ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/10689270

Parameterization of charge model 3 for AM1, PM3, BLYP, and B3LYP

ARTICLE *in* JOURNAL OF COMPUTATIONAL CHEMISTRY · AUGUST 2003

Impact Factor: 3.59 · DOI: 10.1002/jcc.10244 · Source: PubMed

CITATIONS
71

READS
63

3 AUTHORS, INCLUDING:

Christopher J Cramer
University of Minnesota Twin Cities
531 PUBLICATIONS **23,395** CITATIONS
SEE PROFILE

Donald Truhlar
University of Minnesota Twin Cities
1,342 PUBLICATIONS **81,184** CITATIONS
SEE PROFILE

Available from: Christopher J Cramer
Retrieved on: 23 January 2016

Parameterization of Charge Model 3 for AM1, PM3, BLYP, and B3LYP

JASON D. THOMPSON, CHRISTOPHER J. CRAMER, DONALD G. TRUHLAR
Department of Chemistry and Supercomputer Institute, University of Minnesota, Minneapolis, Minnesota 55455-0431

Received 23 August 2002; Accepted 25 November 2002

Abstract: We have recently developed a new Class IV charge model for calculating partial atomic charges in molecules. The new model, called Charge Model 3 (CM3), was parameterized for calculations on molecules containing H, Li, C, N, O, F, Si, S, P, Cl, and Br by Hartree–Fock theory and by hybrid density functional theory (DFT) based on the modified Perdew–Wang density functional with several basis sets. In the present article we extend CM3 to semiempirical molecular orbital theory, in particular Austin Model 1 (AM1) and Parameterized Model 3 (PM3), and to the popular BLYP and B3LYP DFT and hybrid DFT methods, respectively. For the BLYP extension, we consider the 6-31G(d) basis set, and for the B3LYP extension, we consider three basis sets: 6-31G(d), 6-31+G(d), and MIDI!6D. We begin with the previous CM3 strategy, which involves 34 parameters for 30 pairs of elements. We then refine the model to improve the charges in compounds that contain N and O. This modification, involving two new parameters, leads to improved dipole moments for amides, bifunctional H, C, N, O compounds, aldehydes, ketones, esters, and carboxylic acids; the improvement for compounds not containing N results from obtaining more physical parameters for carbonyl groups when the O=C—N conjugation of amides is addressed in the parameterization. In addition, for the PM3 method, we added an additional parameter to improve dipole moments of compounds that contain bonds between C and N. This additional parameter leads to improved accuracy in the dipole moments of aromatic nitrogen heterocycles with five-membered rings.

© 2003 Wiley Periodicals, Inc. J Comput Chem 24: 1291–1304, 2003

Key words: partial atomic charge; dipole moment; semiempirical theory; population analysis

Introduction

Class IV charges, as implemented by Charge Model 1[1] (CM1) and Charge Model 2[2,3] (CM2), have been very successful in predicting useful charge distributions at low cost. Class IV charges are obtained by empirically mapping charges calculated directly from wave functions to new charges that better reproduce charge-dependent observables, like the dipole moment. We have recently introduced Charge Model 3[4] (CM3), which has several advantages over the previous two models:

1. It is based on a much larger training set. Excluding iodine-containing compounds, which are not considered in CM3, the training set is expanded from 204 polar-molecule data[2,3] to 398 polar-molecule data.[5] (We also use the nonpolar molecules ethylene and benzene.) The training set is especially improved for amides, nitrogen-containing heterocycles, and compounds containing C—F, N—O, S—O, H—Si, O—Si, F—Si, H—P, N—S, C—Cl, and C—Br bonds.
2. For basis sets containing diffuse functions, it is designed to be less sensitive to basis-set-dependent variations in bond orders.
3. It is based on redistributed Löwdin population[5] analysis, which is more suitable than Mulliken[6–8] or standard Löwdin[9–12] population analysis for calculations with diffuse basis functions on first row atoms, but which is identical to standard Löwdin population analysis for nonaugmented basis sets.

The CM3 model presented in this work has been further refined to improve charges on N and O when N and O are in the same compound (e.g., for amides, nitro groups, and bifunctional H, C, N, O compounds). For the PM3 method, the CM3 charge model has also been further refined to obtain more accurate charges for compounds containing bonds between C and N.

A key advantage of class IV charges is that one can typically obtain accurate partial charges from economical wave functions,

Correspondence to: C. J. Cramer; e-mail: cramer@chem.umn.edu or D. G. Truhlar; e-mail: truhlar@chem.umn.edu

Contract/grant sponsor: U.S. Department of Defense through the Multidisciplinary Research Program of the University Research Initiative

© 2003 Wiley Periodicals, Inc.

Table 1. Optimized CM3 Parameters of the Methods Used in the Article.

			BLYP	B3LYP		
	AM1	PM3	6-31G(d)	MIDI!6D	6-31G(d)	6-31+G(d)
$C_{ZZ'}$						
C—N		−0.017				
C—O	0.029	0.010	0.052	0.077	0.054	0.031
N—O[a]	0.439	0.274	0.468	0.470	0.503	0.828
O—Si	0.201	−0.044	−0.083	−0.068	−0.086	−0.111
O—P	0.004	−0.076	−0.083	−0.035	−0.082	−0.211
P—S	0.341	0.245	0.020	−0.097	0.015	−0.023
$D_{ZZ'}$						
H—C	−0.009	0.021	−0.045	0.013	−0.051	−0.064
H—N	0.200	0.243	0.048	0.087	0.041	0.082
H—O	0.154	0.153	−0.027	−0.007	−0.035	0.003
H—Si	0.231	0.191	0.004	0.016	0.011	−0.009
H—P	0.078	0.233	0.058	−0.067	0.063	0.052
H—S	0.072	0.118	−0.004	−0.114	−0.003	−0.041
Li—C	[b]	0.192	0.490	−0.009	0.486	0.369
Li—N	[b]	0.403	0.679	0.172	0.684	0.486
Li—O	[b]	0.390	0.673	0.021	0.691	0.481
Li—S	[b]	0.268	0.538	0.016	0.538	0.443
Li—F	[b]	0.430	0.591	0.275	0.599	0.524
Li—Cl	[b]	0.117	0.570	−0.060	0.580	0.433
C—N	0.091	0.126	0.083	0.017	0.079	0.080
C—O	−0.012	0.018	−0.027	−0.147	−0.042	−0.025
C—F	0.006	0.019	0.036	−0.052	0.013	−0.050
C—Si	0.163	0.069	−0.061	0.043	−0.053	0.015
C—P	0.106	0.197	0.111	−0.054	0.115	0.175
C—S	0.103	0.076	0.124	−0.022	0.123	0.162
C—Cl	0.013	0.097	0.086	−0.024	0.082	0.109
C—Br	0.073	0.028	0.051	0.026	0.046	0.104
N—O	−0.132	−0.085	−0.005	0.001	−0.022	−0.052
N—P	−0.067	0.018	−0.017	−0.082	−0.012	0.057
O—Si	0.169	0.141	0.156	0.070	0.178	0.312
O—P	0.033	0.231	0.210	0.035	0.225	0.541
O—S	0.103	0.125	0.094	0.009	0.116	0.188
F—Si	0.228	0.099	0.070	−0.031	0.103	0.219
F—P	0.110	0.175	0.123	−0.013	0.151	0.387
Si—Cl	−0.171	−0.180	−0.018	−0.040	−0.031	−0.037
P—S	−0.685	−0.534	−0.054	0.192	−0.051	0.044
P—Cl	−0.252	−0.394	−0.069	0.069	−0.078	−0.099
B_0						
N—O	0.482	0.460	0.113	0.131	0.111	0.088

[a]See eq. (4) for an explanation of this parameter.
[b]In this work, CM3 is not defined for Li for the AM1 method.

even from semiempirical molecular orbital theory.[1,3] Therefore, in the present article we present CM3 parameters for the two most popular semiempirical molecular orbital methods, namely Austin Model 1[13–17] (AM1) and Parameterized Model 3[18] (PM3). Note that lithium parameters for AM1 do not exist. In many electronic structure programs, however, AM1 calculations on systems containing lithium use MNDO parameters for lithium[19] and AM1 parameters for the other elements. Such a method has not been extensively tested here, and furthermore, it has been shown[20] that higher accuracy for quantities like heats of formation and, to a lesser extent, geometries and dipole moments can be achieved by using PM3 rather than using MNDO for lithium and AM1 for other elements. Therefore, in this work we choose not to define a CM3 charge model for lithium for the AM1 method where the MNDO parameters are used for lithium.

