

Accurate dipole moments from Hartree–Fock calculations by means of class IV charges

Jiabo Li, Jianhua Xing, Christopher J. Cramer, and Donald G. Truhlar^{a)}

Department of Chemistry and Supercomputer Institute, University of Minnesota, 207 Pleasant St. SE, Minneapolis, Minnesota 55455-0431

(Received 1 February 1999; accepted 6 April 1999)

Charge Model 2 (CM2) is parameterized for Hartree–Fock calculations with the correlation-consistent polarized valence double zeta (cc-pVDZ) basis set. The resulting charge model has an RMS error of 0.18 D over a training set of 198 polar molecules. The charge model is additionally applied to 8 nucleic acid bases and methyl azide to test its performance for nitrogen-containing compounds not found in the training set. The results demonstrate that this new CM2 model based on *ab initio* Hartree–Fock calculations is robust in predicting the charge distributions of such molecules. Comparison of CM2 results for the nitrogen-containing test set with those from a previous charge model, charge model 1 (CM1) based on AM1 (Austin model 1) and PM3 (parameterized model 3) wave functions, indicate that the CM2 charges are more accurate than those from the previous model. © 1999 American Institute of Physics. [S0021-9606(99)30125-2]

I. INTRODUCTION

Partial atomic charge is a very useful concept for the interpretation of molecular properties.¹ However, the partial atomic charge on an atom in a molecule cannot be unambiguously defined. Instead, to discuss the partial charges within a molecule, some model must be specified. Two of the most widely used schemes for assigning partial charges, their popularity owing in part to their simplicity, are Mulliken^{2,3} and Löwdin⁴ population analysis. Because of the mathematical (rather than physical) basis of these models, partial charges derived from them can be more sensitive to basis set size and composition than is warranted by the variation in the three-dimensional electron density functions corresponding to the wave functions.

We have previously presented a new class IV charge model, namely charge model 2 (CM2), for calculating partial atomic charges in molecules,⁵ and we parameterized for four basis sets with *ab initio* Hartree–Fock (HF) theory,⁶ for two semiempirical molecular orbital models (AM1^{7–11} and PM3¹²), for four basis sets with density functional theory (DFT),^{13–16} and for one basis set with hybrid Hartree–Fock density-functional theory.¹⁷ We have also¹⁸ parameterized the model for the intermediate neglect of differential overlap for spectroscopy models INDO/S^{19,20} and INDO/S2,¹⁸ where its performance was robust for molecular excited states. In the present paper we present a parameterization for a fifth basis set for Hartree–Fock theory for ground states, namely the cc-pVDZ basis.^{21,22} Unlike the previous basis sets employed with Hartree–Fock theory for which CM2 has been defined, the cc-pVDZ basis set has polarization functions on hydrogen atoms. Moreover, the cc-pVDZ basis set has been developed to provide high-quality results in correlated calculations,^{21,22} so a Hartree–Fock model may prove useful

in further developments based on correlated cc-pVDZ charge distributions.

Section II discusses the parameterization itself. Section III summarizes the results of the parameterization for the 198 molecules in the CM2 training set. Section IV provides a discussion of the results and a test of the CM2 charge model for methyl azide and 8 nucleic acid bases. These compounds contain several different kinds of nitrogen atoms (including nitrogen–nitrogen bonds in methyl azide) and were not included in the training set.

II. CM2 PARAMETERIZATION

Since the concept of class IV charges²³ and the formalism for CM2⁵ have already been presented in detail in the original papers, we recapitulate here only their most critical aspects. A class IV charge model takes as its input a set of charges from some population analysis of a wave function (or any class II or class III scheme) and, in a parameterized procedure, maps them to optimally reproduce *charge-dependent observables obtained from experiment* (or, in special instances, highly accurate theoretical calculations). The parameterization procedure is designed to correct for (i) systematic errors attributable to the particular population analysis (or other input) method and (ii) systematic errors attributable to an incomplete basis set and/or level of electronic structure theory.

There are two choices that define a specific Class IV charge model:²³ (i) The method for obtaining the lower-level charges to be mapped and (ii) the functional form of the mapping itself. Charge model 1 (CM1²³) was the first Class IV charge model to be developed and was based on mapping Mulliken² charges of Austin model 1 (AM1^{7–11}) or parameterized model 3 (PM3¹²) wave functions. (Mulliken charges are so-called Class II²³ charges because they are derived from an arbitrary population analysis of a wave function.) Charge model 2 (CM2) represents the second generation of

^{a)}Electronic mail: truhlar@umn.edu

Class IV charge model, and it differs from CM1 for each of the two major choices associated with a Class IV charge model. First, the input charges are based on a Löwdin⁴ population analysis of the relevant wave functions. For semiempirical methods assuming a unit overlap matrix, as most do, Mulliken and Löwdin charges are identical; both reduce to what we call zero-overlap Mulliken charges and what some others call Coulson²⁴ charges; but this is not the case when *ab initio* Hartree–Fock or DFT wave functions are employed. Second, a simpler and more stable mapping functional was adopted. Such a mapping functional is expected to be more reliable when applied to molecules outside of the training set.

