpG [15] method.

All ab initio dipole moments and partial charges were calculated using GAUSSIAN92 [29]. All semiempirical calculations were performed with a locally modified version of AMSOL, version 4.0 [30].

The primary database of 186 neutral compounds is given in Tables 1–9. The parameterization of neutral molecules was based on experimental dipole moments, which are listed in Tables 1–8, together with those calculated from the AM1 and PM3 Mulliken partial charges and the CM1A and CM1P results, which are discussed further below. The parameterization of ions was based on high-level ab initio partial charges. Tables 10–14 give the results of CM1A and CM1P for the set of 25 ions used here. The ions are divided into four different groups: those containing H, C and N; those containing H, C and O; those containing H, C, N and O; and those containing H, C and S. The results listed in these tables include the dipole moments, partial charges of the heteroatoms, and partial charges of the hydrogens attached to heteroatoms. Errors in the dipole moments are given relative to the MP2/cc-pVTZ wave function, and errors in the partial charges are given relative to ChelpG partial charges at the MP2/cc-pVTZ level. The secondary database will be given below.

Parameterization

The parameterization was carried out in three stages.

Stage 1
In the first stage we considered 103 neutral molecules and 22 ions with H-, C-, N- and O-containing functionalities.

The dipole moments were nonlinearly related to the predicted charges and hence to the parameters by Eq. 10, and the final values of the parameters were obtained by a combination of locally minimizing the rms deviation between the theoretical and experimental dipole moments using the Levenberg–Marquardt method [31] for the neutral molecules and comparing CM1 partial charges with MP2/cc-pVTZ [21–23] ChelpG [15] partial charges for ions. In the following, we present the most important issues that arose during this parameterization.

Recognizing the need for more negative charge on N and O, we began by allowing the constant offsets, $\hat{d}_k$, for these atoms to vary while all $d_{kk'}$ remained zero; this resulted in a considerable improvement in the rms error in the dipole moment. Including scale factors, $\hat{c}_k$, for N and O produced little further improvement (< 0.01 D) beyond that accomplished by the offsets. However, using AM1 Mulliken charges, the offset on N resulted in unreasonably large partial charges for N in the protonated amines. Thus, for CM1A a scale factor is used for nitrogen instead of an offset.

At this point large errors remained on carboxylic acids, alcohols, amines and amides. These were reduced by

TABLE 2
DIPOLE MOMENTS (D) OF MOLECULES CONTAINING H, C AND N AND POSSIBLY O

Compound	Experiment	AM1		PM3	
		Mulliken	CM1A	Mulliken	CM1P
Amines					
Ammonia	1.47	0.64	1.75	0.00	1.32
Methylamine	1.31	0.47	1.40	0.08	1.21
Aziridine	1.89	0.83	1.78	0.39	1.48
Dimethylamine	1.01	0.34	1.09	0.10	0.98
Ethylamine	1.22	0.47	1.37	0.18	1.14
Cyclopropylamine	1.19	0.35	1.19	0.27	1.01
2-Aminopropane	1.19	0.61	1.49	0.23	1.13
Trimethylamine	0.612	0.18	0.83	0.05	0.56
1,2,5,6-Tetrahydropyridine (NH *eq.*)	0.99	0.34	1.06	0.21	0.96
1,2,5,6-Tetrahydropyridine (NH *ax.*)	1.007	0.22	0.92	0.23	0.89
Aniline	1.53	0.85	1.33	0.63	1.49
Piperidine (NH *eq.*)	0.82	0.47	1.09	0.31	0.78
Piperidine (NH *ax.*)	1.189	0.42	1.13	0.19	0.95
Imines and nitrogen aromatics					
N-Methylformaldimine	1.53	0.63	1.38	0.39	1.35
Z-Acetaldimine	2.058	1.23	2.27	0.81	1.90
E-Acetaldimine	2.56	1.13	2.09	0.67	1.67
N-Methylacetaldimine	1.498	0.68	1.50	0.44	1.44
Pyrrole	1.74	1.37	1.26	1.89	2.12
Pyridine	2.15	0.92	1.54	0.63	1.65
2-Methylpyrimidine	1.676	0.23	1.02	0.11	1.14
Multifunctional nitrogen-containing compounds					
Hydroxylamine	0.59	0.65	0.29	1.11	0.98
Cyanamide	4.32	2.22	3.75	2.43	4.44
Aminoacetonitrile	2.64	1.88	2.52	2.36	3.05
Isoxazole	1.503	0.63	1.17	0.58	1.33
4,5-Dihydroisoxazole	1.77	0.63	1.26	0.44	1.08
2-Cyanopyridine	5.78	3.16	5.01	2.96	5.25
3-Cyanopyridine	3.66	2.31	3.71	2.37	3.71
4-Cyanopyridine	1.96	1.52	2.55	1.88	2.45
Nitriles					
Hydrogen cyanide	2.985	1.49	2.99	1.54	3.08
Acetonitrile	3.925	2.24	3.79	2.29	3.89
Cyanoacetylene	3.724	2.23	3.86	2.27	3.92
Dicyanomethane	3.73	2.13	3.89	2.18	3.99
Acrylonitrile	3.56	2.35	3.92	2.36	3.96
Propionitrile	4.01	2.32	3.87	2.38	3.96
Acetyl cyanide	3.45	2.55	3.56	2.39	3.51
Allene carbonitrile	4.28	2.64	4.19	2.68	4.27
E-2-Butenenitrile	4.75	2.84	4.40	2.80	4.40
Z-2-Butenenitrile	4.08	2.54	4.07	2.54	4.13
Methacrylonitrile	3.69	2.36	3.94	2.37	3.97
Cyclopropane carbonitrile	4.131	2.64	4.24	2.64	4.29
Isobutyronitrile	4.29	2.45	4.04	2.52	4.15
Cyclobutane carbonitrile	4.11	2.45	3.99	2.47	4.05
2,2-Dimethylpropionitrile	3.95	2.43	3.98	2.45	4.03
Cyclopentadiene-1-carbonitrile	4.25	2.62	4.19	2.63	4.23
Benzonitrile	4.18	2.72	4.30	2.75	4.36
Amides					
Formamide	3.73	3.11	3.11	2.70	3.56
Z-N-Methylformamide	3.86	2.94	2.96	2.51	3.56
Acetamide	3.76	3.21	3.27	2.86	3.68

introducing a positive $d_{kk'}$ for H. Because of the local-charge-balancing methodology in Eq. 7, the corresponding heteroatoms accept the excess negative charge required to accommodate the more positive protons, thereby improving the predicted dipole moments. In the case of hydrogen, $f^{(d)}(B_{Hk'})$ was simply set equal to the bond order, $B_{Hk'}$. Because all $\hat{d}_H$ are zero, the offset for hydrogen atoms is given by Eq. 11.

$$d_k\big|_{k=H} = \sum_{k'\neq k} B_{kk'} d_{Hk'} \qquad (11)$$

We set $d_{HC} = d_{HH} = 0$, and we allowed d_{HN} and d_{HO} to be non-zero (Table 16).

The main problem remaining with C, N, O and H compounds at this point was the CM1A treatment of nitriles. As shown in Table 2, AM1 underestimates the dipole moments for nitriles by a considerable amount. Typically, an additional 0.3 electrons on the nitrogen are required to correct the dipole moment to experiment. Because of the small Mulliken charge calculated for nitrile nitrogens by AM1, an unreasonably high scale factor would have been required to accomplish this transformation. Thus, we allow d_{NC} to become non-zero. Because this offset applies to nitriles only, we write $f^{(d)}(B_{kk'})$ from Eq. 6 as

$$f^{(d)}(B_{kC})\big|_{k=N} = \frac{1}{2}\tanh\left(\frac{B_{kk'}-2.3}{0.1}\right)+\frac{1}{2} \qquad (12)$$

The form of $f^{(d)}(B_{NC})$ ensures that it increases smoothly from 0 to 1 for individual NC bond orders greater than 2.2 and is essentially 1 for individual NC bond orders greater than 2.4. The value of d_{NC} is given in Table 16. Although this treatment is sufficient for neutral nitriles, further consideration is required for ionic nitriles. Examination of the partial charges in the nitrile ions revealed that it was necessary to eliminate the scale factor for those molecules as they begin with an AM1 Mulliken charge which is far more negative than in neutral nitriles. Thus, for nitrogen in CM1A, $f^{(c)}(B_{NC})$ takes the same form as $f^{(d)}(B_{NC})$, and $c_{kk'}$ is set equal to $-\hat{c}_k$. This yields Eq. 13:

$$c_k\big|_{k=N} = \hat{c}_N + \left[\sum_{k'=C}\left(\frac{1}{2}\tanh\left(\frac{B_{kk'}-2.3}{0.1}\right)+\frac{1}{2}\right)\right]c_{NC} \qquad (13)$$

where $\hat{c}_N$ is the previously determined scale factor for nitrogens. Thus, Eq. 13 is always equal to $\hat{c}_N$ unless the nitrogen has a bond order greater than about 2.2 to a single carbon, in which case the scale factor tends to zero. It should be noted that Eqs. 12 and 13 affect only nitrogens in CM1A. CM1P requires no special effort for nitriles.

For the molecules in the primary database, the CM1A parameters on H, N and O make N and O atoms more

TABLE 3
DIPOLE MOMENTS (D) OF FLUORINE COMPOUNDS CONTAINING H, C, N, O AND F

Compound	Experiment	AM1		PM3	
		Mulliken	CM1A	Mulliken	CM1P
Fluoromethane	1.858	1.57	1.48	1.14	1.38
Difluoromethane	1.978	2.02	1.91	1.46	1.74
Trifluoromethane	1.652	2.12	2.00	1.56	1.80
Cyanogen fluoride	2.17	0.18	2.12	0.56	2.36
Perfluoropropyne	1.71	1.41	1.65	1.33	1.81
Trifluoroethylene	1.32	1.62	1.40	1.35	1.42
Pentafluoroethane	1.54	2.02	1.87	1.59	1.77
1,1-Difluoroethylene	1.384	1.73	1.49	1.42	1.55
Z-1,2-Difluoroethylene	2.42	2.51	2.19	2.03	2.20
Fluoroethylene	1.427	1.45	1.28	1.17	1.33
Acetyl fluoride	2.96	3.02	3.29	2.73	3.24
1,1-Difluoroethane	2.27	2.26	2.14	1.75	2.06
1,1,1-Trifluoroethane	2.347	2.71	2.58	2.11	2.38
Trifluoroacetonitrile	1.262	0.58	1.14	0.19	1.15
Bis(trifluoromethyl)ether	0.54	0.40	0.69	0.55	0.72
3,3,3-Trifluoropropyne	2.36	2.83	2.67	2.43	2.70
1,1,1,2,2,3,3-Heptafluoropropane	1.62	2.23	2.08	1.71	1.90
Fluoroallene	1.97	1.68	1.49	1.44	1.58
1,1-Difluoroallene	2.07	2.42	2.16	2.14	2.27
Z-1-Fluoropropene	1.46	1.42	1.25	1.19	1.33
2-Fluoropropene	1.61	1.65	1.48	1.37	1.51
2-Fluoropropane	1.96	1.71	1.61	1.37	1.63
3,3,3-Trifluoropropene	2.433	3.01	2.87	2.44	2.71
2-Fluoro-2-methylpropane	1.959	1.76	1.65	1.42	1.69
Fluorobenzene	1.60	1.74	1.55	1.49	1.62
m-Difluorobenzene	1.58	1.74	1.54	1.48	1.60
o-Difluorobenzene	2.59	2.98	2.61	2.56	2.70
o-Fluorotoluene	1.37	1.55	1.36	1.34	1.46
m-Fluorotoluene	1.86	1.95	1.76	1.67	1.80
p-Fluorotoluene	2.00	2.08	1.88	1.79	1.92
1,2,3,4-Tetrafluorobenzene	2.42	2.94	2.51	2.51	2.57

TABLE 4
DIPOLE MOMENTS (D) OF SILICON COMPOUNDS CONTAINING H, C AND Si

Compound	Experiment	AM1		PM3	
		Mulliken	CM1A	Mulliken	CM1P
Methylsilane	0.735	0.28	0.73	0.18	0.94
Vinylsilane	0.66	0.40	0.81	0.24	0.55
Ethylsilane	0.81	0.26	0.71	0.23	0.99
Dimethylsilane	0.713	0.29	0.79	0.16	1.00
Trimethylsilane	0.52	0.23	0.63	0.10	0.81
Benzylsilane	0.845	0.23	0.71	0.54	0.25

negative by an average of 0.34 and 0.10 e⁻, respectively and the CM1P method (which uses fewer bond-order-dependent values $c_{kk'}$ and $d_{kk'}$) makes these atoms more negative by an average of 0.38 and 0.13 e⁻, respectively.