We also extend CM3 to the most widely used hybrid density functional theory (DFT) method, namely the combination of a Becke-exchange[21] 3-parameter[22] method with a Lee–Yang–Parr[23] correlation functional in the popular B3LYP[24] formulation, and also to a pure DFT method, namely the Becke-exchange[21] functional with a Lee–Yang–Parr correlation functional,[23] i.e., BLYP. We develop the CM3 model for the B3LYP method with three

Table 2. Root-Mean-Square (RMS) Errors (in Debyes) of CM3 Dipole Moments Calculated by the Methods Used in this Article over Various Subsets of the CM3 Training Set.

				BLYP	B3LYP		
Type of compound	No.[a]	AM1	PM3	6-31G(d)	MIDI!6D	6-31G(d)	6-31+G(d)
Inorganic compounds[b]	10	0.46	0.42	0.32	0.27	0.32	0.38
Alcohols, phenol	13	0.19	0.19	0.15	0.15	0.15	0.26
Ethers	11	0.22	0.20	0.18	0.22	0.15	0.24
Aldehydes	5	0.14	0.13	0.19	0.16	0.16	0.28
Ketones	11	0.22	0.20	0.26	0.33	0.25	0.20
Carboxylic acids	9	0.20	0.22	0.16	0.24	0.18	0.34
Esters	6	0.16	0.15	0.19	0.09	0.19	0.25
Other C, H, O compounds	12	0.28	0.27	0.30	0.30	0.28	0.25
Aliphatic amines, aniline	13	0.24	0.11	0.24	0.35	0.24	0.30
5-Membered aromatic N heterocycles	3	0.59	0.32	0.43	0.45	0.40	0.59
6-Membered aromatic N heterocycles	8	0.33	0.27	0.19	0.17	0.19	0.26
Nitriles	12	0.15	0.15	0.23	0.19	0.23	0.29
Imines	6	0.34	0.35	0.40	0.51	0.40	0.44
Other C, H, N compounds	14	0.27	0.35	0.19	0.21	0.17	0.28
Amides and phenylurea	17	0.12	0.18	0.08	0.11	0.07	0.33
Nitrohydrocarbons	5	0.35	0.24	0.16	0.19	0.15	0.28
Bifunctional H, C, N, O compounds	12	0.37	0.22	0.23	0.29	0.21	0.23
H, C, N, O polar compounds	163	0.27	0.24	0.23	0.26	0.22	0.29
Fluorine-containing compounds	39	0.31	0.34	0.22	0.23	0.21	0.40
Chlorine-containing compounds	33	0.14	0.12	0.15	0.15	0.14	0.29
Bromine-containing compounds	14	0.16	0.14	0.17	0.16	0.16	0.33
Halogenated bifunctional compounds	23	0.29	0.25	0.24	0.30	0.22	0.27
Thiols	8	0.23	0.13	0.12	0.10	0.12	0.22
Sulfides, disulfides	9	0.27	0.20	0.19	0.12	0.21	0.30
Other sulfur-containing compounds	23	0.75	0.50	0.50	0.49	0.51	0.60
Phosphorus	10	0.31	0.30	0.38	0.14	0.41	0.38
Multifunctional phosphorus	13	0.89	0.71	0.48	0.16	0.48	0.30
Compounds with S and P	7	0.77	0.45	0.17	0.08	0.18	0.30
C, H, Si	9	0.28	0.33	0.08	0.08	0.08	0.20
C, H, O, Si	9	0.48	0.40	0.26	0.26	0.25	0.35
C, H, Si, halogen	18	0.59	0.49	0.16	0.19	0.15	0.62
All polar compounds excluding lithium	382	0.39	0.41	0.26	0.25	0.25	0.36
Lithium compounds	16	[c]	0.32	0.33	0.60	0.30	0.39
All polar compounds	398	[c]	0.32	0.26	0.27	0.25	0.36

[a]Number of data in the training set for this row.
[b]Six of the 10 inorganic compounds contain at most H, C, N, O.
[c]AM1 is not parameterized for Li.

basis sets, namely MIDI!6D,[25–27] 6-31G(d),[28–30] and 6-31+G(d).[28–31] For the BLYP method, we develop a CM3 model with the 6-31G(d) basis set.

We critically evaluate the CM3 charge model parameterization for AM1 and PM3 by comparing root-mean-square (RMS) errors of the CM3 dipole moments obtained by AM1 and PM3 over various subsets of a training set of 398 polar data to RMS errors obtained by CM2, Löwdin, and density dipole moments. (Note that we define a density dipole moment as a dipole moment obtained by integrating the electron density and dipole moment operator over all space.) We critically evaluate the quality of the CM3 model for BLYP and B3LYP by making comparisons of CM3 dipoles calculated by BLYP and B3LYP to CM3 dipole moments calculated by mPWPW91[32,33] and mPW1PW91.[32,33]

Charge Model 3 (CM3) Background

In our previous development of the CM2 and CM3 models,[2-4] the charge, q_k, on an atom k was defined to be

$$q_k = q_k^0 + \sum_{k' \neq k} B_{kk'}(D_{Z_kZ_{k'}} + C_{Z_kZ_k}B_{kk'}) \quad (1)$$

where the summation goes over all atoms in the molecule, q_k^0 is the charge on atom k obtained from either a Löwdin population analysis (LPA) or a redistributed Löwdin population analysis (RLPA) of the wave function, $B_{kk'}$ is the Mayer bond order[34–36] between atoms k and k', Z_k is the atomic number of atom k, and $C_{Z_kZ_{k'}}$ and $D_{Z_kZ_{k'}}$ are pairwise atomic parameters. For the AM1 and

Table 3. Root-Mean-Square (RMS) Errors (in Debyes) of Various Types of Dipole Moments Calculated by AM1 over Subsets of the Reduced CM2 Training Set Obtained by Removing Iodine-Containing Compounds.

Type of compound	No.[a]	Density	PA	CM2	CM3
Alcohols and water	12	0.10	0.52	0.14	0.13
Esters and lactones	8	0.33	0.52	0.17	0.19
Aldehydes and ketones	15	0.34	0.59	0.22	0.21
Carboxylic acids	9	0.14	0.48	0.30	0.20
Ethers	9	0.16	0.35	0.20	0.24
Amines and ammonia	13	0.42	0.68	0.22	0.25
Nitriles and HCN	17	0.96	1.59	0.26	0.17
Imines and N aromatics	7	0.29	1.05	0.31	0.32
Amides	3	0.45	0.92	0.18	0.12
Multifunctional N	7	0.94	1.51	0.28	0.32
H, C, N, O compounds	100	0.53	0.94	0.23	0.22
Fluorides	31	0.25	0.48	0.27	0.26
Chlorides	22	0.25	0.37	0.19	0.17
Bromides	10	0.25	0.65	0.17	0.16
Silicon	6	0.44	0.37	0.20	0.14
Sulfur	9	0.62	0.74	0.41	0.54
Phosphorus	13	0.78	1.08	0.62	0.84
Primary training set	191	0.49	0.81	0.29	0.32
Secondary training set	13	0.59	1.15	0.90	0.86
Primary and secondary sets	204	0.49	0.83	0.36	0.38

[a]Number of data in the training set for this row.

PM3 methods, which neglect differential overlap, Löwdin charges are the same as Mulliken charges. Throughout this article, we will denote q_k^0 and the dipole moment calculated from the q_k^0 charge distribution as the population analysis (PA) values of these quantities.

We will sometimes interchange the atomic symbol for the atomic number in $C_{Z_kZ_{k'}}$ and $D_{Z_kZ_{k'}}$, so $D_{78} \equiv D_{\mathrm{NO}}$, for example. Note that to maintain a constant total charge in the molecule, the charge transferred from one atom to another should be (and is) equal in magnitude and opposite in sign, i.e.,

$$C_{ZZ'} = -C_{Z'Z} \tag{2}$$

$$D_{ZZ'} = -D_{Z'Z} \tag{3}$$

In the present work, we wish to improve the description of charges for amides, a biologically significant functional group, and for compounds that contain bonds between N and O, a class of compounds that that includes many high-energy materials. So, when atom k is a nitrogen atom and atom k' is an oxygen atom, we define the CM3 charge to be

$$q_{\mathrm{N}} = q_k^0 + \sum_{\substack{k' \neq \mathrm{N}, \\ k' \neq \mathrm{O}}} B_{\mathrm{N}k'}(D_{\mathrm{N}k'} + C_{\mathrm{N}k'}B_{\mathrm{N}k'}) + \sum_{k'=\mathrm{O}} B_{\mathrm{NO}}(D_{\mathrm{NO}} + C_{\mathrm{NO}}e^{-(B_{\mathrm{NO}}/B_0)^2}) \tag{4}$$

where B_0 is a new parameter to be optimized. To maintain a constant total charge, an analogous definition of the CM3 charge on oxygen exists (and keeping in mind that $C_{\mathrm{NO}} = -C_{\mathrm{ON}}$ and $D_{\mathrm{NO}} = -D_{\mathrm{ON}}$).