The CM2 charge on atom k is defined as

$$q_k^{\text{CM2}} = q_k^L + \sum_{k' \neq k} T_{kk'}(B_{kk'}), \quad (1)$$

where q_k^L is the Löwdin charge on atom k . The mapping function $T_{kk'}$ transfers charge between atoms k and k' according to

$$T_{kk'} = B_{kk'}(D_{Z_k Z_{k'}} + C_{Z_k Z_{k'}} B_{kk'}), \quad (2)$$

where $D_{Z_k Z_{k'}}$ and $C_{Z_k Z_{k'}}$ are parameters depending on the atomic numbers Z_k and $Z_{k'}$ of atoms k and k' , and $B_{kk'}$ is the Mayer generalization²⁵ of the bond order of Armstrong *et al.*,²⁶ defined between atoms k and k' as

$$B_{kk'} = \sum_{i \in k} \sum_{j \in k'} \sum_m (\mathbf{PS})_{im} (\mathbf{PS})_{mj}, \quad (3)$$

where i , j , and m run over basis function indices, \mathbf{P} is the density matrix, and \mathbf{S} is the overlap matrix. Equations (2) and (3) dictate that charge redistribution effectively takes place only between chemically bonded atoms, and make that redistribution quadratic in bond order. In order to preserve the total charge, the relationships $C_{Z_k Z_{k'}} = -C_{Z_{k'} Z_k}$ and $D_{Z_k Z_{k'}} = -D_{Z_{k'} Z_k}$ are enforced.

Parameterization thus consists of finding optimal values for $C_{Z_k Z_{k'}}$ and $D_{Z_k Z_{k'}}$ for a set of atomic numbers. In the present paper and Ref. 4, Z takes on values 1,6,7,... corresponding to H, C, N, O, F, Si, P, S, Cl, Br, and I, although not all pairwise combinations are considered. As described in more detail previously,⁴ parameters are optimized in a non-linear fashion so as to minimize the root-mean-square (RMS) deviation between experimental molecular dipole moments and those computed from the CM2 partial atomic charges after fixing C–H parameters based on an analysis of the quadrupole moment of benzene and the C and H partial charges in a set of prototype aromatic molecules. The parameterization set consists of 198 polar molecules spanning a wide range of molecular functional groups.²³ Molecular geometries are optimized at the HF/MIDI! level; the MIDI! basis set²⁷ was designed in part to permit reliable prediction of geometries at the HF level. The MIDI! basis is defined as the MIDI basis to which specially optimized polarization functions are added for all atoms except H and C.²⁷ The MIDI basis set is in turn defined as a mid-sized split-valence basis set by Tatewaki and Huzinaga.²⁸

TABLE I. CM2 parameters (a.u.) for HF/cc-pVDZ calculations.

Parameter type	$Z_k, Z_{k'}$	Parameter value	
$C_{Z_k Z_{k'}}$	H–C	–0.050	
	H–N	0.063	
	H–O	0.083	
	H–Si	–0.084	
	H–S	0.117	
	C–N	0.010	
	C–O	0.041	
	C–Si	0.052	
	C–S	–0.063	
	N–O	–0.085	
	$D_{Z_k Z_{k'}}$	H–P	0.073
		C–N	0.046
		C–O	–0.042
		C–F	–0.019
C–P		0.040	
C–S		0.229	
C–Cl		0.121	
C–Br		0.260	
C–I		0.320	
O–P		0.165	
F–P		0.282	
N–O	–0.082		
O–S	0.186		
P–S	0.000		

The cc-pVDZ basis set was employed with d subshells consisting of five functions, consistent with the manner in which it was originally defined.^{21,22} The cc-pVDZ basis set was not available for Br and I, so we substituted the MIDI! basis set²⁸ for these elements.

The present paper also includes semiempirical AM1 and PM3 calculations for comparison.

Semiempirical calculations were carried out using AMSOL-version 6.5.3²⁹ and GAMESS-version 06JAN98³⁰ as extended by GAMESOL-version 2.0.³¹ HF calculations were carried out using GAUSSIAN94³² as extended by MN-GSM98.2.3.³³

III. RESULTS AND DISCUSSION

Table I lists the parameters defining this new CM2 model. Note that certain parameters $C_{Z_k Z_{k'}}$ do not have corresponding parameters $D_{Z_k Z_{k'}}$ and vice versa. Values for the missing parameters were found to be statistically indistinguishable from zero.

Note also that not all possible atom–atom combinations are present in the training set. Thus, for instance, there are no parameters for $Z_k, Z_{k'} = \text{N, F}$, and this is also the case, with one exception, for all other combinations that fail to appear anywhere in Table I. The exception is $Z_k, Z_{k'} = \text{P, S}$; while there *are* P–S bonds in the training set, the mapping does not require any charge redistribution between these two atoms—a value of 0.000 is recorded in Table I for D_{PS} to emphasize that P–S bonds were considered in the parameterization.

Experimental and computed CM2/HF/cc-pVDZ dipole moments for the complete training set are provided as supplementary material. Results for various classes of functionalization are provided in Table II. This table illustrates

TABLE II. RMS errors (D) for dipole moments of polar compounds in the training set organized by functional group.^a