Thus, in stage 1, Δq_k used in Eqs. 3 and 7 can be written for CM1A as

$$\Delta q_k = \left[\hat{c}_k + \delta_{kN}\left(\sum_{k'=C}\left(\frac{1}{2}\tanh\left(\frac{B_{kk'} - 2.3}{0.1}\right) + \frac{1}{2}\right)\right)c_{NC}\right]q_k^{(0)}$$
$$+ \hat{d}_k + \delta_{kH}\sum_{k'\neq k}B_{kk'}d_{Hk'} \tag{14}$$
$$+ \delta_{kN}\sum_{k'=C}\left(\frac{1}{2}\tanh\left(\frac{B_{kk'} - 2.3}{0.1}\right) + \frac{1}{2}\right)d_{NC}$$

while in CM1P it takes on the simpler form of

$$\Delta q_k = \hat{c}_k q_k^{(0)} + \hat{d}_k + \delta_{kH}\sum_{k'\neq k}B_{kk'}d_{Hk'} \tag{15}$$

where δ_{kX} is a Kronecker delta that is 1 if atom k has the atomic number of atom X and 0 otherwise. Although the form of the CM1A charge appears far more complex, it should be emphasized that the nitrogen–carbon specific parameters only affect nitrile nitrogens.

Stage 2

In stage 2, with the previously optimized H, C, N and O parameters held frozen, we considered parameterization

of the 83 compounds containing F, Cl, Br, I, Si and S. Optimizations of parameters were carried out again by using the Levenberg–Marquardt method [31]. First we considered the 68 halogen-containing molecules in Tables 3 and 6–8. The parameterization of the halogenated hydrocarbons proceeded best when both scale factors and offsets for the halogen atoms were allowed to vary. For halides, since the largest hydrogen–halogen bond order for any molecule in the data set is only 0.02, we set $d_{HF} = d_{HCl} = d_{HBr} = d_{HI} = 0$.

Compounds containing Si and S, listed in Tables 4 and 5, were optimized similarly. An investigation into allowing d_{HS} to be non-zero gave no improvement; the opposite is true for d_{HSi} for silicon. Thus, we took $d_{HS} = 0$, but we allowed d_{HSi} to vary.

For most compounds, the calculated dipole moments both before and after mapping are approximately parallel to the experimental ones, and so we used the + sign in Eq. 10. One exception is iodoethylene (Table 8), which in AM1 (using $q_k^{(0)}$) is initially antiparallel to the direction determined by high-level ab initio methods for iodomethane, but is made parallel by the CM1A mapping.

Stage 3

The molecules in the primary database contain no S-O bonds and only one N-O bond, the latter occurring in H_2N-OH. In stage 3, we developed a secondary database of compounds with C-S-O and C-N-O linkages. In addition to dipole moments, discussed above, we calculated ChelpG [15] partial charges from MP2 densities at MP2-

TABLE 5
DIPOLE MOMENTS (D) OF SULFUR COMPOUNDS CONTAINING H, C, N, O AND S

Compound	Experiment	AM1		PM3	
		Mulliken	CM1A	Mulliken	CM1P
Hydrogen sulfide	0.97	0.41	1.12	0.14	0.61
Methanethiol	1.52	0.41	1.28	0.55	1.23
Thioformaldehyde	1.647	0.12	1.70	0.40	1.69
Thioacetaldehyde	2.33	0.84	2.32	1.11	2.36
Ethanethiol	1.58	0.63	1.37	0.79	1.39
Dimethyl sulfide	1.50	0.24	1.36	0.53	1.43
Dicyanogen sulfide	3.02	3.10	2.85	2.61	3.07
Thietane	1.85	0.81	2.15	0.92	2.06
Thiophene	0.55	2.16	0.43	1.01	0.69

TABLE 6
DIPOLE MOMENTS (D) OF CHLORIDES CONTAINING H, C, N, O AND Cl

Compound	Experiment	AM1		PM3	
		Mulliken	CM1A	Mulliken	CM1P
Chloromethane	1.892	1.47	1.74	0.81	1.76
Dichloromethane	1.60	1.48	1.77	0.68	1.83
Trichloromethane	1.04	1.16	1.39	0.41	1.46
Cyanogen chloride	2.80	1.41	2.79	2.74	2.86
Chloroacetylene	0.44	0.26	0.47	0.85	0.60
Pentachloroethane	0.92	1.02	1.23	0.24	1.33
1,1-Dichloroethylene	1.34	1.24	1.52	0.11	1.43
Z-1,2-Dichloroethylene	1.90	1.47	1.89	0.07	2.00
Chloroethylene	1.452	1.17	1.44	0.32	1.44
1,1,1-Trichloroethane	1.78	1.84	2.11	0.78	1.90
Acetyl chloride	2.72	2.65	2.99	2.23	2.90
Chloroethane	2.05	1.67	1.96	1.03	1.97
3-Chloropropyne	1.68	1.49	1.77	0.85	1.82
Z-1-Chloropropene	1.67	1.23	1.49	0.45	1.47
E-1-Chloropropene	1.97	1.48	1.75	0.60	1.70
2-Chloropropene	1.647	1.33	1.60	0.46	1.53
3-Chloropropene	1.94	1.47	1.76	0.81	1.77
o-Dichlorobenzene	2.50	2.03	2.48	0.38	2.45
m-Dichlorobenzene	1.72	1.27	1.54	0.34	1.52
Chlorobenzene	1.69	1.34	1.61	0.42	1.58
o-Chlorotoluene	1.56	1.17	1.43	0.36	1.42
p-Chlorotoluene	2.21	1.70	1.97	0.72	1.88

optimized geometries. Neutral molecules were treated using the cc-pVDZ basis set, and ion calculations used the cc-pVTZ basis set. We compared CM1 dipole moments to ab initio wave-function-derived dipole moments, and CM1 charges to ChelpG [15] charges.

First, we considered applying the CM1 parameterization developed in stage 2 to the compounds with C-S-O and C-N-O linkages. The PM3-CM1P model performed acceptably with no change of parameters, and thus no bond-order-dependent parameters were introduced for S, N or O. However, the performance of the stage 2 AM1-CM1A model was not satisfactory for either S-O linkages or N-O linkages. Therefore, for AM1 oxygen bonded to sulfur, we write

$$d_k\big|_{k=O} = \hat{d}_O + \sum_{k'=S} B_{kk'} d_{OS} \qquad (k = O) \qquad (16)$$

where $\hat{d}_O$ is the O offset determined in stage 1. We then optimized d_{OS} by considering the rms deviation from the MP2/cc-pVDZ dipole moments and also the deviation from ChelpG [15] partial charges. For the six compounds with C-S-O linkages in the secondary database, this yielded $d_{OS} = -0.06$. Similarly, for AM1 nitrogen bonded to oxygen, we add an offset using Eq. 17:

$$d_k\big|_{k=N} = \hat{d}_N + \sum_{k'=C} \left(\frac{1}{2} \tanh\left(\frac{B_{kk'} - 2.3}{0.1} \right) + \frac{1}{2} \right) d_{NC} + \sum_{k'=O} B_{kk'} d_{NO} \qquad (17)$$

Again, $\hat{d}_N$ and d_{NC} remain the same values that were found in stage 1. The results from this secondary analysis are discussed below.

TABLE 7
DIPOLE MOMENTS (D) OF BROMIDES CONTAINING H, C AND Br

Compound	Experiment	AM1		PM3	
		Mulliken	CM1A	Mulliken	CM1P
Bromomethane	1.821	0.91	1.73	1.36	1.66
Dibromomethane	1.43	0.69	1.47	1.13	1.55
Tribromomethane	0.99	0.40	0.96	0.60	1.05
Bromoacetylene	0.23	0.18	0.23	0.53	0.19
Bromoethylene	1.42	0.74	1.46	1.17	1.49
Bromoethane	2.03	1.16	2.02	1.71	2.01
3-Bromopropyne	1.54	0.92	1.74	1.52	1.83
1-Bromopropane	2.18	1.21	2.07	1.69	1.99
2-Bromopropane	2.21	1.35	2.24	1.96	2.26
Bromobenzene	1.70	0.91	1.63	1.09	1.42

TABLE 8
DIPOLE MOMENTS (D) OF IODIDES CONTAINING H, C AND I

Compound	Experiment	AM1		PM3	
		Mulliken	CM1A	Mulliken	CM1P
Iodomethane	1.647	0.00	1.68	1.32	1.61
Iodoethylene	1.30	−0.36	1.46	1.07	1.59
Iodoethane	1.91	0.25	1.96	1.77	1.94
1-Iodopropane	2.04	0.32	2.01	1.73	1.93
Iodobenzene	1.71	0.01	1.56	0.82	1.54

Final parameters

With the additions described above, Eq. 14 is replaced by

$$\Delta q_k = \left[\hat{c}_k + \delta_{kN} \left(\sum_{k'=C} \left(\frac{1}{2} \tanh\left(\frac{B_{Nk'} - 2.3}{0.1} \right) + \frac{1}{2} \right) \right) c_{NC} \right] q_k^{(0)}$$

$$+ \hat{d}_k + \delta_{kH} \sum_{k' \neq k} (B_{kk'} d_{Hk'}) + \delta_{kO} \sum_{k'=S} B_{kk'} d_{OS} \qquad (18)$$

$$+ \delta_{kN} \left[\sum_{k'=C} \left(\frac{1}{2} \tanh\left(\frac{B_{kk'} - 2.3}{0.1} \right) + \frac{1}{2} \right) d_{NC} + \sum_{k'=O} B_{kk'} d_{NO} \right]$$

where δ_{kX} is a Kronecker delta that is 1 if atom k has the atomic number of atom X and 0 otherwise. Equation 15 remains valid for the CM1P model. Both models, as defined by Eqs. 3, 7, 15 and 18, are valid for compounds containing C, H, N, O, F, Si, S, Cl, Br and I.

Table 16 gives the CM1A and CM1P parameters used to generate the mappings of the AM1 or PM3 Mulliken charges to the class IV charges. The scale factors and offsets are listed by atom type and charge model. These parameters define the CM1 models using Eqs. 3, 7, 15 and 18. We emphasize that the scale factors and offsets can only be applied in the context of the AM1 and PM3 wave functions, because Eqs. 15 and 18 also employ the AM1 or PM3 Mulliken charges, $q_k^{(0)}$, as well as the covalent bond indices [26], $B_{kk'}$, derived from the semi-empirical wave function.