Our motivation for introducing the new, final term in eq. (4) is primarily to improve our treatment of amide functional groups. The guiding philosophy of the CM*x* charge models is, in essence, to make systematic corrections to bond dipoles to improve the molecular dipole moment. However, functional groups like the amide group, and also the nitro group, involve at least two bonds that are coupled through π conjugation. Careful analysis of model performance indicates that in the absence of the new final term, amides and other N,O-containing functionalities tend to prefer different values of C_{NO} and D_{NO}, so that the final parameters simply reflect the weight given to these different functionalities in the training set. By adding the final term, which goes to zero at low bond order, to $B_0(C_{\mathrm{NO}} + D_{\mathrm{NO}})$ at some characteristic bond order B_0, and to B_0D_{NO} for higher bond orders, we may conveniently fit a wider range of functionalities.

Parameter Optimization

We use a training set of dipole moment data to determine the CM3 parameters used in eqs. (1) and (4). The training set, which consists of 398 experimental and high-quality theoretical dipole moment data for a diverse set of 390 polar compounds containing H, Li, C, N, O , F, Si, P, S, Cl, and Br, is the same training set used in our previous development of CM3.[4] This training set is fully described elsewhere,[5] but we note here that it consists of 382 non-amides for which we have experimental dipole moments in 323 cases and theoretical dipole moments in 59 cases. For molecules where an

Table 4. Root-Mean-Square (RMS) Errors (in Debyes) of Various Types of Dipole Moments Calculated by AM1 over Subsets of the CM3 Training Set Obtained by Removing Lithium-Containing Compounds.

Type of compound	No.[a]	Density	PA	CM2	CM3
Inorganic compounds	10	0.75	0.87	0.97	0.46
Alcohols, phenol	13	0.19	0.49	0.20	0.19
Ethers	11	0.20	0.32	0.19	0.22
Aldehydes	5	0.15	0.48	0.22	0.14
Ketones	11	0.34	0.63	0.19	0.22
Carboxylic acids	9	0.14	0.48	0.30	0.20
Esters	6	0.28	0.29	0.18	0.16
Other C, H, O compounds	12	0.38	0.58	0.35	0.28
Aliphatic amines, aniline	13	0.43	0.66	0.20	0.24
5-Membered aromatic N heterocycles	3	0.19	1.00	0.63	0.59
6-Membered aromatic N heterocycles	8	0.33	1.49	0.35	0.33
Nitriles	12	0.98	1.61	0.22	0.15
Imines	6	0.29	0.94	0.31	0.34
Other C, H, N compounds	14	0.95	1.55	0.28	0.27
Amides and phenylurea	17	0.45	0.94	0.22	0.12
Nitrohydrocarbons	5	0.66	0.74	0.21	0.35
Bifunctional H, C, N, O compounds	12	0.29	0.71	0.60	0.37
H, C, N, O polar compounds	163	0.49	0.94	0.37	0.27
Fluorine-containing compounds	39	0.22	0.32	0.32	0.31
Chlorine-containing compounds	33	0.19	0.17	0.14	0.14
Bromine-containing compounds	14	0.28	0.69	0.16	0.16
Halogenated bifunctional compounds	23	0.47	0.71	0.28	0.29
Thiols	8	0.39	0.69	0.28	0.23
Sulfides, disulfides	9	0.40	1.12	0.27	0.27
Other sulfur-containing compounds	23	0.58	1.07	0.85	0.75
Phosphorus	10	0.75	0.75	0.50	0.31
Multifunctional phosphorus	13	0.95	1.42	1.39	0.89
Compounds with S and P	7	2.32	1.95	2.34	0.77
C, H, Si	9	0.48	0.80	0.99	0.28
C, H, O, Si	9	0.68	0.87	0.85	0.48
C, H, Si, halogen	18	0.53	0.51	0.30	0.59
All polar compounds excluding lithium	382	0.59	0.86	0.59	0.39

[a]Number of data in the training set for this row.

experimental dipole moment is used, except for molecules containing Li and P, we use mPW1PW91/MIDI! geometries. For molecules containing Li and P, we use mPW1PW91/MG3S[37] geometries. The MG3S[37] basis set is obtained from the MG3[38] basis set, which is also called G3LargeMP2,[39] by deleting diffuse functions on H. When experimental dipole moments are not used or are not available, mPW1PW91/MG3S density dipole moments are used instead.

We also have eight amides, and we use mPW1PW91/MG3S density dipole moments for both a planar and a pyramidal conformation of each amide. For each conformation of each amide, we use HF[40]/MIDI! geometries. At the HF/MIDI! level of theory, all planar amides in our training set are first-order saddlepoints, and all pyramidal amides in our training set are minima.

This training set also includes two nonpolar molecules, namely ethene and benzene (we used mPW1PW91/MG3S geometries for these two compounds), which we use to determine the C—H parameter, as described below.

The first step in the parameterization is to determine D_{HC}. We carry out the determination of this parameter in the same manner as before,[4] i.e., we require the average charge on H in benzene and ethylene be 0.11. Also, as before,[4] we set $C_{CH} = 0$.

With the exception of the CM3 parameter between C and H, the optimal set of CM3 parameters are determined by minimizing the error function, ϵ, which is defined to be

$$\varepsilon = \sum_{i=1} (u_i(\text{Calc.}) - u_i(\text{Target}))^2 \tag{5}$$

with respect to the CM3 parameters. In eq. (5), the summation goes over all molecules in a training set, which is the relevant subset of our entire training set of polar molecules, and μ_i(Target) is the experimental or theoretical dipole moment of molecule i. The calculated dipole moment is given by

$$\mu_i(\text{Calc.}) = \left(\left(\sum_k q_{i,k} x_{i,k} \right)^2 + \left(\sum_k q_{i,k} y_{i,k} \right)^2 + \left(\sum_k q_{i,k} z_{i,k} \right)^2 \right)^{1/2} \tag{6}$$

Table 5. Root-Mean-Square (RMS) Errors (in Debyes) of Various Types of Dipole Moments Calculated by PM3 over Subsets of the Reduced CM2 Training Set Obtained by Removing Iodine-Containing Compounds.

Type of compound	No.[a]	Density	PA	CM2	CM3
Alcohols and water	12	0.12	0.69	0.13	0.11
Esters and lactones	8	0.38	0.64	0.17	0.20
Aldehydes and ketones	15	0.41	0.69	0.20	0.19
Carboxylic acids	9	0.11	0.52	0.31	0.22
Ethers	9	0.21	0.51	0.22	0.20
Amines and ammonia	13	0.30	0.99	0.17	0.11
Nitriles and HCN	17	0.69	1.54	0.23	0.16
Imines and N aromatics	7	0.35	1.27	0.27	0.27
Amides	3	0.47	0.88	0.36	0.18
Multifunctional N	7	0.68	1.47	0.13	0.22
H, C, N, O compounds	100	0.43	1.02	0.21	0.18
Fluorides	31	0.23	0.48	0.27	0.27
Chlorides	22	0.51	1.01	0.18	0.14
Bromides	10	0.27	0.41	0.16	0.16
Silicon	6	0.44	0.44	0.28	0.31
Sulfur	9	0.64	0.75	0.27	0.37
Phosphorus	13	0.39	1.38	0.51	0.67
Primary training set	191	0.42	0.93	0.25	0.27
Secondary training set	13	0.50	0.88	0.73	0.47
Primary and secondary sets	204	0.42	0.93	0.31	0.29

[a]Number of data in the training set for this row.

where $x_{i,k}$, $y_{i,k}$, and $z_{i,k}$ are the Cartesian coordinates of atom k in molecule i, and $q_{i,k}$ is the charge on atom k defined in eqs. (1) and (4). The parameter optimization is carried out using standard NAG Fortran 90 routines,[41] which optimize the parameters using the Newton–Raphson method. This optimization method uses numerical derivatives with respect to the CM3 parameters.