Type of compound	No.	RMS error			
		Mulliken	Löwdin	Density	CM2
H,C,N,O compounds					
Alcohols, water	12	0.40	0.30	0.26	0.18
Esters, lactones	8	0.34	0.26	0.32	0.17
Aldehydes, ketones	16	0.44	0.39	0.33	0.18
Acids	9	0.69	0.31	0.36	0.14
Ethers	10	0.34	0.29	0.18	0.19
Amines, ammonia	13	0.51	0.25	0.17	0.17
Nitriles, HCN	17	0.74	1.12	0.28	0.17
Amides	3	1.01	0.40	0.28	0.16
Imines, N-aromatics	7	0.84	0.65	0.26	0.30
Multifunctional N	7	1.08	0.97	0.18	0.24
subtotal	102	0.63	0.61	0.27	0.19
Other polar compounds					
Fluorides	31	0.99	0.51	0.19	0.17
Chlorides	22	0.81	0.80	0.33	0.12
Bromides	10	0.31	1.75	0.09	0.17
Iodides	5	1.08	1.27	0.10	0.24
Silicon compounds	6	0.56	0.12	0.08	0.10
Sulfur compounds	9	0.49	1.20	0.33	0.16
Phosphorus compounds	13	1.38	1.43	0.43	0.18
All polar compounds	198	0.78	0.73	0.28	0.18

^aAll results in this table are based on HF/cc-pVDZ calculations at HF/MIDI! geometries. Errors are with respect to experiment.

how robust the CM2/HF/cc-pVDZ model is. In particular, it exhibits an RMS error that is at least a factor of 1.5 lower than that obtained by Class II Mulliken or Löwdin analysis for all 17 classes of compounds. Furthermore the overall RMS error (0.18 D) over all 198 polar compounds is a factor of 4.1 lower than obtained by Löwdin analysis and a factor of 4.3 better than obtained by Mulliken analysis. It is also a factor of 1.6 lower than is obtained from application of the dipole moment operator to the HF/cc-pVDZ wave function (0.28 D, column labeled “Density” in Table II). This new CM2 model is very similar in accuracy to previous CM2 models—on the same test set, RMS errors for eight other CM2 models based on *ab initio* and density functional methods using basis sets that like cc-pVDZ do not contain diffuse functions are between 0.17 and 0.21 D.⁵

The accuracy of the CM2 model is not particularly dependent on molecular functionality; the smallest RMS error is 0.10 D for silicon compounds, and the largest RMS error is 0.30 D for the class of compounds consisting of imines and nitrogen-containing aromatics, i.e., a range of only 0.2 D in RMS error over disparate functionalities. By comparison, the ranges over functionality of density-, Mulliken-, and Löwdin-derived dipole moment RMS errors are 0.35, 1.07, and 1.63 D, respectively. Note that the CM2 mapping particularly corrects for the rather large errors in dipole moments exhibited by Class II charges for multifunctional nitrogen-containing molecules, and for molecules containing halogen atoms, sulfur, or phosphorus.

The decrease of the RMS error from 0.28 D in the electrostatic-potential-fitted results to 0.18 D in the CM2 model is small (0.10 D) in absolute terms but large (36%) in relative terms. It is more significant, however, to consider the

improvement as compared to Löwdin charges. Table II shows that this is large in *both* absolute (0.55 D) and relative (75%) terms. This comparison is more apt because electrostatic-potential-fitted charges do not lend themselves readily to use in dynamics and are unstable for buried charges, whereas CM2 charges retain three major advantages of Löwdin charges, namely low cost, easily coded analytic gradients, and equal applicability to atoms in the interior of a molecule or on its perimeter.

We note that our goal is not merely the prediction of dipole moments, but rather the prediction of whole charge distributions. We use the dipole moment, which is only one component of the multipole series, just as a convenient way to compare theory to experiment.

IV. CM2 ANALYSIS OF NUCLEIC ACID BASES AND METHYL AZIDE

One test of the quality of any parameterization is its performance on a test set of molecules not used in the parameterization process. To test the new CM2 model, we chose a particularly demanding test set comprised of eight nucleic acid bases and methyl azide. The nucleic acid bases contain several functional groups each, and, moreover, this test set includes nitrogen atoms having a variety of hybridizations. Methyl azide uniquely contains nitrogen–nitrogen bonds.

The nucleic acid bases were chosen because of the availability of high-quality geometries and dipole moments. Johnson *et al.*³⁴ have optimized the geometries of the bases with Møller–Plesset second-order perturbation theory⁶ (MP2) and the 6-31G(d,p) basis set and computed dipole