Results and Discussion

Tables 1–8 list five values of the dipole moment for each of the 186 neutral molecules in the primary data-base. The tables are divided by both functional group type and atom type. The first dipole moment listed for

TABLE 9
RMS ERRORS (D) FOR GROUPS OF COMPOUNDS WITH VARIOUS FUNCTIONAL GROUPS USED IN STAGES 1 AND 2 OF THE PARAMETERIZATION

Compound	No.	AM1			PM3		
		Density	Mulliken	CM1A	Density	Mulliken	CM1P
H, C, N, O compounds							
Alcohols	12	0.19	0.59	0.21	0.31	0.85	0.28
Esters, lactones	8	0.26	0.36	0.24	0.31	0.56	0.20
Aldehydes, ketones	16	0.21	0.45	0.18	0.38	0.70	0.22
Acids	9	0.31	0.63	0.41	0.22	0.62	0.39
Ethers	10	0.21	0.30	0.18	0.23	0.49	0.20
Amines	13	0.29	0.73	0.17	0.26	1.01	0.16
Nitriles	17	0.93	1.55	0.17	0.66	1.53	0.18
Amides	3	0.14	0.71	0.69	0.47	1.11	0.20
Imines, N-aromatics	7	0.28	1.07	0.43	0.35	1.33	0.47
Multifunctional N	8	0.83	1.41	0.47	0.64	1.42	0.41
All H, C, N, O	103	0.49	0.92	0.29	0.42	1.03	0.27
Other compounds							
Fluorides	31	0.37	0.48	0.24	0.29	0.46	0.21
Silicon compounds	12	0.29	0.45	0.11	0.46	0.48	0.32
Sulfur compounds	9	0.49	1.17	0.17	0.58	0.91	0.19
Chlorides	22	0.46	0.44	0.18	0.67	1.17	0.20
Bromides	10	0.29	0.75	0.08	0.24	0.37	0.16
Iodides	5	0.38	1.68	0.10	0.44	0.46	0.16
All other compounds	83	0.40	0.73	0.19	0.47	0.76	0.21
All compounds	186	0.45	0.84	0.25	0.45	0.92	0.25

each molecule is from experiment. The second and third are calculated from the AM1 Mulliken and CM1A class IV charges, respectively, at the AM1-optimized geometries. The fourth and fifth values of the dipole moment are derived from PM3-optimized geometries and the PM3 Mulliken and CM1P class IV charges, respectively. Thus, all of the calculated dipole moments in Tables 1–8 are calculated from the optimized Cartesian coordinates of each molecule and a specified set of partial atomic charges.

An analysis of the rms errors relative to experiment over the set of 186 compounds in the primary database is given in Table 9. The same organic functional groups found in Tables 1–8 are listed in Table 9, together with the number of entries in the database followed by six rms errors given relative to experiment. The six entries correspond to the four theoretical columns in Tables 1–8, plus two additional columns labeled AM1 density and PM3 density. The 'density' columns give the rms deviations between experiment and the AM1 or PM3 dipole moments derived from the full continuous probability densities. Additionally, there are three summarizing rows in which the cumulative rms errors are reported over the set of 103 H-, C-, N- and O-containing compounds, the

TABLE 10
DIPOLE MOMENTS (D) AND PARTIAL CHARGES FOR IONS CONTAINING H, C AND N

Compound	MP2/cc-pVTZ[a]	AM1		PM3	
		Mulliken	CM1A	Mulliken	CM1P
Dipole moments					
CH$_2$CN$^-$	1.33	0.33	1.25	0.46	2.08
CH$_3$CNH$^+$	1.39	0.41	0.42	2.05	1.67
CH$_3$NH$^-$	2.32	1.04	3.45	1.90	2.55
CH$_3$NH$_3^+$	2.19	1.87	2.22	2.24	2.95
CN$^-$	0.77	0.13	-1.30	0.09	-1.39
HCNH$^+$	0.19	0.98	1.75	0.26	0.03
(CH$_3$)$_2$NH$_2^+$	1.50	1.09	1.51	1.45	2.26
(CH$_3$)$_3$NH$^+$	0.87	0.50	0.87	0.76	1.47
NH$_2^-$	1.58	0.06	2.64	0.07	1.73
NH$_4^+$ [b]	0.00	0.00	0.00	0.00	0.00
Mean deviation		-0.67	0.07	-0.32	0.13
Mean unsigned deviation		0.84	0.77	0.49	0.65
Rms deviation		0.93	1.06	0.68	0.88
N charges					
CH$_2$CN$^-$	-0.87	-0.43	-0.68	-0.52	-0.77
CH$_3$CNH$^+$	-0.34	-0.10	-0.47	0.45	-0.06
CH$_3$NH$^-$	-1.26	-0.75	-1.51	-0.82	-1.20
CH$_3$NH$_3^+$	-0.34	-0.06	-0.36	0.88	-0.03
CN$^-$	-0.44	-0.51	-0.77	-0.52	-0.79
HCNH$^+$	-0.18	-0.06	-0.43	0.49	-0.03
(CH$_3$)$_2$NH$_2^+$	-0.14	-0.03	-0.21	0.78	0.06
(CH$_3$)$_3$NH$^+$	-0.01	0.00	-0.06	0.69	0.16
NH$_2^-$	-1.28	-0.90	-1.75	-0.90	-1.45
NH$_4^+$	-0.77	-0.09	-0.51	1.00	-0.11
Mean deviation		0.27	-0.11	0.72	0.14
Mean unsigned deviation		0.29	0.20	0.73	0.24
Rms deviation		0.35	0.24	0.76	0.26
H bonded to N charges					
CH$_3$CNH$^+$	0.52	0.41	0.53	0.20	0.45
CH$_3$NH$^-$	0.22	-0.02	-0.02	-0.01	0.26
CH$_3$NH$_3^+$ [c]	0.34	0.26	0.35	0.02	0.29
HCNH$^+$	0.51	0.41	0.53	0.20	0.45
(CH$_3$)$_2$NH$_2^+$ [c]	0.31	0.25	0.34	0.04	0.31
(CH$_3$)$_3$NH$^+$	0.32	0.25	0.32	0.05	0.32
NH$_2^-$ [c]	0.14	-0.05	0.38	-0.05	0.22
NH$_4^+$ [c]	0.44	0.27	0.38	0.00	0.28
Mean deviation		-0.13	0.01	-0.32	-0.05
Mean unsigned deviation		0.13	0.07	0.32	0.08
Rms deviation		0.14	0.11	0.33	0.10

Deviations are relative to MP2/cc-pVTZ calculations.
[a] Dipole moments are density-derived; partial charges are derived using the ChelpG algorithm.
[b] The dipole moment of ammonium is not counted in the deviation statistics, as it is zero by symmetry.
[c] Multiple H bonded to N were counted the corresponding number of times in computing average deviations but are listed only once.

set of 83 F, Si, S, Cl, Br and I compounds, and the full set of 186 compounds. The results of the CM1 models on the 25 ions used in the parameterization are presented in Tables 10–14.

For a representative subset of 23 compounds, a comparison is made in Table 15 with respect to results from various levels of ab initio theory. Dipole moments are calculated from both the MP2/6-31G*//HF/6-31G* [24] and HF/6-31G* [22] densities. The dipole moments calculated with the dipole moment operator and the continuous charge density are labeled 'density'. The remaining four columns under the heading HF/6-31G* give dipole moments calculated from partial charges calculated from the wave function, and the subheading indicates how those partial charges were determined.

The Mulliken charge on atom k for an ab initio wave function is given by

$$q_k = Z_k - \sum_{\mu \in k} (\mathbf{PS})_{\mu\mu} \qquad (19)$$

where $\mathbf{S}$ is the overlap matrix. In addition to the dipole moments predicted by Mulliken partial charges, we have included those from ChelpG [15], Merz–Kollman (MK) [16] and natural population analysis (NPA) [17]. The ab initio charges and dipole moments were calculated at either the MP2/6-31G*//HF/6-31G* or HF/6-31G* levels [21,22,24].

The ChelpG and Merz–Kollman methods used to calculate charges from the HF/6-31G* electron density are similar in that they generate charges which best repro-

TABLE 11
DIPOLE MOMENTS (D) AND PARTIAL CHARGES FOR IONS CONTAINING H, C AND O

Compound	MP2/cc-pVTZ[a]	AM1		PM3	
		Mulliken	CM1A	Mulliken	CM1P
Dipole moments					
$CH_3COCH_2^-$	2.91	3.45	3.58	3.54	3.74
CH_3COO^-	3.26	3.13	3.43	3.27	3.73
CH_3O^-	1.63	0.82	1.17	1.22	1.68
H_3O^+	1.53	1.64	2.33	1.25	1.99
OH^-	1.35	0.36	1.14	0.51	1.36
HOO^-	2.65	1.99	2.53	2.18	2.63
$(CH_3)_2COH^+$	1.55	2.30	2.61	1.77	2.26
$(CH_3)_2OH^+$	1.20	0.91	1.45	0.89	1.53
$CH_3OH_2^+$	1.90	1.43	2.14	1.50	2.19
Mean deviation		−0.22	0.27	−0.21	0.35
Mean unsigned deviation		0.53	0.44	0.40	0.35
Rms deviation		0.60	0.54	0.46	0.45
O charges					
$CH_3COCH_2^-$	−0.78	−0.54	−0.59	−0.57	−0.66
CH_3COO^-	−0.79	−0.59	−0.64	−0.64	−0.71
CH_3COO^-	−0.80	−0.60	−0.64	−0.65	−0.72
CH_3O^-	−0.98	−0.76	−0.81	−0.85	−0.91
H_3O^+	−0.45	−0.11	−0.50	0.05	−0.42
OH^-	−1.19	−1.02	−1.19	−1.06	−1.24
HOO^-	−0.75	−0.80	−0.81	−0.77	−0.77
$H\underline{O}O^-$	−0.54	−0.23	−0.39	−0.35	−0.53
$(CH_3)_2COH^+$	−0.35	−0.28	−0.43	−0.25	−0.45
$(CH_3)_2OH^+$	−0.24	−0.10	−0.28	0.01	−0.22
$CH_3OH_2^+$	−0.42	−0.11	−0.40	0.02	−0.33
Mean deviation		0.19	0.05	0.20	0.03
Mean unsigned deviation		0.20	0.10	0.21	0.06
Rms deviation		0.22	0.12	0.25	0.07
H bonded to O charges					
$H_3O^{+\,b}$	0.48	0.37	0.50	0.32	0.47
OH^-	0.19	0.02	0.19	0.06	0.24
HOO^-	0.29	0.03	0.20	0.11	0.30
$(CH_3)_2COH^+$	0.46	0.35	0.48	0.33	0.47
$(CH_3)_2OH^+$	0.47	0.33	0.46	0.29	0.45
$CH_3OH_2^{+\,b}$	0.48	0.34	0.48	0.30	0.46
Mean deviation		−0.14	0.00	−0.16	0.00
Mean unsigned deviation		0.14	0.02	0.16	0.02
Rms deviation		0.15	0.03	0.16	0.02

Deviations are relative to MP2/cc-pVTZ calculations.
[a] Dipole moments are density-derived; partial charges are derived using the ChelpG algorithm.
[b] Multiple H bonded to O were counted the corresponding number of times in computing average deviations but are listed only once.

TABLE 12
DIPOLE MOMENTS (D) AND PARTIAL CHARGES FOR IONS CONTAINING C, N AND O

Compound	MP2/cc-pVTZ[a]	AM1		PM3	
		Mulliken	CM1A	Mulliken	CM1P
Dipole moments					
NO_2^-	0.16	1.51	0.76	1.95	1.52
OCN^-	1.49	0.00	1.67	0.69	1.72
$NO_3^{-\,b}$	0.00	0.00	0.00	0.00	0.00
Mean deviation		−0.07	0.39	0.49	0.79
Mean unsigned deviation		1.42	0.39	1.29	0.79
Rms deviation		1.42	0.44	1.38	0.97
N charges					
NO_2^-	−0.17	0.20	−0.05	0.31	0.18
OCN^-	−0.91	−0.50	−0.81	−0.62	−0.85
NO_3^-	0.96	0.70	0.92	1.43	1.26
Mean deviation		0.17	0.06	0.41	0.23
Mean unsigned deviation		0.34	0.09	0.41	0.23
Rms deviation		0.35	0.09	0.42	0.27
O charges					
$NO_2^{-\,c}$	−0.41	−0.60	−0.47	−0.66	−0.59
OCN^-	−0.70	−0.54	−0.55	−0.54	−0.59
$NO_3^{-\,c}$	−0.65	−0.57	−0.64	−0.81	−0.75
Mean deviation		0.01	0.01	−0.13	−0.09
Mean unsigned deviation		0.13	0.05	0.19	0.12
Rms deviation		0.14	0.07	0.19	0.13

Deviations are relative to MP2/cc-pVTZ calculations.

[a] Dipole moments are density-derived; partial charges are derived using the ChelpG algorithm.

[b] The dipole moment of nitrate is not counted in the deviation statistics, as it is zero by symmetry.

[c] Multiple O bonded to N were counted the corresponding number of times in computing average deviations but are listed only once.

duce the molecular electrostatic potential (calculated from the densities) at various points away from the center of the molecule. Table 15 shows that, as expected, the dipole moments derived from these two methods are similar.