In this work, with two exceptions, we optimize the same set of $C_{ZZ'}$ and $D_{ZZ'}$ parameters as the ones that were nonzero in our previous development of CM3, namely D_{HC}, D_{HN}, D_{HO}, D_{HSi}, D_{HP}, D_{HS}, D_{LiC}, D_{LiN}, D_{LiO}, D_{LiF}, D_{LiS}, D_{LiCl}, D_{CN}, D_{CO}, D_{CF}, D_{CSi}, D_{CP}, D_{CS}, D_{CCl}, D_{CBr}, D_{NO}, D_{NP}, D_{OSi}, D_{OP}, D_{OS}, D_{FSi}, D_{FP}, D_{SiCl}, D_{PS}, D_{PCl}, C_{CO}, C_{OSi}, C_{OP}, and C_{PS}. The two exceptions are as follows: First, for the PM3 method, we use a C_{CN} parameter in addition to the D_{CN} parameter. Second, for all six parameterizations in this article, we treat the N—O charge map using eq. (4) instead of eq. (1); for this we optimize a C_{NO} parameter, a D_{NO} parameter, and a new B_0 parameter.

The application of eq. (5) is carried out in several stages. First, the parameters for molecules containing at most H, C, N, and O are determined. Then, with the parameters for H, C, N, and O fixed, the parameters for bonds between C—F, C—Cl, and C—Br are optimized. Then, these parameters are fixed, and the parameters involving sulfur are determined. Next, the parameters for silicon are determined. The parameter optimization for Si is carried out in two steps. First, the D_{HSi} and the D_{CSi} parameters are optimized using the subset of the training set that contains at most C, H, and Si. Second, these parameters are held fixed, and the remaining Si parameters (C_{OSi}, D_{OSi}, D_{FSi}, and D_{SiCl}) are optimized using the rest of the Si training set. With all of these parameters fixed, the parameters for Li, except for the AM1 method, and P are optimized.

Results and Discussion

The CM3 parameters for all six parameterizations are shown in Table 1. Table 2 gives the root-mean-square (RMS) errors in debyes of the resulting dipole moments over various subsets of the CM3 training set.

To evaluate the performance of each method, we compare CM3 for AM1 and PM3 to CM2 for AM1 and PM3. We also compare CM3 for BLYP to CM3 for mPWPW91 and CM3 for B3LYP to CM3 for mPW1PW91. Both CM3 for mPWPW91 and CM3 for mPW1PW91 were developed in our previous work.[4] Table 3 lists the RMS errors of the density, PA, CM2, and CM3 dipole moments calculated by the AM1 method over subsets of the CM2 training set, and Table 4 lists the RMS errors of the same types of dipole moments over the CM3 training set. Overall, the CM3 dipole moments are more accurate than the density dipole moments, particularly for inorganic compounds, aldehydes, ketones, esters, amines, nitriles, other C, H, N compounds, amides, and compounds containing both sulfur and phosphorus. CM3 dipole moments are also more accurate than PA dipole moments for both the CM2 and the CM3 training sets, as expected.

The RMS error of the CM2 dipole moments over the entire CM2 training set is 0.02 D smaller than the corresponding RMS error of the CM3 dipole moments, which may not be too surprising because the CM2 parameters were trained using this set of dipole moment data. However, CM3 still outperforms CM2 for carboxylic acids, nitriles, and amides. The sulfur and phosphorus subsets are two subsets of the CM2 training set where the CM2 dipoles moments are noticeably more accurate than the CM3 dipole moments.

Table 6. Root-Mean-Square (RMS) Errors (in Debyes) of Various Types of Dipole Moments Calculated by PM3 over Subsets of the CM3 Training Set.

Type of compound	No.[a]	Density	PA	CM2	CM3
Inorganic compounds	10	0.50	1.25	0.51	0.42
Alcohols, phenol	13	0.19	0.65	0.19	0.19
Ethers	11	0.20	0.47	0.21	0.20
Aldehydes	5	0.22	0.58	0.22	0.13
Ketones	11	0.40	0.73	0.18	0.20
Carboxylic acids	9	0.11	0.52	0.31	0.22
Esters	6	0.23	0.26	0.20	0.15
Other C, H, O compounds	12	0.44	0.73	0.26	0.27
Aliphatic amines, aniline	13	0.31	0.94	0.18	0.11
5-Membered aromatic N heterocycles	3	0.31	0.76	0.37	0.32
6-Membered aromatic N heterocycles	8	0.32	1.75	0.13	0.27
Nitriles	12	0.68	1.55	0.18	0.15
Imines	6	0.32	1.25	0.41	0.35
Other C, H, N compounds	14	0.69	1.49	0.37	0.35
Amides and phenylurea	17	0.41	0.84	0.34	0.18
Nitrohydrocarbons	5	0.33	0.30	0.25	0.24
Bifunctional H, C, N, O compounds	12	0.35	0.87	0.31	0.22
H, C, N, O polar compounds	163	0.41	1.01	0.28	0.24
Fluorine-containing compounds	39	0.26	0.44	0.34	0.34
Chlorine-containing compounds	33	0.49	1.00	0.16	0.12
Bromine-containing compounds	14	0.27	0.40	0.14	0.14
Halogenated bifunctional compounds	23	0.43	0.70	0.28	0.25
Thiols	8	0.48	0.77	0.21	0.13
Sulfides, disulfides	9	0.52	1.08	0.22	0.20
Other sulfur-containing compounds	23	0.59	0.87	0.65	0.50
Phosphorus	10	0.23	0.60	0.58	0.30
Multifunctional phosphorus	13	0.98	1.85	1.32	0.71
Compounds with S and P	7	1.38	0.81	1.37	0.45
C, H, Si	9	0.44	0.72	0.73	0.33
C, H, O, Si	9	0.85	0.55	0.63	0.40
C, H, Si, halogen	18	0.70	0.63	0.75	0.49
All polar compounds excluding lithium	382	0.51	0.92	0.48	0.32
Lithium compounds	16	2.43	2.11	[b]	0.41
All polar compounds	398	0.70	0.99	[b]	0.32

[a]Number of data in the training set for this row.
[b]The CM2 model was not parameterized for Li.

Using the CM3 training set, we find that CM3 dipole moments are more accurate than CM2 dipole moments by 0.2 D. Notable improvements of CM3 over CM2 are observed for inorganic compounds, carboxylic acids, other C, H, O compounds, nitriles, amides, bifunctional H, C, N, O compounds, phosphorus compounds, and compounds containing silicon. The more robust coverage of functionality in the CM3 training set compared to that in the CM2 one suggests that CM3 will be a more physical choice for application to molecules containing diverse combinations of functional groups.

Tables 5 and 6 provide the same RMS error data that Tables 3 and 4 do, but for the PM3 method. For PM3, the accuracy of the CM3 dipole moments relative to the accuracy of the density dipole moments and the PA dipole moments follow similar trends to those observed for AM1. The RMS error over the five-membered aromatic nitrogen heterocycles subset of the CM3 training set obtained from PM3 is, however, noticeably smaller than the corresponding RMS error obtained from AM1. We were able to achieve this by using the C_{CN} parameter in addition to the D_{CN} parameter for PM3. For all other methods used in this work, we did not observe significant improvements in the accuracy of the CM3 dipole moments when using this extra parameter.

Table 7 gives the RMS errors of the density, PA, and CM3 dipole moments calculated by BLYP/6-31G(d) over various subsets of the CM3 training set. Also listed are the RMS errors of the CM3 dipole moments calculated by mPWPW91/6-31G(d).[4] The accuracy obtained with the CM3 dipole moments is comparable to the accuracy obtained by using the density. (We note, however, that the density dipoles are not associated with partial charges, so they are not useful for many modeling applications.) However, improved accuracy in the dipole moments of fluorine-containing compounds and of compounds containing lithium is observed when CM3 charges are used.

Table 8 gives the range of bond orders observed in various types of bonds, and it helps to rationalize the usefulness of various

Table 7. Root-Mean-Square (RMS) Errors (in Debyes) of Density, Löwdin, and CM3 Dipole Moments Calculated by BLYP/6-31G(d) and CM3 Dipole Moments Calculated by mPWPW91/6-31G(d) over Various Subsets of the CM3 Training Set.