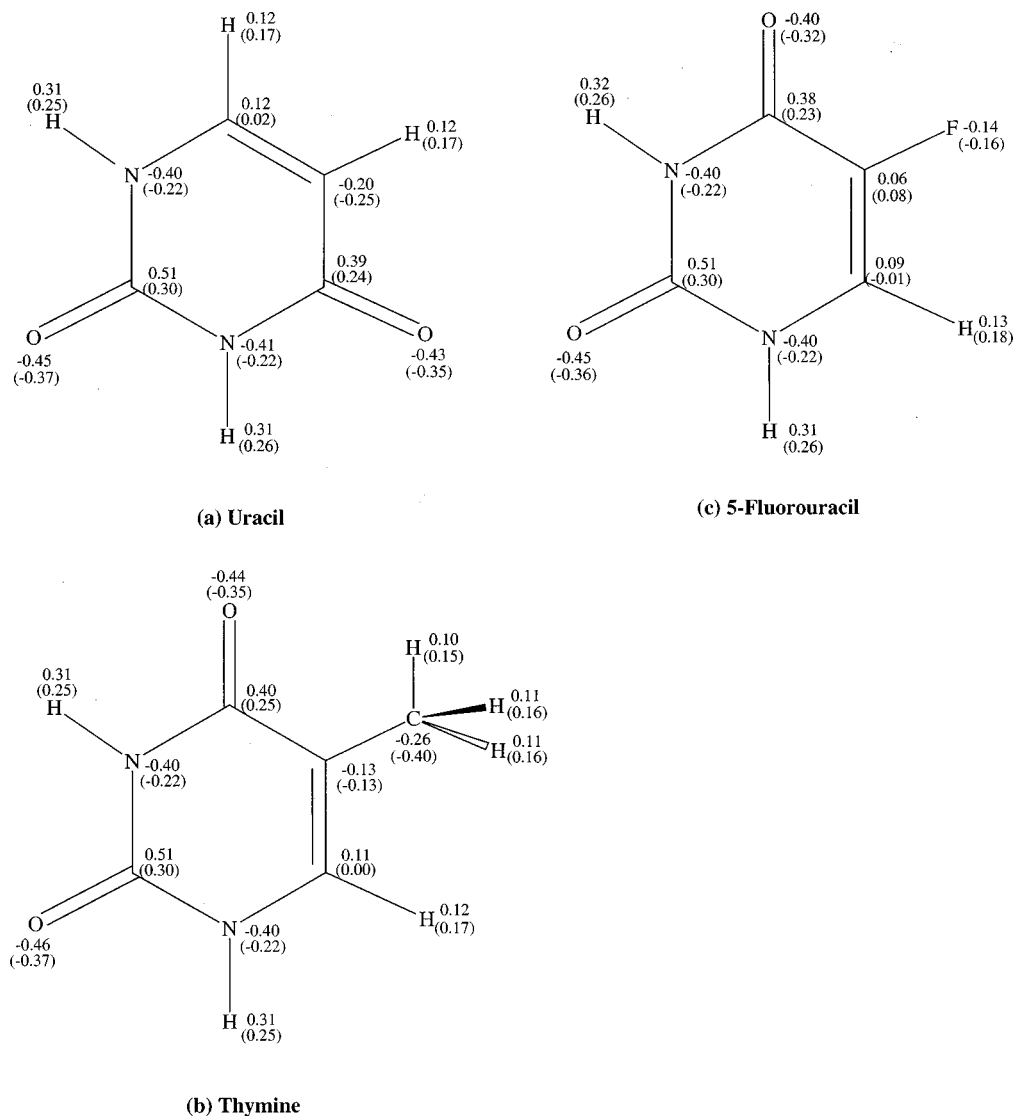


FIG. 1. CM2/HF/cc-pVDZ charges and antecedent HF/cc-pVDZ Löwdin charges (in parentheses) of (a) uracil, (b) thymine, (c) 5-fluorouracil, (d) cytosine, (e) 5-methylcytosine, (f) adenine, (g) hypoxanthine, (h) guanine, and (i) methyl azide.

moments for these molecules at the MP2 level using a $[5s3p2d]/[3s2p]$ basis set; we have also calculated a presumably even more accurate set of dipole moments at the B3LYP/cc-pVTZ level (B3LYP is a hybrid level of DFT and HF^{14–17}), a level shown by Kwiatkowski and Leszczynski³⁵ to give very similar results to MP2 for several tautomers of cytosine and some heterosubstituted cytosine analogs. We have employed those same geometries to explore the predictive accuracy of a number of different CM2 models, and also the CM1/AM1 and CM1/PM3 models. For methyl azide, the HF/MIDI! geometry was used.

In addition to comparing the quality of CM2 to CM1 predictions, we test whether *ab initio* CM2 models perform better than semiempirical CM2 models. One might expect *ab initio* Hartree–Fock methods using polarized basis sets to give more reliable charge distributions than semiempirical molecular orbital theory, particularly for cases where the minimal basis sets employed by the semiempirical methods are insufficiently flexible for the molecule in question, as might be the case for highly functionalized molecules.

The test set molecules themselves, together with their computed HF/cc-pVDZ Löwdin and CM2 charges, are provided in Fig. 1. Several trends are apparent in the charge mappings. For instance, C–H polarities tend to be reduced on mapping. For example, in uracil [Fig. 1(a)] the charge separations across the C5–H and C6–H bonds are reduced from 0.42 and 0.15 at the Löwdin/HF/cc-pVDZ level to 0.32 and 0.00 at the CM2/HF/cc-pVDZ level. This same trend is seen in the methyl groups of 5-methylcytosine and methyl azide [Figs. 1(e) and 1(h)]. In some instances, the bond polarity is even reversed, as is seen for cytosine [Fig. 1(d)] where the charge separation across the C6–H bond goes from carbon being less positive by 0.12 units with Löwdin charges to carbon being more positive by 0.04 units with CM2 charges. These trends in C–H polarities can, however, be outweighed by larger charge redistributions between the carbon atom and attached heteroatoms, so that the C–H bond polarity becomes larger in magnitude. Thus, for instance, the CM2 charges on hydrogen at C8 in the purines [Figs. 1(f)–1(h)] are always smaller than the antecedent Löwdin charges