Natural population analysis orthogonalizes the atomic and molecular orbitals in separate steps in order to reduce

TABLE 13
DIPOLE MOMENTS (D) AND PARTIAL CHARGES FOR IONS CONTAINING H, C AND S

Compound	MP2/cc-pVTZ[a]	AM1		PM3	
		Mulliken	CM1A	Mulliken	CM1P
Dipole moments					
H_3S^+	1.58	0.10	2.22	0.48	1.10
HS^-	0.65	0.56	0.29	0.00	0.02
$CH_3SH_2^+$	1.30	1.46	1.46	1.07	0.82
Mean deviation		−0.47	0.15	−0.66	−0.53
Mean unsigned deviation		0.57	0.39	0.66	0.53
Rms deviation		0.86	0.44	0.75	0.53
S charges					
H_3S^+	0.40	0.88	0.25	1.07	0.55
HS^-	−1.07	−1.06	−1.02	−0.97	−0.97
$CH_3SH_2^+$	0.27	1.01	0.32	0.92	0.44
Mean deviation		0.41	−0.01	0.47	0.14
Mean unsigned deviation		0.41	0.08	0.47	0.14
Rms deviation		0.51	0.10	0.54	0.14
H bonded to S charges					
$H_3S^{+\,b}$	0.20	0.04	0.25	−0.02	0.15
$CH_3SH_2^{+\,b}$	0.22	−0.01	0.21	−0.01	0.15
Mean deviation		−0.19	0.03	−0.23	−0.06
Mean unsigned deviation		0.19	0.03	0.23	0.06
Rms deviation		0.19	0.04	0.23	0.06

Deviations are relative to MP2/cc-pVTZ calculations.

[a] Dipole moments are density-derived; partial charges are derived using the ChelpG algorithm.

[b] Multiple H bonded to S were counted the corresponding number of times in computing average deviations but are listed only once.

TABLE 14

SUMMARY OF AVERAGE DEVIATIONS FROM MP2/cc-pVTZ DENSITY-DERIVED DIPOLE MOMENTS (D) AND MP2/cc-pVTZ ChelpG PARTIAL CHARGES FOR 25 IONS

Parameter	AM1		PM3	
	Mulliken	CM1A	Mulliken	CM1P
Dipole moments				
Mean deviation	−0.37	0.17	−0.23	0.17
Mean unsigned deviation	0.66	0.51	0.50	0.49
Rms deviation	0.82	0.74	0.68	0.68
N charges				
Mean deviation	0.25	−0.07	0.65	0.16
Mean unsigned deviation	0.30	0.17	0.66	0.24
Rms deviation	0.35	0.22	0.79	0.29
O charges				
Mean deviation	0.13	0.04	0.08	−0.01
Mean unsigned deviation	0.18	0.08	0.20	0.08
Rms deviation	0.20	0.10	0.23	0.10
S charges				
Mean deviation	0.51	0.10	0.54	0.14
Mean unsigned deviation	0.41	−0.01	0.47	0.14
Rms deviation	0.41	0.08	0.47	0.14
H bonded to N, O, S charges				
Mean deviation	0.16	0.09	0.27	0.07
Mean unsigned deviation	−0.15	0.00	−0.25	−0.03
Rms deviation	0.15	0.06	0.25	0.06

the influence of the overlap matrix S on the partitioning of charge between atoms in molecules [17]. The inherently arbitrary nature of such a partitioning is what makes every partial charge methodology unique. Dipole moments calculated from these partial charges were used in the final column of the HF/6-31G* section of Table 15.

TABLE 15

GENERAL COMPARISON BETWEEN EXPERIMENTAL AND CALCULATED DIPOLE MOMENTS (D)

Molecule	Exp.	MP2[a] density	HF/6-31G*					AM1			PM3		
			Density	Mulliken	ChelpG	MK	NPA	Density	Mulliken	CM1A	Density	Mulliken	CM1P
Water	1.85	2.16	2.20	2.39	2.25	2.25	2.63	1.86	1.09	2.02	1.74	0.97	1.92
Methanol	1.70	1.77	1.87	2.68	1.85	1.86	3.00	1.62	1.14	1.63	1.49	0.93	1.59
Methylformate	1.77	1.87	2.04	1.85	2.03	2.05	2.24	1.43	1.42	1.56	1.52	1.49	1.70
Formaldehyde	2.33	2.17	2.67	3.15	2.67	2.66	3.97	2.32	2.02	2.42	2.16	1.83	2.41
Acetone	2.88	2.64	3.11	3.80	3.15	3.14	4.65	2.91	2.55	2.95	2.77	2.33	2.92
Cyclopropanone	2.67	2.58	3.15	3.65	3.19	3.16	2.43	2.39	1.97	2.37	2.21	1.67	2.28
Acetic acid	1.70	1.46	1.79	2.00	1.83	1.83	2.37	1.89	2.30	1.98	1.84	2.16	1.96
Dimethylether	1.30	1.50	1.62	2.72	1.63	1.63	3.04	1.68	1.31	1.53	1.49	1.11	1.44
Tetrahydrofuran	1.63	1.84	1.95	2.96	1.92	1.95	3.36	1.92	1.54	1.76	1.71	1.32	1.67
Furan	0.66	0.60	0.77	1.96	0.74	0.75	1.75	0.49	0.36	0.66	0.22	0.07	0.54
Ammonia	1.47	1.92	1.92	1.77	1.96	1.97	1.98	1.83	0.64	1.75	1.52	0.00	1.32
Methylamine	1.31	1.50	1.53	1.60	1.48	1.52	1.81	1.50	0.47	1.40	1.39	0.08	1.21
Hydrogen cyanide	2.99	2.96	3.21	3.65	3.16	3.18	3.08	2.36	1.49	2.99	2.70	1.54	3.08
Acetonitrile	3.93	3.75	4.04	4.75	4.03	4.07	3.79	2.89	2.24	3.79	3.21	2.29	3.89
Formamide	3.73	3.73	4.10	4.27	4.10	4.10	5.16	3.70	3.11	3.10	3.11	2.70	3.43
Acetamide	3.76	3.63	4.03	4.26	4.05	4.05	5.21	3.76	3.21	3.26	3.28	2.86	3.63
Cyanamide	4.32	4.36	4.56	3.91	4.58	4.60	3.92	3.31	2.22	3.75	3.44	2.43	4.44
Fluoromethane	1.86	1.78	1.99	3.41	2.00	1.99	3.68	1.62	1.57	1.48	1.44	1.14	1.38
Methylsilane	0.74	0.72	0.68	0.03	0.66	0.67	0.78	0.37	0.28	0.73	0.43	0.18	0.94
Hydrogen sulfide	0.97	1.49	1.41	0.96	1.52	1.56	1.14	1.86	0.41	1.12	1.78	0.14	0.61
Methanethiol	1.52	1.78	1.79	−0.86	1.75	1.85	0.71	1.76	0.41	1.28	1.95	0.55	1.23
Thioformaldehyde	1.65	1.70	2.23	1.64	2.27	2.28	0.88	1.84	0.12	1.70	2.07	0.40	1.69
Chloromethane	1.89	2.08	2.25	1.99	2.30	2.31	2.00	1.51	1.47	1.74	1.38	0.81	1.76
Rms error		0.21	0.31	0.93	0.33	0.34	1.05	0.44	0.89	0.27	0.43	1.00	0.20

[a] MP2/6-31G*//HF/6-31G*

Figures 1–11 illustrate the charge distributions in selected molecules found in Tables 1–6. Note that there are three rows of atomic partial charges for each molecule. The first row gives Mulliken and ChelpG charges, in that order, calculated from the HF/6-31G* wave function for each molecule. The second row contains the AM1 Mulliken and CM1A class IV charges, and the third row contains the PM3 Mulliken and CM1P class IV charges. Then, below each molecule is the experimental dipole moment and the set of dipole moments calculated from the six displayed sets of charges in the same order as noted above.

Discussion of results from the primary database

Methanol and alcohols Partial charges calculated for alcohols by both AM1 and PM3 are not negative enough at oxygen. The CM1 models place an additional charge of about –0.2 on oxygen at the expense, primarily, of the hydroxyl hydrogen, which becomes more positive by nearly as much; the attached carbon makes up the difference, as illustrated in Fig. 1. The CM1 dipole moments for methanol agree with experiment to within 0.1 D. The ab initio Mulliken and ChelpG dipole moments are larger as a result of a more negative oxygen and a more positive hydroxyl hydrogen. Since carbon and hydrogens attached to carbon are unaffected by the primary change of the CM1 mapping, their charges vary, if at all, only as a result of the renormalization step insofar as they have a non-zero bond order to oxygen. Table 9 shows that the rms errors of the CM1A and CM1P models for alcohol

dipole moments are about the same as for dipole moments calculated from the AM1 and PM3 densities, respectively. It appears that the CM1 mappings primarily make up for the error of replacing the continuous charge distribution by a distributed set of monopoles.

Formaldehyde, acetone, aldehydes, ketones and esters
Formaldehyde acts as an illustrative molecule containing a carbonyl group, as shown in Fig. 2. Table 9 shows low rms errors (between 0.24 and 0.18 D) in CM1 mappings for esters, lactones, aldehydes and ketones. In the CM1A mapping of formaldehyde, however, a charge of +0.14 on each hydrogen remains since they are not altered by the model. The CM1A and CM1P partial charges at C are quite different, but they do strike a compromise between the ChelpG and ab initio Mulliken charges. Thus, formaldehyde is complicated by its widely varying hydrogen charges, but we may generalize that the CM1 charges show less polarization of the carbonyl bond than Mulliken and ChelpG charges while achieving a better agreement with the experimental dipole moment. Acetone, also listed in Table 15, is treated well by both CM1A and CM1P, with dipole moment errors of less than 0.10 D. In fact, both CM1A and CM1P are more accurate than any of the ab initio methods in the case of acetone.

Formic acid The results for *s-cis*-formic acid are presented in Fig. 3. To make up for the generally too weakly polarized character of C-O bonds in AM1 and PM3, the carbonyl and hydroxyl oxygens in formic acid are made more negative by the charge models. The acid hydrogen becomes 0.05 more positive in the acid relative

TABLE 16
PARAMETERS OF THE CHARGE MODELS[a]

Atom type	AM1-CM1A		PM3-CM1P	
	$\hat{c}_k$ or $c_{kk'}$	$\hat{d}_k$ or $d_{kk'}$	$\hat{c}_k$	$\hat{d}_k$ or $d_{kk'}$
Stage 1				
H	0	0	0	0
H bonded to N	0	0.0850[b]	0	0.1854[b]
H bonded to O	0	0.1447[b]	0	0.1434[b]
C	0	0	0	0
N	0.3846	0	0	–0.0909
N bonded to C	–0.3846[c]	–0.0880[b]	0	0
O	0	–0.0283	0	–0.0449
Stage 2				
F	0.1468	0.0399	0.3381	0.0148
H bonded to Si	0	0.0640[b]	0	–0.1004[b]
Si	0	0	0	0
S	–0.1311	–0.0956	–0.0834	–0.0848
Cl	0.0405	–0.0276	–0.1080	–0.1168
Br	0.1761	–0.0802	–0.0116	–0.0338
I	0.2380	–0.1819	–0.3213	–0.0636
Stage 3				
O bonded to S	0	–0.0600[b]	0	0
N bonded to O	0	–0.0630[b]	0	0

[a] In most cases, the quantity tabulated is the single subscripted variable; exceptions are indicated by a footnote.
[b] $d_{kk'}$.
[c] $c_{kk'}$.