		BLYP			mPWPW91
Type of compound	No.[a]	Density	PA	CM3	CM3
Inorganic compounds	10	0.27	0.35	0.32	0.32
Alcohols, phenol	13	0.15	0.19	0.15	0.15
Ethers	11	0.14	0.22	0.18	0.18
Aldehydes	5	0.21	0.45	0.19	0.25
Ketones	11	0.30	0.72	0.26	0.25
Carboxylic acids	9	0.20	0.41	0.16	0.19
Esters	6	0.11	0.22	0.19	0.23
Other C, H, O compounds	12	0.30	0.55	0.30	0.29
Aliphatic amines, aniline	13	0.23	0.27	0.24	0.25
5-Membered aromatic N heterocycles	3	0.14	0.51	0.43	0.38
6-Membered aromatic N heterocycles	8	0.19	0.98	0.19	0.21
Nitriles	12	0.21	1.20	0.23	0.23
Imines	6	0.33	0.64	0.40	0.40
Other C, H, N compounds	14	0.18	1.09	0.19	0.19
Amides and phenylurea	17	0.37	0.58	0.08	0.14
Nitrohydrocarbons	5	0.31	0.29	0.16	0.15
Bifunctional H, C, N, O compounds	12	0.28	0.63	0.23	0.29
H, C, N, O polar compounds	163	0.25	0.66	0.23	0.23
Fluorine-containing compounds	39	0.46	0.25	0.22	0.22
Chlorine-containing compounds	33	0.14	0.62	0.15	0.15
Bromine-containing compounds	14	0.11	0.40	0.17	0.17
Halogenated bifunctional compounds	23	0.30	0.45	0.24	0.25
Thiols	8	0.15	0.71	0.12	0.12
Sulfides, disulfides	9	0.17	1.27	0.19	0.18
Other sulfur-containing compounds	23	0.26	0.74	0.50	0.52
Phosphorus	10	0.19	0.91	0.38	0.38
Multifunctional phosphorus	13	0.35	0.84	0.48	0.47
Compounds with S and P	7	0.37	0.86	0.17	0.17
C, H, Si	9	0.14	0.52	0.08	0.08
C, H, O, Si	9	0.37	0.43	0.26	0.26
C, H, Si, halogen	18	0.24	0.40	0.16	0.16
All polar compounds excluding lithium	382	0.28	0.63	0.26	0.26
Lithium compounds	16	0.85	4.84	0.33	0.30
All polar compounds	398	0.32	1.15	0.26	0.26

[a]Number of data in the training set for this row.

terms in the mapping. For example, when the N—O bond order is small, as it is in amides, it is primarily the C_{NO} term that is used to transfer charge between N and O. For nitrohydrocarbons, where the bond order between N and O is approximately 1.5, it is the D_{NO} parameter that determines the amount of charge transfer between N and O.

Table 8 also shows that there is a larger variation in the bond order for basis sets with diffuse functions [6-31+G(d)] than for tight basis sets [MIDI!6D and 6-31G(d)] and minimal basis sets (AM1 and PM3). In addition, Table 8 shows that the Mayer bond order is more sensitive to basis set size than it is to the electronic structure method employed. For example, there is more similarity between the bond orders calculated by BLYP/6-31G(d) and B3LYP/6-31G(d) than there is between B3LYP/MIDI!6D and B3LYP/6-31G(d).

The CM3 model in this work uses a different functional form to map the population analysis charges on N and O, so it is worthwhile to compare CM3 dipole moments obtained here from BLYP/6-31G(d) to the CM3 dipole moments obtained from mPWPW91/6-31G(d), where the functional form of eq. (1) was employed for all atomic pairs.[4] Table 7 shows that the CM3 dipole moments of amides and bifunctional H, C, N, O compounds are more accurate when BLYP and the new treatment of compounds with N and O are used than when the corresponding dipole moments are calculated by the previous CM3 model for mPWPW91/6-31G(d). Furthermore, the dipole moments of aldehydes, carboxylic acids, and esters are slightly more accurate than the previous CM3 model. Over the entire CM3 training set, both BLYP and mPWPW91 give very comparable RMS errors.

Tables 9 and 10 give the RMS errors of the density, PA, and CM3 dipole moments calculated by B3LYP/MIDI!6D (Table 9) and B3LYP/6-31G(d) (Table 10) over various subsets of the CM3 training set. Also given in Tables 9 and 10 are the RMS errors of

Table 8. Ranges of Observed Bond Orders.

			BLYP	B3LYP		
Type of bond	AM1	PM3	6-31G(d)	MIDI!6D	6-31G(d)	6-31+G(d)
N—O bond						
In amides	0.14–0.18	0.14–0.20	0.07–0.13	0.10–0.15	0.06–0.11	−0.04–0.05
In nitrohydrocarbons	1.45–1.47	1.49–1.51	1.48–1.54	1.49–1.52	1.49–1.53	1.81–2.18
C—N bond						
In amides	0.93–1.11	0.97–1.18	0.92–1.17	0.88–1.08	0.90–1.16	0.83–1.26
In nitrohydrocarbons	0.81–0.87	0.88–0.92	0.78–0.84	0.72–0.80	0.79–0.83	0.64–0.79
In amines	0.99–1.07	0.99–1.09	0.96–1.06	0.90–1.03	0.95–1.04	0.59–1.55
In nitriles	2.89–2.97	2.89–2.98	2.79–2.86	2.64–2.82	2.80–2.87	2.03–2.70
In imines and N aromatics	1.17–1.27	1.23–1.30	1.12–1.30	1.06–1.19	1.11–1.29	1.25–1.33

Table 9. Root-Mean-Square (RMS) Errors (in Debyes) of Density, Löwdin, and CM3 Dipole Moments Calculated by B3LYP/MIDI!6D and CM3 Dipole Moments Calculated by mPW1PW91/MIDI!6D over Various Subsets of the CM3 Training Set.

		B3LYP			mPW1PW91
Type of compound	No.[a]	Density	PA	CM3	CM3
Inorganic compounds	10	0.27	0.65	0.27	0.26
Alcohols, phenol	13	0.15	0.45	0.15	0.15
Ethers	11	0.17	0.71	0.22	0.23
Aldehydes	5	0.36	0.26	0.16	0.23
Ketones	11	0.57	0.22	0.33	0.31
Carboxylic acids	9	0.24	0.43	0.24	0.29
Esters	6	0.20	0.25	0.09	0.11
Other C, H, O compounds	12	0.45	0.52	0.30	0.29
Aliphatic amines, aniline	13	0.21	0.16	0.35	0.36
5-Membered aromatic N heterocycles	3	0.21	0.67	0.45	0.39
6-Membered aromatic N heterocycles	8	0.28	0.42	0.17	0.20
Nitriles	12	0.35	0.48	0.19	0.20
Imines	6	0.31	0.39	0.51	0.51
Other C, H, N compounds	14	0.33	0.50	0.21	0.21
Amides and phenylurea	17	0.51	0.28	0.11	0.19
Nitrohydrocarbons	5	0.43	0.19	0.19	0.21
Bifunctional H, C, N, O compounds	12	0.39	0.18	0.29	0.33
H, C, N, O polar compounds	163	0.35	0.42	0.26	0.27
Fluorine-containing compounds	39	0.50	0.39	0.23	0.23
Chlorine-containing compounds	33	0.44	0.17	0.15	0.15
Bromine-containing compounds	14	0.19	0.39	0.16	0.15
Halogenated bifunctional compounds	23	0.45	0.41	0.30	0.30
Thiols	8	0.39	0.38	0.10	0.10
Sulfides, disulfides	9	0.40	0.38	0.12	0.12
Other sulfur-containing compounds	23	0.44	0.69	0.49	0.49
Phosphorus	10	0.19	0.49	0.14	0.14
Multifunctional phosphorus	13	0.43	0.68	0.16	0.16
Compounds with S and P	7	0.48	0.37	0.08	0.10
C, H, Si	9	0.13	0.32	0.08	0.08
C, H, O, Si	9	0.54	0.78	0.26	0.26
C, H, Si, halogen	18	0.46	0.40	0.19	0.19
All polar compounds excluding lithium	382	0.40	0.44	0.25	0.25
Lithium compounds	16	0.58	0.67	0.60	0.63
All polar compounds	398	0.41	0.46	0.27	0.28

[a]Number of data in the training set for this row.

Table 10. Root-Mean-Square (RMS) Errors (in Debyes) of Density, Löwdin, and CM3 Dipole Moments Calculated by B3LYP/6-31G(d) and CM3 Dipole Moments Calculated by mPW1PW91/6-31G(d) over Various Subsets of the CM3 Training Set.