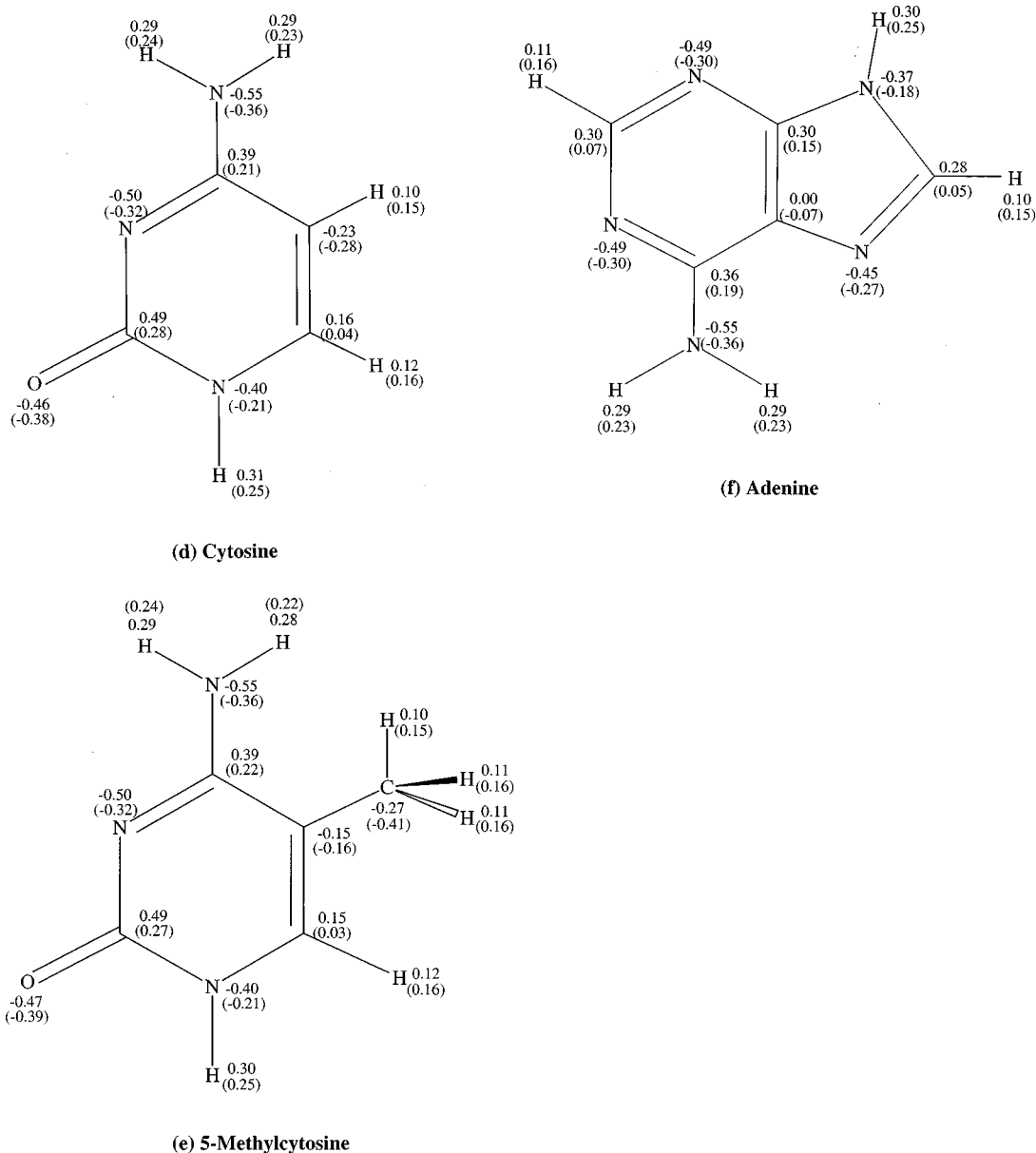


FIG. 1. (Continued.)

by 0.05 units, but the change in the C8 charge is a much larger 0.22 or 0.23 units because of charge redistribution across the two C–N bonds.

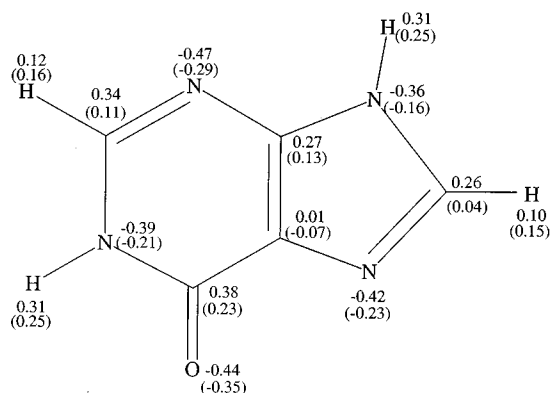
In general, CM2 heteroatom charges are larger than antecedent Löwdin charges, and this increase in magnitude is quite uniform, at least in the heterocycles. For example, all nitrogen atoms in the nucleic acid bases become more negative by 0.18 or 0.19 units, and all oxygen atoms become more negative by from 0.07 to 0.09 units.

Table III details the relative performance of the different models taking the B3LYP/cc-pVTZ results as the standard. The highest accuracy is provided by the CM2/HF/6-31G* model, which has an RMS error only 0.03 D larger than that for the MP2/[5s3p2d]/[3s2p] level, where the dipole moments are determined as expectation values of the dipole moment operator. Thus, the accuracy of this CM2 model is basically within the range of expected theoretical accuracy from different high-quality methods. Importantly, the CM2

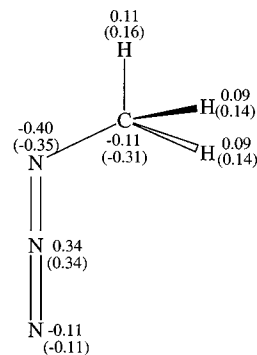
calculations for these molecules, which rely on Hartree–Fock densities for the original population analysis, take only a small fraction of the time required to compute the MP2 densities.