TABLE 17

DIPOLE MOMENTS, DEVIATIONS RELATIVE TO MP2/cc–pVDZ, AND PARTIAL CHARGES FOR MOLECULES WITH C-S-O LINKAGES

Level	μ (D)	Deviation	Partial Charge			
			C[a]	S	=O	–O or N
Dimethyl sulfone [(CH₃)₂SO₂]						
MP2/cc-pVDZ[b]	4.16		–0.36	1.00	–0.52	
AM1[c]	2.61	–1.55	–0.95	2.83	–0.93	
CM1A	2.83	–1.33	–0.64	1.30	–0.50	
CM1A//MP2/cc-pVDZ	3.23	–0.93	–0.51	1.07	–0.47	
PM3[c]	3.88	–0.28	–0.57	2.20	–0.84	
CM1P	4.33	0.17	–0.37	1.27	–0.58	
CM1P//MP2/cc-pVDZ	4.05	0.11	–0.37	1.25	–0.59	
Dimethyl sulfoxide [(CH₃)₂SO]						
MP2/cc-pVDZ[b]	3.89		–0.21	0.22	–0.43	
AM1[c]	4.92	1.03	–0.68	1.39	–0.78	
CM1A	4.09	0.20	–0.45	0.72	–0.58	
CM1A//MP2/cc-pVDZ	4.19	0.30	–0.39	0.61	–0.56	
PM3[c]	4.85	0.96	–0.38	0.94	–0.69	
CM1P	4.50	0.61	–0.24	0.51	–0.56	
CM1P//MP2/cc-pVDZ	4.46	0.57	–0.26	0.54	–0.57	
Methanesulfonic acid [CH₃SO₃H]						
MP2/cc-pVDZ[b]	2.33		–0.40	1.00	–0.47	–0.57
AM1[c]	1.77	–0.56	–0.97	2.85	–0.95	–0.78
CM1A	1.42	–0.91	–0.65	1.38	–0.51	–0.70
CM1A//MP2/cc-pVDZ	1.55	–0.78	–0.50	1.13	–0.47	–0.59
PM3[c]	2.71	0.38	–0.55	2.36	–0.86	–0.46
CM1P	2.49	0.16	–0.34	1.41	–0.58	–0.66
CM1P//MP2/cc-pVDZ	2.10	0.23	–0.36	1.36	–0.58	–0.61
Methyl methanesulfenate [CH₃S-OCH₃]						
MP2/cc-pVDZ[b]	0.28		–0.17	–0.12		–0.22
AM1[c]	1.28	1.00	–0.35	0.16		–0.31
CM1A	0.77	0.49	–0.24	0.00		–0.37
CM1A//MP2/cc-pVDZ	0.51	0.23	–0.22	–0.01		–0.36
PM3[c]	0.81	0.53	–0.20	0.09		–0.03
CM1P	0.17	–0.11	–0.10	–0.05		–0.31
CM1P//MP2/cc-pVDZ	0.22	0.06	–0.12	–0.04		–0.31
Methyl methanesulfinate [CH₃S(O)OCH₃]						
MP2/cc-pVDZ[b]	2.46		–0.28	0.31	–0.37	–0.31
AM1[c]	3.72	1.26	–0.48	1.47	–0.79	–0.55
CM1A	2.08	–0.38	–0.48	0.83	–0.57	–0.46
CM1A//MP2/cc-pVDZ	2.21	–0.25	–0.43	0.74	–0.55	–0.41
PM3[c]	3.66	1.20	–0.41	1.10	–0.71	–0.44
CM1P	2.94	0.48	–0.26	0.66	–0.54	–0.38
CM1P//MP2/cc-pVDZ	2.66	0.20	–0.29	0.68	–0.55	–0.37
Methanesulfonamide [CH₃SO₂NH₂]						
MP2/cc-pVDZ[b]	2.87		–0.27	1.03	–0.50	–0.81
AM1[c]	3.17	0.30	–0.96	2.83	–0.94	–0.96
CM1A	3.81	0.94	–0.65	1.38	–0.50	–1.19
CM1A//MP2/cc-pVDZ	2.45	–0.42	–0.50	1.13	–0.49	–1.02
PM3[c]	3.73	0.86	–0.55	2.17	–0.84	–0.46
CM1P	3.77	0.90	–0.34	1.31	–0.58	–0.87
CM1P//MP2/cc-pVDZ	3.07	0.20	–0.35	1.30	–0.59	–0.84

[a] In CH₃S-OCH₃ and CH₃S(O)OCH₃, we tabulate the partial charge of the C bonded to S.

[b] Dipole moments are density-derived; partial charges are derived using the ChelpG algorithm.

[c] AM1 and PM3 dipole moments are calculated from AM1 and PM3 partial charges.

to the change found for hydroxyl hydrogens in alcohols. This shift is similar to the changes between acids and alcohols exhibited by the ab initio charges. In formic acid, the CM1 dipole moments decrease with respect to those of AM1 and PM3. This is achieved by simultaneously increasing the opposing carbonyl and hydroxyl bond dipoles.

Tetrahydrofuran, furan and ethers The only parameterized atom in tetrahydrofuran, illustrated in Fig. 4, is the ring oxygen. Here, like all other oxygen atoms in

TABLE 18
DIPOLE MOMENTS, DEVIATIONS RELATIVE TO MP2/cc-pVDZ AND PARTIAL CHARGES FOR MOLECULES WITH C-N-O LINKAGES

Level	μ (D)	Deviation	Partial charge		
			C	N	O
Oxazole [C₃H₃NO]					
MP2/cc-pVDZ[a]	2.71		0.29	−0.34	−0.06
AM1[b]	1.63	−1.08	−0.16	−0.02	−0.07
CM1A	2.83	0.12	−0.03	−0.22	−0.05
CM1A//MP2/cc-pVDZ	2.84	0.13	−0.01	−0.23	−0.06
PM3[b]	1.33	−1.38	−0.07	−0.04	−0.07
CM1P	2.82	0.11	0.09	−0.27	−0.08
CM1P//MP2/cc-pVDZ	2.86	0.15	0.06	−0.27	−0.06
Nitromethane [CH₃NO₂]					
MP2/cc-pVDZ[a]	3.24		−0.29	0.73	−0.40
AM1[b]	4.13	0.89	−0.20	0.51	−0.36
CM1A	4.46	1.22	−0.22	0.65	−0.42
CM1A//MP2/cc-pVDZ	4.33	1.09	−0.20	0.59	−0.40
PM3[b]	3.39	0.15	−0.36	1.24	−0.59
CM1P	3.69	0.45	−0.28	1.02	−0.52
CM1P//MP2/cc-pVDZ	3.61	0.37	−0.27	0.99	−0.52
Formaldoxime [CH₂NOH]					
MP2/cc-pVDZ[a]	0.11		0.16	−0.29	−0.36
AM1[b]	0.58	0.47	−0.19	−0.06	−0.26
CM1A	0.57	0.46	−0.03	−0.29	−0.36
CM1A//MP2/cc-pVDZ	0.86	0.75	−0.03	−0.25	−0.40
PM3[b]	0.67	0.56	−0.11	−0.27	−0.01
CM1P	0.74	0.63	0.06	−0.40	−0.24
CM1P//MP2/cc-pVDZ	0.60	0.49	0.06	−0.24	−0.41

[a] Dipole moments are density-derived; partial charges are derived using the ChelpG algorithm.
[b] AM1 and PM3 dipole moments are calculated from AM1 and PM3 partial charges.

the CM1 mapping, the charge is made more negative. This results in an increased dipole moment with an error of only 0.1–0.2 D. Although furan and tetrahydrofuran are similar in structure, the experimental dipole moments differ by 1.0 D. In furan, as in tetrahydrofuran, only the ring oxygen is parameterized, and in both molecules the oxygen is singly bonded to two carbon atoms. However, CM1A and CM1P perform quite well on these molecules; both methods predict a 1.1 D difference between the aromatic and nonaromatic compounds. In general, ether dipoles are improved 30–50% by the CM1 charges.

Methylamine, amines and ammonium ion AM1 and PM3 Mulliken charges for amines place particularly little negative charge on nitrogen; see, e.g., methylamine in Fig. 5. The CM1 mappings correct this tendency by making nitrogen more negative and any attached hydrogen more positive. The CM1A and CM1P dipole moments for amines are much improved over their AM1 and PM3 predecessors. The CM1A dipole moment is larger by 0.2 D, as might be expected on the basis of its greater charge separation relative to CM1P. This difference derives from the more polarized methyl group from AM1 Mulliken

TABLE 19
RMS DEVIATIONS IN THE DIPOLE MOMENT (D) FOR COMPOUNDS USED IN STAGE 3 OF THE PARAMETERIZATION

Level	Stage 3		All stages (195 compounds[c])
	C-S-O compounds[a]	C-N-O compounds[b]	
AM1[d]	1.04	0.85	0.85
CM1A	0.81	0.75	0.30
CM1A//MP2/cc-pVDZ	0.56	0.77	_[e]
PM3[d]	0.78	0.86	0.92
CM1P	0.50	0.45	0.26
CM1P//MP2/cc-pVDZ	0.28	0.36	_[e]

[a] Six compounds in Table 17; deviations with respect to MP2/cc-pVDZ.
[b] Three compounds in Table 18; deviations with respect to MP2/cc-pVDZ.
[c] Combining results from stage 3 with those from stages 1 and 2; errors in stages 1 and 2 are with respect to experiment.
[d] From Mulliken charges.
[e] Not available.

TABLE 20
DIPOLE MOMENTS, DEVIATIONS RELATIVE TO MP2/cc-pVDZ AND PARTIAL CHARGES FOR NITRIC OXIDE

Level	μ (D)	Deviation	Partial charge	
			N	O
Nitric oxide [NO]				
MP2/cc-pVDZ[a]	0.12		−0.02	0.02
AM1[b]	−0.27	0.39	0.05	−0.05
CM1A	0.71	0.59	−0.13	0.13
CM1A//MP2/cc-pVDZ	0.58	0.46	−0.11	0.11
PM3[b]	−0.25	0.37	0.05	−0.05
CM1P	0.28	0.16	−0.05	0.05
CM1P//MP2/cc-pVDZ	0.24	0.12	−0.05	0.05

[a] Dipole moment calculated from the density, partial charges are ChelpG charges.
[b] AM1 and PM3 dipole moments are calculated from AM1 and PM3 Mulliken charges.

which remains unaltered by the charge model. The mapped dipole moments are larger than the AM1 and PM3 results by 0.9 and 1.1 D, respectively. Over the set of 13 amines, the rms errors were reduced from 0.73 to 0.17 and from 1.01 to 0.16 D by the CM1A and CM1P models, respectively. The CM1 dipole moments have errors about a factor of two smaller than those even from the full AM1 and PM3 wave functions.

The ammonium ion provides an excellent example of the power of the CM1 approach. Although the PM3-CM1P charge model was parameterized entirely based on data for neutral molecules, ions are also treated well. In the PM3 method, the partial charges for this ion are very nonphysical, in particular, $q_N = 1.00$ and $q_H = 0.00$. In the PM3-CM1P charge model, $q_N = -0.11$ and $q_H = 0.28$. These values are far more reasonable than those of PM3, albeit less polarized than the MP2/cc-pVTZ ChelpG
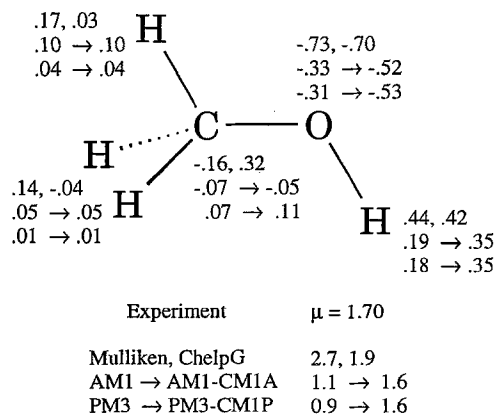

Fig. 1. Partial charges and dipole moments (D) for methanol. Six partial charges are shown next to each atom. Row 1 contains partial charges calculated by Mulliken analysis (given first) and ChelpG analysis of the HF/6-31G* wave function at the HF/6-31G*-optimized geometry. Row 2 contains partial charges calculated by Mulliken and CM1A analysis of the AM1 wave function at the AM1-optimized geometry. Row 3 contains partial charges calculated by Mulliken and CM1P analysis of the PM3 wave function at the PM3-optimized geometry. The experimental dipole moment is given below the molecule, and below that the dipole moments calculated from each of the six sets of partial charges described above are given in the same order.

charges where $q_N = -0.77$ and $q_H = 0.44$. AM1-CM1A, which starts from a more reasonable set of Mulliken charges is even more successful, producing charges of $q_N = -0.51$ and $q_H = 0.38$.