		B3LYP			mPW1PW91
Type of compound	No.[a]	Density	PA	CM3	CM3
Inorganic compounds	10	0.27	0.32	0.32	0.31
Alcohols, phenol	13	0.18	0.25	0.15	0.15
Ethers	11	0.09	0.32	0.15	0.16
Aldehydes	5	0.13	0.32	0.16	0.22
Ketones	11	0.17	0.56	0.25	0.23
Carboxylic acids	9	0.27	0.36	0.18	0.22
Esters	6	0.10	0.21	0.19	0.22
Other C, H, O compounds	12	0.27	0.43	0.28	0.27
Aliphatic amines, aniline	13	0.25	0.22	0.24	0.25
5-Membered aromatic N heterocycles	3	0.10	0.47	0.40	0.35
6-Membered aromatic N heterocycles	8	0.11	0.90	0.19	0.21
Nitriles	12	0.19	1.11	0.23	0.23
Imines	6	0.30	0.58	0.40	0.40
Other C, H, N compounds	14	0.17	1.00	0.17	0.16
Amides and phenylurea	17	0.24	0.45	0.07	0.16
Nitrohydrocarbons	5	0.17	0.15	0.15	0.17
Bifunctional H, C, N, O compounds	12	0.17	0.51	0.21	0.27
H, C, N, O polar compounds	163	0.20	0.58	0.22	0.23
Fluorine-containing compounds	39	0.30	0.30	0.21	0.21
Chlorine-containing compounds	33	0.19	0.53	0.14	0.14
Bromine-containing compounds	14	0.12	0.31	0.16	0.15
Halogenated bifunctional compounds	23	0.27	0.48	0.22	0.23
Thiols	8	0.18	0.70	0.12	0.12
Sulfides, disulfides	9	0.20	1.24	0.21	0.20
Other sulfur-containing compounds	23	0.28	0.78	0.51	0.54
Phosphorus	10	0.15	0.92	0.41	0.41
Multifunctional phosphorus	13	0.22	0.94	0.48	0.48
Compounds with S and P	7	0.31	0.94	0.18	0.18
C, H, Si	9	0.13	0.53	0.08	0.08
C, H, O, Si	9	0.26	0.43	0.25	0.25
C, H, Si, halogen	18	0.24	0.50	0.15	0.15
All polar compounds excluding lithium	382	0.23	0.61	0.25	0.26
Lithium compounds	16	0.64	4.69	0.30	0.27
All polar compounds	398	0.26	1.11	0.25	0.26

[a]Number of data in the training set for this row.

the CM3 dipole moments calculated by mPW1PW91/MIDI!6D and mPW1PW91/6-31G(d), respectively. Tables 9 and 10 show trends similar to those observed in Table 7. In particular, the CM3 model for B3LYP is slightly more accurate for several of the subsets of the training set that contain both N and O and for several subsets of the training set that contain at most C, H, and O.

Although the accuracies of CM3 dipole moments when calculated by B3LYP/MIDI!6D and B3LYP/6-31G(d) are similar, the time required to calculate a CM3 dipole moment using these two basis sets is different. For example, it takes approximately 1.8 times longer to carry out a single-point calculation of either pyramidal benzamide or phenylurea when using 6-31G(d) than when using MIDI!6D in Gaussian98.[42] Note that Gaussian98 takes advantage of the fact that 6-31G(d) uses sp shells to describe the valence space (i.e., for a particular valence shell, the s and p orbital exponents are identical). On the other hand, with the exception of Si, Br, and I, MIDI!6D uses different s and p orbital exponents for each valence shell. (If sp-type shells are not assumed to be identical in the 6-31G(d) basis set, then it takes approximately 2.7 times longer to carry out a single point calculation of either pyramidal benzamide or phenylurea when using 6-31G(d) than when using MIDI!6D.)

Table 11 gives the RMS errors of the density, PA, and CM3 dipole moments calculated by B3LYP/6-31+G(d) over various subsets of the CM3 training set. Also given in Table 10 are the RMS errors of the CM3 dipole moments calculated by mPW1PW91/6-31+G(d). The RMS error for the amide data is 0.04 D smaller for B3LYP than for mPW1PW91, and there is also similar increased accuracy for bifunctional H, C, N, O compounds.

Table 11. Root-Mean-Square (RMS) Errors (in Debyes) of Density, Löwdin, and CM3 Dipole Moments Calculated by B3LYP/6-31+G(d) and CM3 Dipole Moments Calculated by mPW1PW91/6-31+G(d) over Various Subsets of the CM3 Training Set.

		B3LYP			mPW1PW91
Type of compound	No.[a]	Density	PA	CM3	CM3
Inorganic compounds	10	0.32	0.35	0.38	0.39
Alcohols, phenol	13	0.34	0.35	0.26	0.26
Ethers	11	0.20	0.42	0.24	0.25
Aldehydes	5	0.40	0.20	0.28	0.30
Ketones	11	0.29	0.20	0.20	0.20
Carboxylic acids	9	0.42	0.43	0.34	0.35
Esters	6	0.18	0.24	0.25	0.26
Other C, H, O compounds	12	0.40	0.33	0.25	0.25
Aliphatic amines, aniline	13	0.34	0.17	0.30	0.27
5-Membered aromatic N heterocycles	3	0.12	0.79	0.59	0.59
6-Membered aromatic N heterocycles	8	0.16	0.72	0.26	0.24
Nitriles	12	0.31	1.06	0.29	0.27
Imines	6	0.38	0.53	0.44	0.43
Other C, H, N compounds	14	0.28	0.85	0.28	0.27
Amides and phenylurea	17	0.09	0.37	0.33	0.37
Nitrohydrocarbons	5	0.37	0.46	0.28	0.23
Bifunctional H, C, N, O compounds	12	0.17	0.35	0.23	0.28
H, C, N, O polar compounds	163	0.29	0.53	0.29	0.30
Fluorine-containing compounds	39	0.18	0.69	0.40	0.39
Chlorine-containing compounds	33	0.22	0.69	0.29	0.27
Bromine-containing compounds	14	0.21	0.83	0.33	0.29
Halogenated bifunctional compounds	23	0.23	1.04	0.27	0.26
Thiols	8	0.19	0.67	0.22	0.22
Sulfides, disulfides	9	0.22	1.31	0.30	0.28
Other sulfur-containing compounds	23	0.39	1.15	0.60	0.63
Phosphorus	10	0.10	0.72	0.38	0.38
Multifunctional phosphorus	13	0.18	1.70	0.30	0.28
Compounds with S and P	7	0.11	1.58	0.30	0.41
C, H, Si	9	0.07	0.27	0.20	0.21
C, H, O, Si	9	0.18	1.03	0.35	0.36
C, H, Si, halogen	18	0.22	1.37	0.62	0.65
All polar compounds excluding lithium	382	0.26	0.84	0.36	0.36
Lithium compounds	16	0.20	3.72	0.39	0.42
All polar compounds	398	0.25	1.11	0.36	0.36

[a]Number of data in the training set for this row.

Tables 9, 10, and 11 also show that the density and PA dipole moments are more sensitive to basis set size than are the CM3 dipole moments. For the 382 polar data that do not include lithium, the RMS error of the density dipole moments changes from 0.40 D when using MIDI!6D to 0.23 D when using 6-31G(d) to 0.26 D when using 6-31+G(d). The RMS error of the PA dipole moments over the 382 data is 0.44 D, 0.61 D, and 0.84 D when using MIDI!6D, 6-31G(d), and 6-31+G(d), respectively. On the other hand, the RMS error of CM3 dipole moments over the 382 polar data is 0.25 D, 0.25 D, and 0.36 D when using MIDI!6D, 6-31G(d), 6-31+G(d), respectively. The density dipole moments of lithium-containing compounds are less sensitive to the difference between the MIDI!6D and 6-31G(d) basis sets than the corresponding CM3 and PA dipole moments are. However, for the lithium data the RMS error of the CM3 dipole moments only changes from 0.30 D when using 6-31G(d) to 0.39 D when using 6-31+G(d), whereas the RMS error of the density dipole moments changes by 0.44 D.