For the cc-pVDZ basis set, the RMS error of dipole moments determined directly from the HF Löwdin charges is only 0.33 D. This suggests that the charge mapping does not need to make a large difference in the charges, since there is only small room for improvement. This is consistent with the parameter values found in Table I, all of which are fairly small in magnitude. Nevertheless, the CM2 model does reduce the RMS error by almost a third, and furthermore the mean signed error (MSE) is reduced to 0.04; the Löwdin-charge-derived dipole moments are always underestimated and show a MSE of -0.26 .

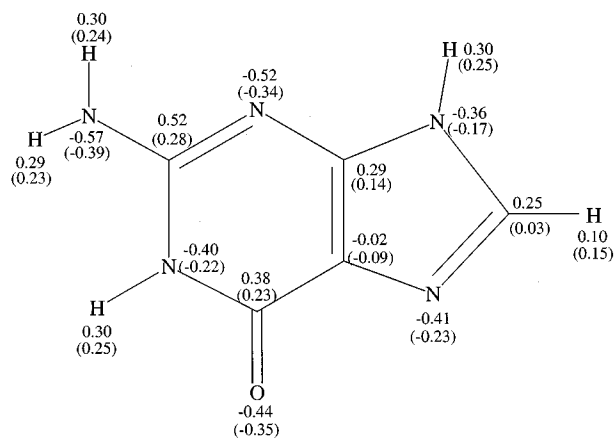
The other two CM2/HF models, using the MIDI! and cc-pVDZ basis sets, perform about equivalently. While it may seem surprising that the smaller MIDI! basis set is com-



(g) Hypoxanthine



(i) Methyl azide



(h) Guanine

FIG. 1. (Continued.)

petitive in this regard, it must be recalled that the MIDI! basis set was designed in part to provide high-quality charge distributions at low cost even prior to mapping. The results in Table III provide further validation of the MIDI! basis set.

Note that methyl azide is not particularly well handled by any of the CM2/HF models. This is a difficult molecule in part because no charge redistribution is possible across homonuclear bonds (the charge normalization constraints re-

TABLE III. Dipole moments (D) of eight nucleic acid bases and methyl azide from different models.^{a,b}

Model	Molecule										MSE ^c	MUE ^c	RMS ^c
	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>	<i>i</i>				
B3LYP/cc-pVTZ	4.47	4.38	4.13	6.41	6.75	2.53	5.28	6.57	2.08				
MP2/[5s3p2d]/[3s2p]	4.34	4.32	3.94	6.32	6.78	2.69	5.13	6.40	2.16	-0.06	0.12	0.13	
Löwdin/HF/cc-pVDZ	4.47	4.42	4.02	6.09	6.50	1.98	5.18	6.09	1.51	-0.26	0.26	0.33	
CM2/HF/cc-pVDZ	4.63	4.58	4.12	6.66	7.04	2.34	5.48	6.50	1.64	0.04	0.20	0.23	
CM2/HF/MIDI!6D	4.49	4.40	3.94	6.46	6.77	2.24	5.08	6.04	1.76	-0.16	0.18	0.25	
CM2/HF/6-31G*	4.52	4.47	4.07	6.58	6.95	2.33	5.40	6.48	1.79	0.00	0.14	0.16	
CM2/AM1	4.59	4.62	3.99	6.43	6.69	2.04	4.87	5.75	1.72	-0.21	0.30	0.38	
CM2/PM3	4.25	4.26	3.79	6.22	6.50	2.13	4.79	5.74	2.16	0.19	0.30	0.38	
CM1/AM1	4.22	4.26	3.56	6.22	6.47	2.34	3.81	4.22	1.98	-0.61	0.61	0.96	
CM1/AM1(opt)	4.43	4.47	3.75	6.49	6.00	2.21	3.91	4.29	2.00	-0.56	0.60	0.94	
CM1/PM3	4.06	4.00	3.78	5.84	6.10	2.12	5.01	6.30	1.85	-0.39	0.39	0.41	
CM1/PM3(opt)	4.03	4.02	3.65	5.79	6.00	2.20	4.91	6.04	2.05	-0.43	0.43	0.47	

^aMolecular geometries of the nucleic acid bases and methyl azide were optimized at the MP2/6-31G(*d,p*) and HF/MIDI! levels, respectively, except for rows marked "opt," for which geometries were re-optimized with AM1 and PM3.

^bSee Fig. 1 for molecules.

^cErrors are with respect to B3LYP/cc-pVTZ.

quire the *C* and *D* parameters to be identically zero for the homonuclear case). Since there is only appreciable bond order between the methyl carbon and the first nitrogen of the azide, it is not possible to adjust the charges of the outer two nitrogen atoms to any significant extent [Fig. 1(i)], which somewhat limits the ability of the charge model to affect the dipole moment.

The two semiempirical CM2 methods, based on AM1 and PM3, respectively, perform with equivalent accuracy over the 9 molecule test set, and exhibit errors about 0.2 D larger than are found for the *ab initio* CM2 models. Thus, more flexible bases do appear to improve the ability of the CM2 model to predict molecular dipole moments for these highly functionalized molecules.