Acetonitrile and nitriles Nitrile dipole moments are poorly reproduced by AM1 and PM3 Mulliken charges, as illustrated in Fig. 6. These methods, as noted in the case of amines, place too little charge on nitrogen. The charge on nitrogen in each case is accordingly made roughly 0.3 e⁻ more negative by the mappings, at the expense of the nitrile carbon. The error for nitriles falls from 1.5 to 0.17 D in CM1A and from 1.5 to 0.18 D in CM1P, as shown in Table 9.

Formamide and amides In the amides the CM1 mapping gives results in good agreement with the ab initio ChelpG charges for nitrogen and its attached hydrogens, as illustrated for formamide in Fig. 7. However, the CM1 mapping gives a smaller polarization of the carbonyl group than does ChelpG, so the latter predicts a larger dipole moment. The CM1P mapping yields the smaller error in the dipole moment, but for both the CM1A and CM1P mappings the charges are reasonable in the context of the ChelpG results. Clearly, the PM3 charge on nitrogen requires substantial adjustment, as illustrated by the dipole moment which is more than 1 D in error before the mapping.

Fluoromethane and fluorides The CM1A mappings reduce the AM1 partial charge error on 31 organofluorides by 50%, and the CM1P mapping reduces the PM3 partial charge errors by 57%. The final errors are smaller than those calculated directly from the wave functions, so here, as in nitrogen compounds, the mapping makes up for errors in the NDDO approximation and parameterization, as well as for errors in replacing the density by distributed monopoles. Figure 8 shows fluoromethane, which is an outlier in the fluoride data set, exceeding the rms error by 0.2 D. The CM1A mapping makes fluoride less negative than the AM1 Mulliken charge, and the CM1P mapping makes fluoride more negative. Other cases, like 1,2-difluorobenzene, are much improved by the CM1A mapping, as shown in Tables 3 and 9.

Fig. 2. Partial charges and dipole moments (D) for formaldehyde (see legend to Fig. 1).

```
        -0.42, -0.48
        -0.28 → -0.33
        -0.31 → -0.40
             O
             ‖        .14, .50
             C        .14 → .19
            / \       .30 → .39
          H    H
        .14, -.01
        .07 → .07
        .01 → .01

Experiment            μ = 2.33

Mulliken, ChelpG      3.2, 2.7
AM1 → AM1-CM1A        2.0 → 2.4
PM3 → PM3-CM1P        1.8 → 2.4
```

Methylsilane and silanes The six silanes in Table 4 were parameterized by adjusting only the hydrogen atoms attached to silicon. For CM1A, these hydrogens were made less negative, while for CM1P, they were made more negative, as illustrated in Fig. 9. By virtue of the charge balance requirement, the charge on silicon is brought into general accord with the ab initio ChelpG results. CM1 dipole moments are slightly better than those from the Mulliken charges or from the wave function.

Methanethiol and sulfur compounds Sulfur is made more negative by the present CM1 mappings, and the hydrogens attached to sulfur are made more positive by the mechanism of local charge balance, as illustrated in Fig. 10. The dipole moments of nine sulfur compounds are improved by nearly an order of magnitude with CM1A and CM1P.
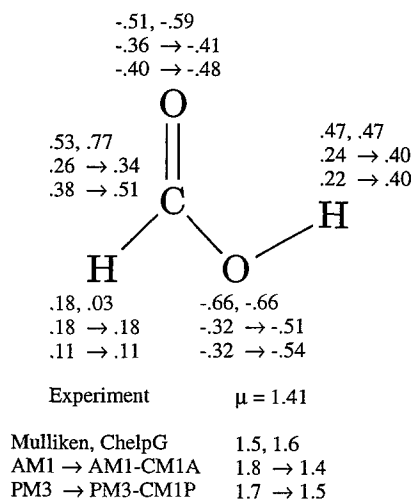
```
        -.51, -.59
        -.36 → -.41
        -.40 → -.48
             O
.53, .77     ‖         .47, .47
.26 → .34              .24 → .40
.38 → .51    C         .22 → .40
            / \         H
          H    O      
        .18, .03   -.66, -.66
        .18 → .18  -.32 → -.51
        .11 → .11  -.32 → -.54

Experiment            μ = 1.41

Mulliken, ChelpG      1.5, 1.6
AM1 → AM1-CM1A        1.8 → 1.4
PM3 → PM3-CM1P        1.7 → 1.5
```

Fig. 3. Partial charges and dipole moments (D) for formic acid (see legend to Fig. 1).
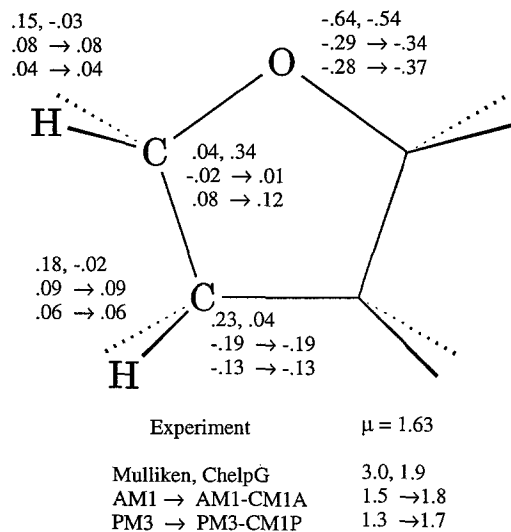
Fig. 4. Partial charges and dipole moments (D) for tetrahydrofuran (see legend to Fig. 1).

```
.15, -.03                        -.64, -.54
.08 → .08            O           -.29 → -.34
.04 → .04                        -.28 → -.37
  H     C    .04, .34
              -.02 → .01
              .08 → .12
.18, -.02
.09 → .09      C
.06 → .06      .23, .04
               -.19 → -.19
    H          -.13 → -.13

Experiment            μ = 1.63

Mulliken, ChelpG      3.0, 1.9
AM1 → AM1-CM1A        1.5 → 1.8
PM3 → PM3-CM1P        1.3 → 1.7
```

Chloromethane and halides The final three data sets comprise 22 alkyl chlorides, 10 alkyl bromides and five alkyl iodides. The C-X bond polarity is not large enough when using AM1 and PM3 partial charges, as illustrated in Fig. 11. This is evidenced by the small dipole moments in Tables 6–8. Both scale factors and offsets were used for halogen parameterization. The error reductions shown in Table 9 are dramatic, especially for the AM1 iodides where the dipole moment is improved by making carbon more positive and iodine more negative. This polarization for iodides agrees with the high-level ab initio dipole direction for methyl iodide.

For 22 chlorides, the CM1A mapping applies a slightly more negative charge to chlorine, changing the rms error from 0.44 to 0.18 D. For example, chlorine in chloromethane shifts from a −0.12 charge to −0.15. Although this is a small change in charge, the C-Cl bond of 1.7 Å is relatively long and the result is a change in dipole moment from 1.47 to 1.74 D (the experimental value is 1.89 D) and the error of 0.15 D is near the rms error over
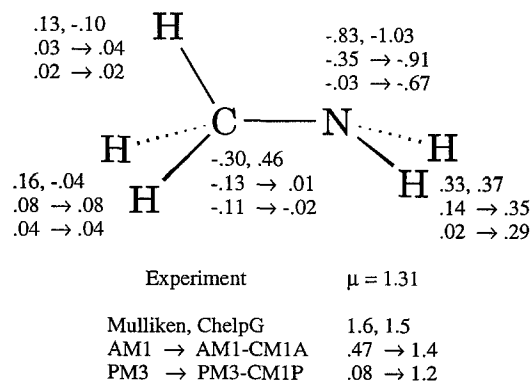
```
.13, -.10      H          -.83, -1.03
.03 → .04                 -.35 → -.91
.02 → .02                 -.03 → -.67
              C———N
   H      -.30, .46      H
.16, -.04  -.13 → .01    H  .33, .37
.08 → .08  -.11 → -.02      .14 → .35
.04 → .04   H              .02 → .29

Experiment            μ = 1.31

Mulliken, ChelpG      1.6, 1.5
AM1 → AM1-CM1A        .47 → 1.4
PM3 → PM3-CM1P        .08 → 1.2
```

Fig. 5. Partial charges and dipole moments (D) for methylamine (see legend to Fig. 1).

H
.23, .10   -.53, -.20
.11 → .11   -.14 → -.14
.06 → .06   .04 → .04
H·······/C————C≡≡≡N
H   -.45, -.51
                         -.05 → -.31
                         -.07 → -.34
                         .30, .47
                         -.14 → .11
                         -.16 → .11

Experiment   μ = 3.925

Mulliken, ChelpG   4.7, 4.0
AM1 → AM1-CM1A   2.2 → 3.8
PM3 → PM3-CM1P   2.3 → 3.9
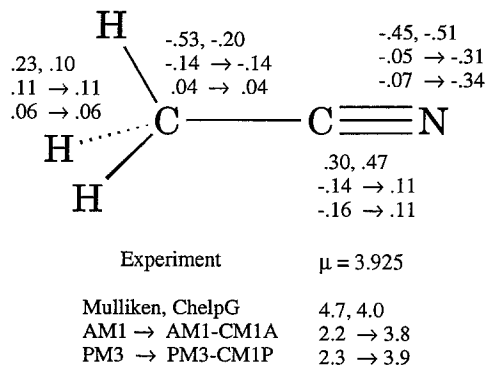
Fig. 6. Partial charges and dipole moments (D) for acetonitrile (see legend to Fig. 1).

the set of 22 compounds. The CM1P mapping also places more negative charge on chlorine. Here the rms error is dramatically reduced from 1.17 D to a final error of 0.20 D. Similar results are obtained for bromides and iodides.

Table 15 The comparison over 23 representative cases from Tables 1–6 shows that the CM1 dipole moments are on the same level of agreement with experiment as those derived using MP2/6-31G* densities calculated at HF/6-31G* geometries. (The major exception is provided by the dipole moment calculated by the CM1A method for amides; CM1P, however, is better for amides.) Dipole moments and charges, however, are considerably more costly to compute from MP2 densities than by the CM1 methods.

Ions For reasons elaborated upon earlier, the dipole moment is a dubious measure of the validity of partial charges in an ion. Therefore, for ions, we stress the importance of agreement between the high-level ab initio
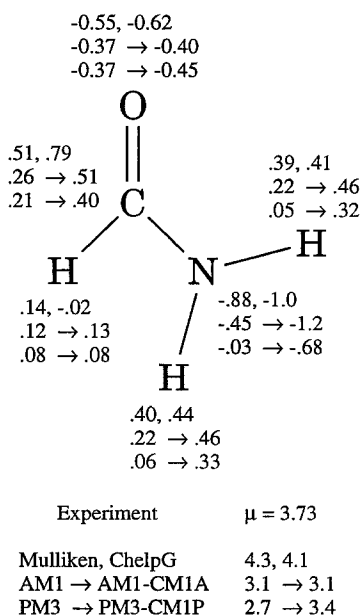

-0.55, -0.62
-0.37 → -0.40
-0.37 → -0.45
O
.51, .79   ‖
.26 → .51   .39, .41
.21 → .40   C   .22 → .46
                        .05 → .32
        /   \   ___H
    H   N
.14, -.02   -.88, -1.0
.12 → .13   -.45 → -1.2
.08 → .08   -.03 → -.68
        H
        .40, .44
        .22 → .46
        .06 → .33

Experiment   μ = 3.73

Mulliken, ChelpG   4.3, 4.1
AM1 → AM1-CM1A   3.1 → 3.1
PM3 → PM3-CM1P   2.7 → 3.4

Fig. 7. Partial charges and dipole moments (D) for formamide (see legend to Fig. 1).

H
.16, .03   -.06, .19
.07 → .07   -.04 → -.06
.03 → .03   .05 → .09
H·······/C————F
H   -.40, -.28
                         -.18 → -.16
                         -.15 → -.19

Experiment   μ = 1.86

Mulliken, ChelpG   3.4, 2.0
AM1 → AM1-CM1A   1.6 → 1.5
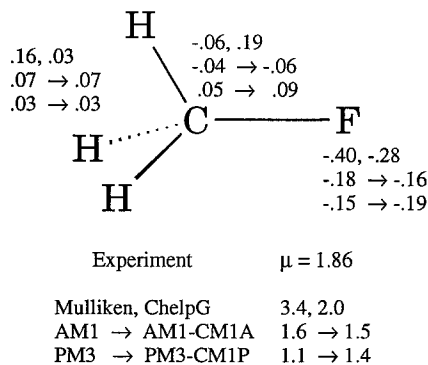PM3 → PM3-CM1P   1.1 → 1.4

Fig. 8. Partial charges and dipole moments (D) for fluoromethane (see legend to Fig. 1).