To show that CM3 charges remain reasonably constant from one level of theory to another, Figures 1–7 present PA and CM3 charges calculated by B3LYP/MIDI!6D and by AM1 for formaldehyde, dimethyl ether, formic acid, methyl formate, planar and pyramidal formamide, nitromethane, and fluorosilane. These examples show that PA charges are more sensitive to the level of theory employed than the corresponding CM3 charges are. For example, for the molecules in Figures 1–6, the difference between the PA charges on oxygen calculated by B3LYP/MIDI!6D and by AM1 range from 0.10 to 0.20, with an average difference of 0.13, while the corresponding differences between the CM3 charges on oxygen calculated by these two levels of theory range from 0.01 to

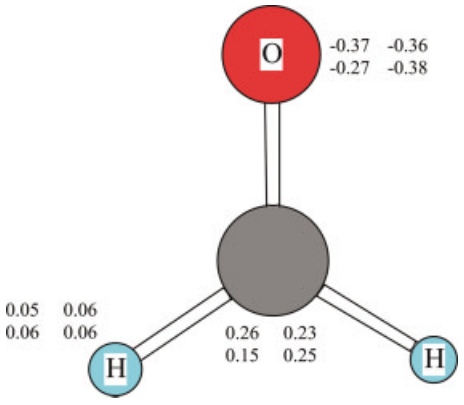


Figure 1. PA and CM3 partial atomic charges in formaldehyde. There are two rows of numbers beside each symmetry-unique atom. The first row of numbers corresponds to the PA charge (given first) and the CM3 charge (given second) of the atom calculated by B3LYP/MIDI!6D. The second row of numbers corresponds to the PA charge (given first) and the CM3 charge (given second) of the atom calculated by AM1.

0.09, with an average difference of 0.04. In addition, the PA charge on Si in fluorosilane is calculated to be 0.30 and 1.16 by B3LYP/MIDI!6D and AM1, respectively, but the corresponding mapped CM3 charges on Si are much closer to each other, with values of 0.29 and 0.34. Figures 5 and 6 show that the differences between the PA charges on N calculated by AM1 and B3LYP/MIDI!6D are, however, smaller than the corresponding differences between CM3 charges on N.

As another point of discussion, we wish to emphasize that the "quality" of a given charge model depends intimately on the use to which it is to be put. The CM3 charge models by design provide dipole moments of very high quality when these moments are computed pointwise from the partial atomic charges. To the extent that the molecular electrostatic potential at long range is dominated by the dipole moment (and, of course, monopole moment if the molecule is charged), we expect CM3 charges to be useful for computing electrostatic interactions at medium to long range. Such interactions are important in explicit molecular simulations, and also in the modeling of solvation effects using continuum self-consistent reaction field models, and we expect CM3 charges to be

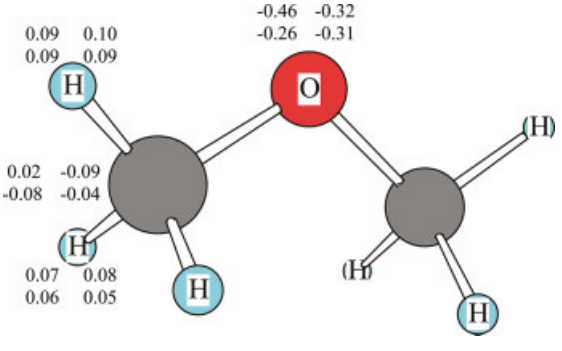


Figure 2. Same as Figure 1 except for dimethyl ether.

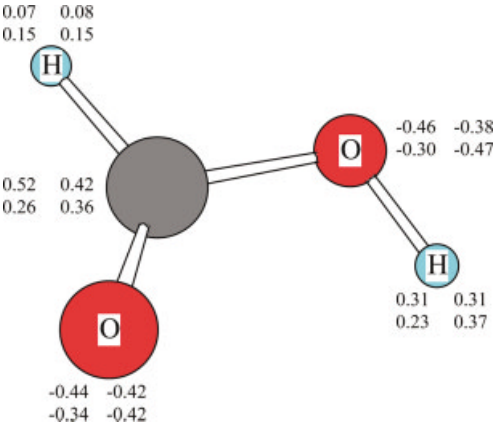


Figure 3. Same as Figure 1 except for formic acid.

particularly useful in these contexts. Because, however, the model is designed to obtain not just accurate overall dipole moments but also accurate bond dipoles throughout the molecule, it should also yield accurate electrostatic potentials at shorter distances, and hence, it should also be useful for solvation models that are sensitive to these shorter range interactions. We have, however, not explicitly evaluated the quality of higher order electrical moments computed from the CM3 charges for most cases, and those interested specifically in these properties would be well advised to test the accuracy.

We also expect CM3 charges to be useful for the qualitative analysis of structural and reactivity properties, in much the same sense that other charge models are used when they are applied in a consistent fashion across a body of data. In such applications, as a rule, drawing conclusions from results obtained from *different* charge models is best avoided, because trends in partial charges are typically more informative than absolute values, and such trends will be more meaningful within a charge model.

Because most of this article has stressed differences of CM3 from CM2 and details of performance of various parameterizations, it is probably useful to close by mentioning a few key

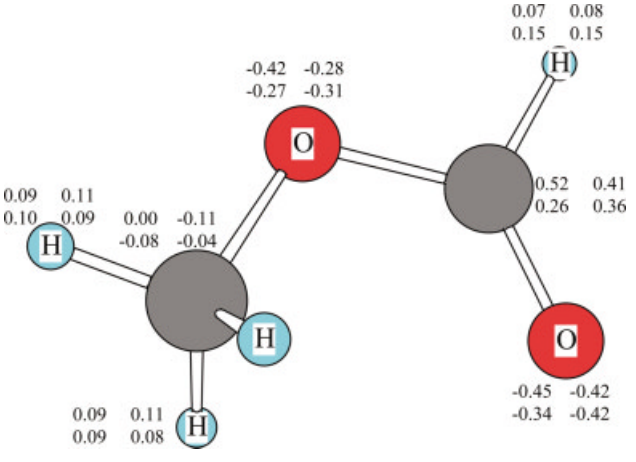
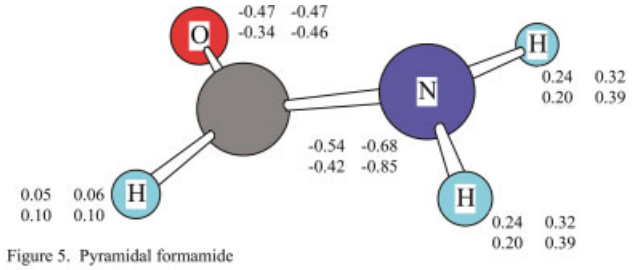


Figure 4. Same as Figure 1 except for methyl formate.



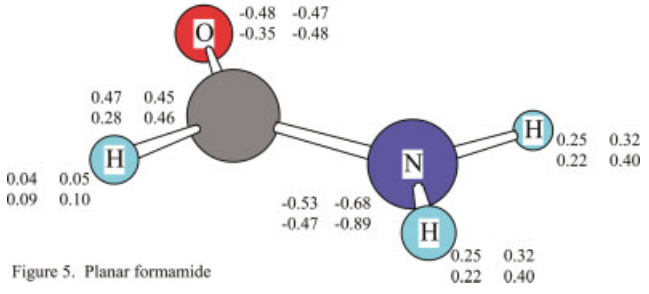


Figure 5. Same as Figure 1 except for (a) pyramidal formamide, with the hydrogens 0.29 Å below the OCN plane, and (b) formamide with the amino group and the carbonyl group in the same plane. The geometries of both of these conformations of formamdie are from HF/MIDI!.

advantages shared by all CM*x* models. The CM*x* models are class IV charge models. They are designed to yield accurate charge distributions for large molecules by employing empirical parameters in the interpretation of the approximate wave function, as contrasted to the more usual procedure of only incorporating parameters in the Hamiltonian. This allows one to obtain more accurate charges than those inherent in the approximate wave function itself. Furthermore, the approach is designed to yield accurate molecular charge distributions in multifunctional molecules, not by parameterizing them directly, but rather by obtaining the overall charge distribution as a superposition of realistically modeled dipolar bonds. The method is equally applicable to atoms on the surface of a large molecule as to atoms buried in its interior, whereas electrostatic fitting models are often unstable for buried atoms and may give quite different charge distributions across a bond, depending on its distance from the fitting points in the region exterior to the molecule. Another factor contributing to the stable nature of CM*x* charge distributions is the local character of the class IV mapping, which may be contrasted with the global nature of electrostatic fitting. Because CM3 is our most mature model with these advantages, we hope it will be useful for a wide variety of chemical modeling purposes.