With respect to comparing CM2 models to their corresponding CM1 models, there is a somewhat surprising difference between AM1 and PM3. The CM2/AM1 model significantly outperforms the analogous CM1 model, the latter having an RMS error of just under 1 D. Since the CM1 model was parameterized using AM1 geometries, we examined whether relaxing the molecular geometries at this level made any difference. While there are sometimes substantial changes in individual dipole moments on relaxation (e.g., 0.47 D in the case of 5-methylcytosine), the overall RMS error is only reduced by 0.02 D. The CM1/PM3 model, on the other hand, is only very slightly less accurate than the CM2/PM3 model, this accuracy being somewhat reduced when PM3 geometries are used. This is remarkable, as noted previously,²³ because it is well known that Class II partial atomic charges for PM3 nitrogen atoms are often very bad indeed.^{1,12,36} Apparently both charge models, CM1 and CM2, are capable of correcting for this deficiency in a systematic way. One additional item meriting note is that errors in the CM1 models are always such that the molecular dipole moments in the test set are underestimated, i.e., the mean signed error (MSE) is negative and equal in magnitude to the mean unsigned error (MUE). This is not the case for any of the CM2 models, although for the CM2/HF/MIDI! model the largest errors are underestimations.

The atomic charges in Fig. 1 may be compared to other sets of charges published elsewhere,^{23,37–47} including those from experiment,^{23,37,41} from CM1/AM1,^{23,37} from molecular mechanics^{37,40,42–47} (including the force fields in AMBER^{40,45–47}), from *ab initio* gas-phase calculations,³⁷ from self-consistent reaction-field calculations,^{37–39} and from CM2 with other wave functions.^{5,38} In general the agreement is very good.

V. CONCLUDING REMARKS

We have parameterized a CM2 charge model for HF/cc-pVDZ calculations.⁴⁸ The resulting RMS error over a training set of 198 polar molecules is 0.18 D. Application of this model to a test set of 8 nucleic acid bases and methyl azide indicates that the model is robust for biologically relevant nitrogen-containing compounds not found in the training set. With semiempirical wave functions, CM2 results are better than CM1 results for AM1 calculations on the test set, and

the two models are about equivalent for PM3 calculations. Furthermore, CM2 results for the test set are found to be more accurate when mapping Löwdin charges from the HF level than when mapping them from semiempirical wave functions.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation, the Army Research Office, and the Alfred P. Sloan Foundation, and it made use of computing resources of the Minnesota Supercomputer Institute and the University of Minnesota-IBM Shared Research Project.

- ¹ S. M. Bachrach, in *Reviews in Computational Chemistry*, edited by K. B. Lipkowitz and D. B. Boyd (VCH, New York, 1993), Vol. 5, p. 171.
- ² R. S. Mulliken, *J. Chem. Phys.* **3**, 564 (1935).
- ³ R. S. Mulliken, *J. Chem. Phys.* **23**, 1833 (1955).
- ⁴ P.-O. Löwdin, *J. Chem. Phys.* **18**, 365 (1950).
- ⁵ J. Li, T. Zhu, C. J. Cramer, and D. G. Truhlar, *J. Phys. Chem. A* **102**, 1820 (1998).
- ⁶ J. Simons and J. Nichols, *Quantum Mechanics in Chemistry* (Oxford University Press, New York, 1997).
- ⁷ M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart, *J. Am. Chem. Soc.* **107**, 3902 (1985).
- ⁸ M. J. S. Dewar and E. G. Zoebisch, *J. Mol. Struct.: THEOCHEM* **180**, 1 (1988).
- ⁹ M. J. S. Dewar and C. Jie, *Organometallics* **6**, 1486 (1989).
- ¹⁰ M. J. S. Dewar and C. Jie, *J. Mol. Struct.: THEOCHEM* **187**, 1 (1989).
- ¹¹ M. S. J. Dewar and Y. Yate-Ching, *Inorg. Chem.* **29**, 3881 (1990).
- ¹² J. J. P. Stewart, *J. Comput. Chem.* **10**, 209 (1989).
- ¹³ R. G. Parr and W. Yang, *Annu. Rev. Phys. Chem.* **46**, 701 (1995).
- ¹⁴ C. Lee, W. Yang, and R. G. Parr, *Phys. Rev. B* **37**, 785 (1988).
- ¹⁵ A. D. Becke, *Phys. Rev. A* **38**, 3098 (1988).
- ¹⁶ A. D. Becke, *J. Chem. Phys.* **98**, 5648 (1993).
- ¹⁷ P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, *J. Phys. Chem.* **98**, 11623 (1994).
- ¹⁸ J. Li, B. Williams, C. J. Cramer, and D. G. Truhlar, *J. Chem. Phys.* **110**, 724 (1999).
- ¹⁹ J. E. Ridley and M. C. Zerner, *Theor. Chim. Acta* **32**, 111 (1973).
- ²⁰ M. C. Zerner, in *Reviews in Computational Chemistry*, edited by K. B. Lipkowitz and D. B. Boyd (VCH, New York, 1990), Vol. 2, p. 313.
- ²¹ T. H. Dunning, *J. Chem. Phys.* **90**, 1007 (1989).
- ²² D. E. Woon and T. H. Dunning, *J. Chem. Phys.* **98**, 1358 (1993).
- ²³ J. W. Storer, D. J. Giesen, C. J. Cramer, and D. G. Truhlar, *J. Comput.-Aided Mol. Des.* **9**, 87 (1995).
- ²⁴ C. A. Coulson and H. C. Longuet-Higgins, *Proc. R. Soc. London* **191**, 39 (1947).
- ²⁵ I. Mayer, *Chem. Phys. Lett.* **97**, 270 (1983).
- ²⁶ D. R. Armstrong, P. G. Perkins, and J. J. P. Stewart, *J. Chem. Soc., Dalton Trans.* **1973**, 838 (1973).
- ²⁷ (a) R. E. Easton, D. J. Giesen, A. Welch, C. J. Cramer, and D. G. Truhlar, *Theor. Chim. Acta* **93**, 281 (1996). (b) J. Li, C. J. Cramer, and D. G. Truhlar, *Theor. Chem. Acc.* **99**, 192 (1998).
- ²⁸ (a) H. Tatewaki and S. Huzinaga, *J. Comput. Chem.* **1**, 205 (1980). (b) S. Huzinaga, J. Andzelm, M. Klobukowski, E. Radzio-Andzelm, Y. Sakai, and H. Tatewaki, in *Gaussian Basis Sets for Molecular Calculations*, edited by S. Huzinaga (Elsevier, Amsterdam, 1984).
- ²⁹ AMSOL-version 6.5.3, G. D. Hawkins, D. J. Giesen, G. C. Lynch, C. C. Chambers, I. Rossi, J. W. Storer, J. Li, T. Zhu, D. Rinaldi, D. A. Liotard, C. J. Cramer, and D. G. Truhlar, University of Minnesota, Minneapolis, 1998, based on AMPAC-version 2.1, D. A. Liotard, E. F. Healy, J. M. Ruiz, and M. J. S. Dewar.
- ³⁰ M. W. Schmidt, K. K. Baldrige, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, *J. Comput. Chem.* **14**, 1347 (1993).
- ³¹ GAMESOL-version 2.0, J. Li, T. Zhu, G. D. Hawkins, D. A. Liotard, D. Rinaldi, C. J. Cramer, and D. G. Truhlar, University of Minnesota, Minneapolis, 1998. This solvation module is based on GAMESS (Ref. 30).
- ³² GAUSSIAN94-RevE. 2, M. J. Frisch, G. W. Trucks, H. B. Schlegel, P. M. W.