ChelpG charges and the CM1 charges. However, as shown in Tables 10–13, both CM1 models improve the dipole moments overall in three of the four categories. More importantly, it is in the agreement with the ChelpG charges that the real power of this method begins to show itself. It should be pointed out that only in the case of CM1A nitrogen and hydrogen attached to nitrogen did the ions actually play a role in determining the value of a parameter. All other atomic parameters in CM1A and all parameters in CM1P were determined using *only* neutral molecules. Nevertheless, all CM1 atomic partial charges show exceedingly good agreement with the ab initio ChelpG partial charges. With the exception of nitrogen in both CM1A and CM1P, all atomic partial charges examined show an rms error of 0.14 or less compared to MP2/cc-pVTZ ChelpG charges, often improving on the error of the AM1 or PM3 Mulliken analysis by a factor of four or five. Even for nitrogen, both of the CM1 models are much improved over AM1 or PM3. CM1P shows a three-fold improvement over PM3 which inevitably predicts large, positive partial charges on N in ammonium cations. The agreement between CM1 and the ChelpG charges is still more impressive when the time involved is considered. For some of the heavier ions, the MP2/cc-pVTZ optimizations and ChelpG calculations took several hours on
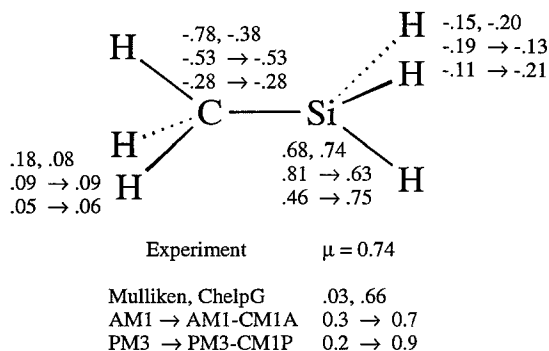

H   -.15, -.20
        -.78, -.38   -.19 → -.13
H\   -.53 → -.53   H -.11 → -.21
        -.28 → -.28 /H
         \   /
          C————Si
    H··/   \
.18, .08   .68, .74
.09 → .09   H   .81 → .63   H
.05 → .06   .46 → .75

Experiment   μ = 0.74

Mulliken, ChelpG   .03, .66
AM1 → AM1-CM1A   0.3 → 0.7
PM3 → PM3-CM1P   0.2 → 0.9

Fig. 9. Partial charges and dipole moments (D) for methylsilane (see legend to Fig. 1).

.09, .10   H
.12 → .12
.08 → .08   -.06, -.34
                             -.01 → -.20
            \   -.01 → -.18
    H·······C——S
.20, .02   / -.63, .03
.10 → .10 H -.32 → -.23
.06 → .06   -.19 → -.11   H   .20, .20
                                     .02 → .11
                                     -.01 → .07

Experiment   μ = 1.52

Mulliken, ChelpG   0.8, 1.8
AM1 → AM1-CM1A   0.4 → 1.3
PM3 → PM3-CM1P   0.5 → 1.2
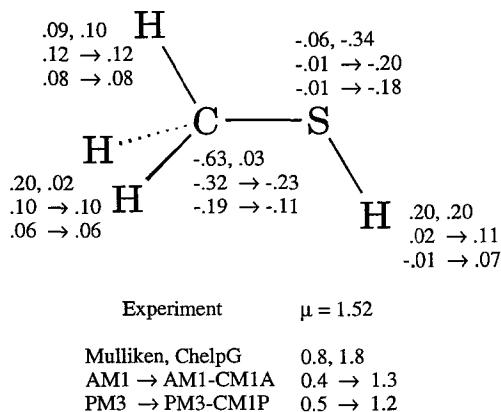
Fig. 10. Partial charges and dipole moments (D) for methanethiol (see legend to Fig. 1).

a Cray Y-MP C90 (starting from reasonable initial structures). For instance, an MP2/cc-pVTZ optimization for trimethylammonium, starting from an AM1-optimized geometry, took 10 h of CPU time. In comparison, the CM1A calculation starting from a highly perturbed geometry took 1 s – more than four orders of magnitude faster than the ab initio calculation. Thus, when semiempirical geometries are reasonably accurate, it should be possible to use the CM1 models to obtain MP2/cc-pVTZ-quality partial charges on systems that are prohibitively large for even simple ab initio calculations.

Table 14 provides a statistical summary for ions. This table illustrates the overall success of the CM1 models in removing both systematic and random deviations of the semiempirical charges from the most accurate available ab initio charges.

Discussion of results on molecules in the secondary database

Tables 17 and 18 give results for the nine neutral molecules in the secondary database. In addition, results for NO (nitric oxide) are reported in Table 20, although it was not used in the parameterization. Table 17 contains six molecules with C-S-O linkages and Table 18 contains three molecules with C-N-O linkages.

We first consider the CM1P results in Table 17. In five out of six cases, the CM1P dipole moment is more accurate than that computed from PM3 partial charges, and in the sixth case these dipole moments differ by only 0.04 D. This shows that in adjusting O in compounds containing H, C and O, and S in compounds containing H, C and S, thereby charge balancing O and S against C, we have also charge-balanced them against one another. Furthermore, in 20 out of 21 cases, the CM1P tabulated partial charges for the constituent atoms of the six molecules are closer than the PM3 ones to the MP2/cc-pVDZ ChelpG charges – an amazing success rate! Without an additional parameter (d_{OS}), the CM1A model did not fare

so well. With d_{OS}, the dipole moment is improved in four out of six cases, and the partial charges were brought closer to the MP2/cc-pVDZ ChelpG ones in 20 out of 21 cases. Table 14 shows that the rms improvement in the dipole moment is 0.3 D for CM1P vs. PM3 and 0.2 D for CM1A vs. AM1.

Table 18 shows analogous comparisons for compounds with C-N-O linkages. These molecules are notoriously difficult for the NDDO level of theory [19,20]; the mappings (with *no* new parameters for CM1P) improve the dipole moments by about 0.4 D for CM1P and 0.1 D for CM1A. Without the additional N-O parameter in CM1A, nitrogens bonded to oxygen were made too positive by the N scale factor. In nitromethane, for example, the scale factor increased the charge from the AM1 value of 0.51 to 1.34, which was too positive compared to the ChelpG charge of 0.73. The N-O offset on the nitrogen brings the charge down to 0.65, in good agreement with the ChelpG charge. Although the dipole moment improvement in CM1A is small in the three neutral molecules listed in Table 18, in eight out of nine cases the partial charges obtained with CM1A are improved over those from AM1 when compared to ChelpG partial charges.

An important consistency check here is to apply the mappings at the ab initio geometries. It can be seen from Tables 17 and 18 that when CM1A and CM1P are applied to the more accurate geometries supplied by the MP2/cc-pVDZ calculations, the dipole moments and partial charges remain about the same as those calculated by CM1A and CM1P using the AM1 and PM3 geometries. This is encouraging insofar as it attests to the physical character of the mapping. If the mapping was unphysical or cancelled geometry errors against charge balance, this trend would not be found. Table 19 summarizes the results for both stages of the process.

Table 20 gives CM1A and CM1P results for nitric oxide at semiempirical and ab initio geometries. Nitric oxide (NO) was not used in the parameterization. However, it is an interesting case for several reasons. First of

        H   -.54, -.09
.22, .10   -.18 → -.15
.10 → .10   -.09 → .02
.05 → .05   \
    H·······C————Cl
         /   -.11, -.21
         H   -.12 → -.15
                          -.06 → -.18

Experiment   μ = 1.89

Mulliken, ChelpG   2.0, 2.3
AM1 → AM1-CM1A   1.5 → 1.7
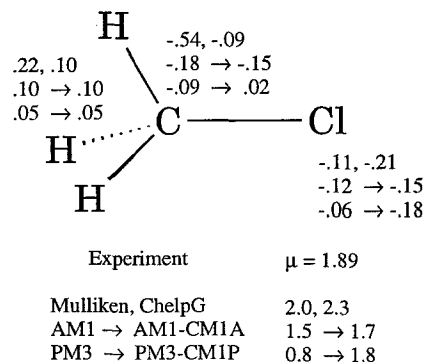PM3 → PM3-CM1P   0.8 → 1.8

Fig. 11. Partial charges and dipole moments (D) for chloromethane (see legend to Fig. 1).

all, it is an open-shell molecule, and there are no open-shell molecules in the database. Second, both AM1 and PM3 predict the direction of the dipole moment reversed relative to the MP2/cc-pVDZ dipole moment. As shown in Table 20, both CM1 methods treat this molecule well. CM1P gives the correct direction on the dipole and improves the errors in the partial charges in spite of the fact that there is no special N-O parameter in CM1P. CM1A does not do quite as well, giving errors that are larger in magnitude for both the dipole moment and the partial charges. However, CM1A does predict the direction of the dipole correctly, and the magnitude of the errors remains small. Again, we see the same trend of slightly improved results when going to ab initio geometries.

Conclusions

In previous developments of semiempirical molecular orbital theory, only the Hamiltonian was parameterized. Wave functions resulting from such Hamiltonians were then used or interpreted by the same techniques as are typically applied to ab initio wave functions. In the present work, we introduce a new kind of semiempiricism, in which semiempirical parameters are introduced in the interpretative step. In future work, such semiempirical parameters could also be developed for interpreting ab initio wave functions, e.g., one could introduce a parameter set for calculating effective point charges from HF/3-21G* wave functions. Charges derived from density functional calculations would also be an option. Partial charges obtained from mappings containing semiempirical parameters are defined to be class IV charges if the parameters are adjusted using physical observables obtained from experiment or well-converged calculations.

One very important point about the philosophy of class IV charge models should be emphasized here. In class IV charge models one can obtain the best possible charges from a given level of wave function because the mapping ensures that physical observables calculated directly from the improved partial charges are as accurate as possible. We contrast this with class III methods, where partial charges are obtained under the criterion that they reproduce the values of the physical observable, typically electrostatic potentials, computed from the wave function; not the criterion that they reproduce experimental physical observables. At suitably high levels of theory the quantitative differences between the two may become quite small, but the application of such high levels of theory may not be practical in many instances. We summarize the new philosophy as follows: if you want *useful* partial charges, optimize the semiempirical mapping so that properties computed directly from the partial charges are as accurate as possible.

We have presented two class IV charge models as examples of the approach. These models are called CM1 mappings, in particular CM1A, based on AM1 wave functions, and CM1P, based on PM3 wave functions. Overall, the rms errors in Tables 9, 14 and 19 show that the CM1 mappings provide general improvement in charge balance across a wide variety of functional groups. The charges generated by the CM1A and CM1P models have two major advantages. First, they make up for errors intrinsic to replacing a continuous charge distribution by a set of distributed point charges because the semiempirical mapping from which they are obtained is chosen to minimize errors in the physical observables predicted from the point charges. This advantage is shared with class III charge models. Second, they make up for deficiencies in the wave function from which they were obtained, because the parameterizations are chosen to minimize deviations from experiment. This advantage is unique to class IV charge models.

Here, the charge models were mapped from the popular AM1 and PM3 wave functions which tend, for example, to underestimate the charge magnitudes at N, O and hydrogens attached to N, O and Si, and the negative charges at S, Cl, Br and I. The mapping does an excellent job of correcting these deficiencies. In typical cases, hydrogens attached to N or O are mapped to be between 0.1 and 0.3 charge units more positive than the AM1 or PM3 Mulliken charges. The final partial charges yield more accurate dipole moments than even the far more expensive HF/6-31G* ab initio ChelpG or MK charges, which are widely used for simulations [32]. Further, the CM1 dipole moments are in about as close agreement with experiment as are those from MP2 densities for 23 representative molecules. Finally, partial charges calculated with CM1 agree quite well with MP2/cc-pVTZ ChelpG partial charges, in spite of the fact that the latter charges take three to four orders of magnitude longer to calculate.