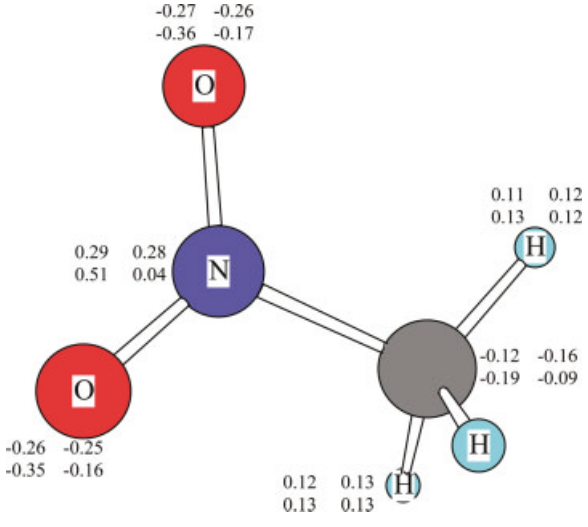


Figure 6. Same as Figure 1 except for nitromethane.

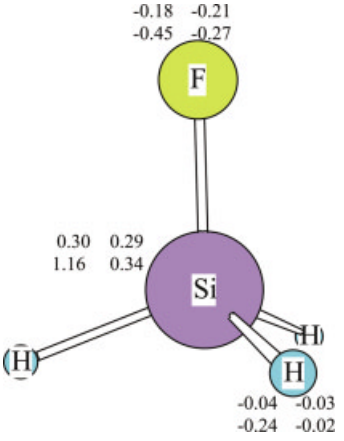


Figure 7. Same as Figure 1 except for fluorosilane.

Conclusions

In this article, we have presented CM3 charge model parameterizations for the AM1, PM3, BLYP, and B3LYP methods. The CM3 model developed in this work is very similar to the CM3 model developed previously,[4] except that it maps charges on N and O using an extended functional form. As a consequence of this improvement, the CM3 charge model for the methods presented here is particularly successful for aldehydes, carboxylic acids, esters, amides, and bifunctional H, C, N, O compounds. We also use an additional C_{CN} parameter for the PM3 method, which yields improved accuracy in dipole moments of 5-membered aromatic nitrogen heterocycles.

Availability of the New Parameters

The new CM3 parameters are available in several electronic structure programs, namely AMSOL,[43] GAMESSPLUS,[44] HONDOPLUS,[45,46] and MN-GSM.[42,47] In particular, the new CM3 parameters for the AM1 and PM3 methods are available in AMSOL and GAMESSPLUS, and the new CM3 parameters for BLYP and B3LYP are available in HONDOPLUS, GAMESSPLUS, and MN-GSM.

References

1. Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. J Comput-Aid Mol Des 1995, 9, 87.
2. Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. J Phys Chem A 1998, 102, 1820.
3. Li, J.; Williams, B.; Cramer, C. J.; Truhlar, D. G. J Chem Phys 1999, 110, 724.
4. Winget, P.; Thompson, J. D.; Xidos, J. D.; Cramer, C. J.; Truhlar, D. G. J Phys Chem A 2002, 106, 10707.
5. Thompson, J. D.; Xidos, J. D.; Sonbuchner, T. M.; Cramer, C. J.; Truhlar, D. G. PhysChemComm 2002, 5, 117.
6. Mulliken, R. S. J Chem Phys 1935, 3, 564.
7. Mulliken, R. S. J Chem Phys 1955, 23, 1833.
8. Mulliken, R. S. J Chem Phys 1962, 36, 3428.
9. Löwdin, P.-O. J Chem Phys 1950, 18, 365.
10. Golebiewski, A.; Rzescowksa, E. Acta Phys Polym A 1974, 45, 563.
11. Baker, J. Theor Chim Acta 1985, 68, 221.
12. Kar, T.; Sannigrahi, A. B. Mukherjee, D. C. J Mol Struct (Theochem) 1987, 153, 93.
13. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. J Am Chem Soc 1985, 107, 3902.
14. Dewar, M. J. S.; Jie, C. Organometallics 1987, 6, 1486.
15. Dewar, M. J. S.; Zoebisch, E. G. J Mol Struct (Theochem) 1988, 49, 1.
16. Dewar, M. J. S.; Lie, C. J Mol Struct (Theochem) 1989, 56, 1.
17. Dewar, M. J. S.; Yuan, Y. C. Inorg Chem 1990, 29, 3981.
18. Stewart, J. J. P. J Comput Chem 1989, 10, 221.
19. Thiel, W. QCPE Bull 1982, 2, 36.
20. Anders, E.; Koch, R.; Freunscht, P. J Comput Chem 1993, 11, 1301.
21. Becke, A. D. Phys Rev 1988, 38, 3098.
22. Becke, A. D. J Chem Phys 1993, 98, 5648.
23. Lee, C.; Yang, W.; Parr, R. G. Phys Rev B 1988, 37, 785.
24. Stephens, P. J.; Devlin, J. F.; Chabalowski, C. F.; Frisch, M. J. J Phys Chem 1994, 98, 11623.
25. Easton, R. E.; Giesen, D. J.; Welch, A.; Cramer, C. J.; Truhlar, D. G. Theor Chim Acta 1996, 93, 281.
26. Li, J.; Cramer, C. J.; Truhlar, D. G. Theor Chem Acc 1998, 99, 192.
27. Thompson, J. D.; Winget, P.; Truhlar, D. G. Phys Chem Commun 2001, 4/16, 1.
28. Ditchfield, R.; Hehre, W. J.; Pople, J. A. J Chem Phys 1971, 54, 724.
29. Hehre, W. J.; Ditchfield, R.; Pople, J. A. J Chem Phys 1972, 56, 2257.
30. Hariharan, P. C.; Pople, J. A. Chem Phys Lett 1972, 66, 217.
31. Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. v. R. J Comput Chem 1983, 4, 294.
32. Adamo, C.; Barone, V. J Chem Phys 1998, 108, 664.
33. Perdew, J. P. In Electronic Structure of Solids '91; Ziesche, P., Eschrig, H., Eds.; Akademie Verlag: Berlin, 1991, p. 11.
34. Mayer, I. Chem Phys Lett 1983, 97, 270.
35. Mayer, I. Chem Phys Lett 1985, 117, 396.
36. Mayer, I. Int J Quantum Chem 1986, 29, 73.
37. Lynch, B. J.; Zhao, Y.; Truhlar, D. G. J Phys Chem A 2003, 107, 1384.
38. Fast, P. L.; Sanchez, M. L.; Truhlar, D. G. Chem Phys Lett 1999, 306, 407.
39. Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Rassolov, V.; Pople, J. A. J Chem Phys 1999, 110, 4703.
40. Roothaan, C. C. J. Rev Mod Phys 1951, 23, 69.
41. NAG Fortran 90 Library; The Numerical Algorithms Group, Inc.: Oxford, 2000, 4th ed.
42. Frisch, M. J.; Trucks, G. W.; Schlegel; H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q .; Morokuma, K.; Salvador, P.; Dannenberg, J. J.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head–Gordon, M.; Replogle, E. S .; and Pople, J. A. Gaussian98 (Revision A.11); Gaussian, Inc.: Pittsburgh, PA, 2002.
43. Hawkins, G. D.; Giesen, D. J.; Lynch, G. C.; Chamber, C. C.; Rossi, I.; Storer, J. W.; Li, J.; Thompson, J. D.; Winget, P.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. Amsol, version 6.8; University of Minnesota: Minneapolis, MN, 2002, based in part on Ampac, version 2.1; Liotard, D. A.; Healy, E. F.; Ruiz, J. M.; Dewar, M. J. S.
44. Li, J.; Zhu, T.; Hawkins, G. D.; Chuang, Y.-Y.; Fast, P. L.; Thompson, J. D.; Xidos, J. D.; Liotard, D. A.; Rinaldi, D.; Cramer, C. J.; Truhlar, D. G. GAMESSPLUS, version 3.9; University of Minnesota, Minneapolis, 2002, based on the General Atomic and Molecular Electronic Structure System (Gamess) as described in Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery J. A. J Comp Chem 1993, 14, 1347.
45. Dupuis, M.; Marquez, A.; Davidson, E. R. HONDO 99.6, 1999, based on Hondo 95.3, Dupuis, M.; Marquez, A.; Davidson, E. R. Quantum Chemistry Program Exchange (QCPE); Indiana University, Bloomington, IN.
46. HONDOPLUS, v.4.2, by Nakamura, H.; Xidos, J. D.; Thompson, J. D.; Li, J.; Hawkins, G. D.; Zhu, T.; Lynch, B. J.; Volobuev, Y.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. University of Minnesota, MN, 2002, based on Hondo, v.99.6.
47. Xidos, J. D.; Li, J.; Thompson, J. D.; Hawkins, G. D.; Winget, P. D.; Zhu, T.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G.; Frisch, M. J. MN-GSM, version 2.2; Minneapolis, MN: University of Minnesota, 2002.