- Gill, B. G. Johnson, M. A. Robb, J. R. Cheeseman, T. A. Keith, G. A. Petersson, J. A. Montgomery, K. Raghavachari, M. A. Al-Laham, V. G. Zakrzewski, J. V. Ortiz, J. B. Foresman, C. Y. Peng, P. A. Ayala, M. W. Wong, J. L. Andres, E. S. Replogle, R. Gomperts, R. L. Martin, D. J. Fox, J. S. Binkley, D. J. Defrees, J. Baker, J. P. Stewart, M. Head-Gordon, C. Gonzalez, and J. A. Pople, Gaussian, Inc., Pittsburgh, PA, 1995.
- ³³MN-GSM98.2.3, J. Li, T. Zhu, G. D. Hawkins, D. A. Liotard, C. J. Cramer, and D. G. Truhlar, University of Minnesota, Minneapolis, 1998. This solvation module is based on GAUSSIAN (Ref. 32).
- ³⁴R. C. Johnson, T. D. Power, J. S. Holt, B. Immaraporn, J. E. Monat, A. A. Sissoko, M. M. Yanik, A. V. Zagorodny, and S. M. Cybulski, *J. Phys. Chem.* **100**, 18875 (1996).
- ³⁵J. S. Kwiatkowski and J. Leszczynski, *J. Phys. Chem.* **100**, 941 (1996).
- ³⁶C. J. Cramer and D. G. Truhlar, *J. Comput.-Aided Mol. Des.* **6**, 629 (1992).
- ³⁷G. D. Hawkins, C. J. Cramer, and D. G. Truhlar, *J. Chim. Phys. Phys.-Chim. Biol.* **94**, 1448 (1997).
- ³⁸J. Li, C. J. Cramer, and D. G. Truhlar, *Biophys. Chem.* **78**, 147 (1999).
- ³⁹M. Orozco and F. J. Luque, *Biopolymers* **33**, 1851 (1993).
- ⁴⁰J. E. Esterowicz, J. L. Miller, and P. A. Kollman, *J. Phys. Chem. B* **101**, 10971 (1997).
- ⁴¹D. A. Pearlman and S.-H. Kim, *J. Mol. Biol.* **211**, 171 (1990).
- ⁴²J. Pranata, S. G. Wierschke, and W. L. Jorgensen, *J. Am. Chem. Soc.* **113**, 2810 (1991).
- ⁴³W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, *J. Am. Chem. Soc.* **118**, 11225 (1996).
- ⁴⁴A. D. Mackerell, J. Wiórkiewicz, and M. Karplus, *J. Am. Chem. Soc.* **117**, 11946 (1995).
- ⁴⁵S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case, *J. Comput. Chem.* **7**, 230 (1986).
- ⁴⁶C. I. Bayly, P. Cieplak, W. D. Cornell, and P. A. Kollman, *J. Phys. Chem.* **97**, 269 (1993).
- ⁴⁷W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmayer, T. Fox, J. W. Caldwell, and P. A. Kollman, *J. Am. Chem. Soc.* **117**, 5179 (1995).
- ⁴⁸See AIP Document No. PAPS JCPSA6-111-301925 for seven pages of tables. These contain experimental and CMZ/cc-pVDZ dipole moments for the molecules in the database. Order by PAPS number and journal reference from American Institute of Physics, Physics Auxiliary Publication Service, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502. Fax: 516-576-2223, e-mail: paps@aip.org. The price is \$1.50 for each microfiche (98 pages) or \$5.00 for photocopies of up to 30 pages, and \$0.15 for each additional page over 30 pages. Make checks payable to the American Institute of Physics.