The development of the CM1 models employed a data set of 195 neutral molecules and 25 ions. The CM1P method achieves better accuracy than the CM1A method, although it has fewer parameters (15 in CM1P vs. 19 in CM1A). In CM1P, 12 parameters depend only on atom type, and three are specific to atom pairs with non-zero bond order, whereas in CM1A, 12 parameters depend on atom type, and there are seven parameters for specific atom pairs with non-zero bond orders.

We believe that the method presented here for predicting partial charges will be useful for a wide variety of applications in which the continuous charge distribution of a molecule is modeled by a set of partial charges centered at the nuclei. Two types of applications that immediately come to mind are force fields and solvation [32,33].

Acknowledgements

References

1 Straatsma, T.P. and McCammon, J.A., Annu. Rev. Phys. Chem., 43 (1992) 407.
2 For recent comparisons of different methods of charge analysis see:
a. Bachrach, S.M., In Lipkowitz, K.B. and Boyd, D.B. (Eds.) Reviews in Computational Chemistry, Vol. 5, VCH Publishers, New York, NY, 1993, pp. 171–227.
b. Wiberg, K.B. and Rablen, P.R., J. Comput. Chem., 14 (1993) 1504.
3 Coppens, P., Annu. Rev. Phys. Chem., 43 (1992) 663.
4 a. Mulliken, R.S., J. Chem. Phys., 3 (1935) 564.
b. Mulliken, R.S., J. Chem. Phys., 23 (1955) 1833.
c. Mulliken, R.S., J. Chem. Phys., 36 (1962) 3428.
5 a. Bader, R.W.F., Acc. Chem. Res., 18 (1985) 9.
b. Bader, R.W.F., Atoms in Molecules. A Quantum Theory, Clarendon Press, Oxford, 1990.
6 Warshel, A., Acc. Chem. Res., 14 (1981) 284.
7 Price, S.L. and Stone, A.J., J. Chem. Phys., 86 (1987) 2859.
8 Shi, X. and Bartell, L.S., J. Am. Chem. Soc., 92 (1988) 5667.
9 Hall, D. and Williams, D.E., Acta Crystallogr., A31 (1975) 56.
10 Momany, F.A., J. Phys. Chem., 82 (1978) 592.
11 a. Kollman, P.A., J. Am. Chem. Soc., 99 (1977) 4875.
b. Kollman, P.A., J. Am. Chem. Soc., 100 (1978) 2974.
12 Smit, P.H., Derissen, J.L. and Van Duijneveldt, F.B., Mol. Phys., 37 (1979) 521.
13 Cox, S.R. and Williams, D.E., J. Comput. Chem., 2 (1981) 304.
14 a. Williams, D.E. and Yan, M.J., Adv. Atomic Mol. Phys., 23 (1988) 87.
b. Williams, D.E., J. Comput. Chem., 9 (1988) 745.
c. Williams, D.E., Biopolymers, 29 (1990) 1367.
d. Williams, D.E., In Lipkowitz, K.B. and Boyd, D.B. (Eds.) Reviews in Computational Chemistry, Vol. 2, VCH Publishers, New York, NY, 1991, pp. 219–271.
15 a. Chirlian, L.E. and Francl, M.M., J. Comput. Chem., 8 (1987) 894.
b. Breneman, C.M. and Wiberg, K.B., J. Comput. Chem., 11 (1990) 361.
16 a. Singh, U.C. and Kollman, P.A., J. Comput. Chem., 5 (1984) 129.
b. Besler, B.H., Merz Jr., K.M. and Kollman, P.A., J. Comput. Chem., 11 (1990) 431.
c. Merz, K.M., J. Comput. Chem., 11 (1992) 749.
17 a. Reed, A.E., Weinstock, R.B. and Weinhold, F., J. Chem. Phys., 83 (1985) 735.
b. Reed, A.E., Weinhold, F. and Curtiss, L.A., Chem. Rev., 88 (1988) 899.
18 Montagnini, R. and Tomasi, J., J. Mol. Struct. (THEOCHEM), 279 (1993) 131.
19 Dewar, M.J.S., Zoebisch, E.G., Healy, E.F. and Stewart, J.J.P., J. Am. Chem. Soc., 107 (1985) 3902.
20 a. Stewart, J.J.P., J. Comput. Chem., 10 (1989) 209.
b. Stewart, J.J.P., J. Comput. Chem., 10 (1989) 221.
21 a. Møller, C. and Plesset, M.S., Phys. Rev., 46 (1934) 618.
b. Pople, J.A., Seeger, R. and Krishnan, R., Int. J. Quantum Chem. Symp., 11 (1977) 49.
c. Krishnan, R. and Pople, J.A., Int. J. Quantum Chem., 14 (1978) 91.
d. Krishnan, R., J. Chem. Phys., 72 (1980) 4244.
e. Hehre, W.J., Radom, L., Schleyer, P.v.R. and Pople, J.A., Ab Initio Molecular Orbital Theory, Wiley, New York, NY, 1986.
22 Hehre, W.J., Ditchfield, R. and Pople, J.A., J. Chem. Phys., 56 (1972) 2257.
23 a. Dunning Jr., T.H., J. Chem. Phys., 90 (1989) 1007.
b. Woon, D.E. and Dunning Jr., T.H., J. Chem. Phys., 98 (1993) 1358.
24 a. X/Y denotes electronic structure level X with basis set Y. HF denotes Hartree–Fock, MP2 denotes second-order Møller–Plesset perturbation theory (Ref. 21a), and cc-pVDZ (Ref. 23) denotes a basis set.
b. X/Y//Z/W denotes that the wave function and energy are calculated by method X with basis set Y at a geometry optimized by method Z with basis set W. X/Y denotes X/Y//X/Y.
25 a. Pople, J.A. and Segal, G.A., J. Chem. Phys., 43 (1965) S129.
b. Dewar, M.J.S. and Thiel, W., J. Am. Chem. Soc., 99 (1977) 4899.
26 Armstrong, D.R., Perkins, P.G. and Stewart, J.J.P., J. Chem. Soc., Dalton Trans., (1973) 838.
27 a. Stark, B., In Hellwege, K.-H. and Hellwege, A.M. (Eds.) Molecular Constants from Microwave Spectroscopy, Landolt-Börnstein, New Series, Group II, Vol. 4, Springer-Verlag, Berlin, 1967, pp. 136–151.
b. Demaison, J., Hütner, W., Stark, B., Buck, I., Tischer, R. and Winnewisser, M., In Hellwege, K.-H. (Ed.) Molecular Constants, Landolt-Börnstein, New Series, Group II, Vol. 6, Springer-Verlag, Berlin, 1974, pp. 261–304.
c. Demaison, J., Hütner, W. and Tiemann, E., In Hellwege, K.-H. and Hellwege, A.M. (Eds.) Molecular Constants, Landolt-Börnstein, New Series, Group II, Vol. 14a, Springer-Verlag, Berlin, 1982, pp. 584–643.
d. Nelson, R.D., Lide, D.R. and Maryott, A.A., Natl. Stand., Ref. Data Ser., United States National Bureau of Standards, NSRDS-NBS 10, 1967.
28 a. Hocking, W.H., Z. Naturforsch., 31A (1976) 1113.
b. Caminati, W., J. Mol. Spectrosc., 86 (1981) 193.
c. Caminati, W. and Corbelli, G., J. Mol. Spectrosc., 90 (1981) 572.
29 Frisch, M.J., Trucks, G.W., Head-Gordon, M., Gill, P.M.W., Wong, M.W., Foresman, J.B., Johnson, B.G., Schlegel, H.B., Robb, M.A., Repolgle, E.S., Gomperts, R., Andres, J.L., Raghavachari, K., Binkley, J.S., Stewart, J.J.P. and Pople, J.A., GAUSSIAN92, Gaussian, Inc., Pittsburgh, PA, 1992.
30 Cramer, C.J., Lynch, G.C., Hawkins, G.D. and Truhlar, D.G., QCPE Bull., 13 (1993) 78. This new code will be made available as AMSOL, version 4.5.
31 a. Marquardt, D.W., J. Soc. Indian Appl. Math., 11 (1963) 431.
b. Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T., Numerical Recipes, Cambridge University Press, Cambridge, 1989.
32 a. Jorgensen, W.L., Chandresekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L., J. Chem. Phys., 79 (1983) 926.
b. Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta Jr., S. and Weiner, P., J. Am. Chem. Soc., 106 (1984) 765.
c. Weiner, S.J., Kollman, P.A., Nguyen, D.T. and Case, D.A., J. Comput. Chem., 7 (1986) 230.
d. Jorgensen, W.L. and Tirado-Rives, J., J. Am. Chem. Soc., 110 (1988) 1657.
33 a. Cramer, C.J. and Truhlar, D.G., J. Comput.-Aided Mol. Design, 6 (1992) 629.
b. Cramer, C.J. and Truhlar, D.G., In Lipkowitz, K.B. and Boyd, D.B. (Eds.) Reviews in Computational Chemistry, Vol. 6, VCH Publishers, New York, NY, in press.

Appendix

Selected geometrical data and dipole moments for nine compounds in the secondary database

Level	Bond length (Å)		Angle (°)		Dipole moment (D)
Me₂SO₂	S=O		O=S=O		
MP2/cc-pVDZ	1.48		124.6		4.17
HF/cc-pVDZ	1.45		122.6		4.95
HF/6-31G*	1.43		120.1		5.10
AM1	1.40		117.7		1.97
PM3	1.47		118.2		3.88
Me₂SO	S=O		C-S=O		
MP2/cc-pVDZ	1.53		103.6		3.89
HF/cc-pVDZ	1.51		103.6		4.57
HF/6-31G*	1.49		106.7		4.47
AM1	1.49		105.7		3.12
PM3	1.56		104.4		4.85
MeSO₂NH₂	S=O	S-N	O=S=O	N-S=O	
MP2/cc-pVDZ	1.47	1.69	126.3	106.6	2.87
HF/cc-pVDZ	1.44	1.65	124.1	106.9	3.68
HF/6-31G*	1.43	1.65	121.2	107.5	3.89
AM1	1.41	1.60	117.8	110.0	3.01
PM3	1.46	1.75	119.7	107.2	3.73
MeSO₃H	S=O	S-O	O=S=O		
MP2/cc-pVDZ	1.47	1.67	124.0		2.33
HF/cc-pVDZ	1.43	1.61	121.9		3.05
HF/6-31G*	1.42	1.59	119.3		3.24
AM1	1.38	1.66	120.1		1.77
PM3	1.43	1.69	119.8		2.71
MeS-OMe	S-O		C-S-O		
MP2/cc-pVDZ	1.72		93.3		0.28
HF/cc-pVDZ	1.68		95.1		0.32
HF/6-31G*	1.67		96.0		0.15
AM1	1.73		98.7		1.28
PM3	1.73		98.5		0.81
MeS(O)OMe	S=O	S-O	O-S=O		
MP2/cc-pVDZ	1.51	1.70	109.6		2.46
HF/cc-pVDZ	1.48	1.63	108.6		3.10
HF/6-31G*	1.46	1.62	109.1		3.27
AM1	1.47	1.71	100.3		3.72
PM3	1.53	1.72	104.0		3.66
Oxazole	N-O	C-N	C-N-O		
MP2/cc-pVDZ	1.38	1.33	105.5		2.71
HF/cc-pVDZ	1.36	1.28	106.3		3.13
HF/6-31G*	1.36	1.28	106.1		3.18
AM1	1.32	1.34	109.0		2.97
PM3	1.43	1.32	109.0		1.33
MeNO₂	N-O	C-N	O-N=O		
MP2/cc-pVDZ	1.23	1.49	126.2		3.24
HF/cc-pVDZ	1.19	1.50	125.9		3.99
HF/6-31G*	1.19	1.48	125.8		4.02
AM1	1.20	1.50	122.5		4.26
PM3	1.21	1.51	122.3		3.39
CH₂NOH	N-O	O-H	C-N-O	N-O-H	
MP2/cc-pVDZ	1.40	0.97	110.3	101.5	0.11
HF/cc-pVDZ	1.36	0.95	112.1	104.5	0.37
HF/6-31G*	1.37	0.95	112.0	104.5	0.38
AM1	1.32	0.98	116.6	104.8	0.58
PM3	1.40	0.95	117.0	117.0	0.67