



Published in final edited form as:

J Comput Chem. 2011 July 30; 32(10): 2127–2139. doi:10.1002/jcc.21795.

Dipole Preserving and Polarization Consistent Charges

Peng Zhang, Peng Bao, and Jiali Gao*

Department of Chemistry, Digital Technology Center and Supercomputing Institute, University of Minnesota, Minneapolis, MN 55455

Abstract

A method for estimating dipole preserving and polarization consistent (DPPC) charges is described, which reproduces exactly the molecular dipole moment as well as the local, atomic hybridization dipoles determined from the corresponding wave function, and can yield accurate molecular polarization. The method is based on a model described by Thole and van Duijnen and a new feature is introduced in the DPPC model to treat molecular polarization. Thus, the DPPC method offers a convenient procedure to describe molecular polarization in applications using semiempirical models and ab initio molecular orbital theory with relatively small basis functions such as 6-31+G(d,p) or without inclusion of electron correlation; these methods tend to underestimate molecular polarizability. The trends of the DPPC partial atomic charges are found to be in good accord with those of the CM2 model, a class IV charge analysis method that has been used in a variety of applications. The DPPC method is illustrated to mimic the correct molecular polarizability in a water dimer test case and in water-halide ion complexes using the explicit polarization (X-Pol) potential with the AM1 Hamiltonian.

1. Introduction

Partial atomic charges are not physical observables that can be measured experimentally,¹ but the concept is extremely useful in our qualitative understanding of structure and reactivity and in sophisticated applications such as biomolecular simulations. In practice, they are obtained according to the specific need and there are many ways of deriving partial charges.² In this article, we describe a method to produce atomic partial charges that can be used in combined quantum mechanical and molecular mechanical (QM/MM)³ or mixed QM/QM potentials,⁴⁻⁵ or in the explicit polarization (X-Pol)⁶⁻¹⁰ quantal force field for biomolecular simulations.¹¹ In this article, we make use of semiempirical methods that are computationally efficient and capable of yielding excellent molecular dipole moments,¹²⁻¹⁴ although we emphasize that any quantum mechanical models can be employed. Storer et al.¹⁵ classified the way by which partial charges are derived into four categories, including (i) extraction from experimental data, (ii) population analysis of the molecular wave function,¹⁶⁻¹⁷ (iii) property fitting¹⁸⁻²⁰ such as the electrostatic potential obtained from quantum mechanical calculations²¹⁻²⁴ or from experiments,¹ and (iv) semiempirical

gao@jialigao.org.

Supporting Information Available: Partial atomic charges for H, C, N and O containing compounds derived using the DPPC charge analysis and the CM2 model with the AM1 Hamiltonian. This material is available free of charge via the Internet at <http://pubs.acs.org>.

mapping that transforms the population charges to best reproduce the experimental dipole moments.^{15,25} Each category has its own strengths and weaknesses and they have been lucidly discussed in the past.^{2,15,26} One particular approach is the Mulliken population analysis¹⁷ which is often criticized for its lack of consistency, especially using large basis sets. Nevertheless, the Mulliken population analysis enjoys its simplicity, which can provide an indication of the relative charge separation in a molecular system.²⁵ Mulliken population charges have been used as the reference point in the class IV charge models,^{15,25,27} which has been adopted in combined QM/MM simulations.²⁸

Thole and van Duijnen¹⁹ described a population analysis that preserves the dipole moment of the molecule by distributing partial atomic charges with the constraint that the molecular dipole moment is identical to that determined from the molecular wave function. The original method of Thole and van Duijnen only utilizes molecular geometry about which the charge reassignment was made;¹⁹ however, as the complexity of a molecule increases, it is difficult to choose a proper weighting function to take into account the properties of functional groups. Subsequently, Swart et al.²⁹ extended that approach to preserve atomic multipole moments that are derived from atomic charge densities using density functional theory. Although the latter approach is very appealing especially in calculations using high quality electron densities with augmented basis functions, it is less useful in situations when relatively small basis functions are used such as 6-31+G(dp) or the minimal basis in semiempirical models.³⁰⁻³¹ In the later case, Mulliken population analysis in fact provides an adequate starting point, as shown by Truhlar and Cramer et al. in their development of class IV charge models^{15,25} to map partial charges that can yield molecular dipole moments in good agreement with experiments. This strategy is particularly attractive for applications using semiempirical methods such as the Austin model 1 (AM1)¹² or the recently developed parameterization model 6 (PM6)¹³ because they have been parameterized to reproduce experimental dipole moments.

Our goal is to use partial atomic charges to represent the molecular electrostatic moments of the molecular wave function of individual fragments in solution or in a macromolecular system in combined QM/MM,³ multilayer QM/QM,⁴⁻⁵ and the X-Pol quantal force field.^{7,11} The procedure described by Thole and van Duijnen¹⁹ and later revised by Swart et al.^{19,29} meet this criterion. In this article, we present a method for deriving partial atomic charges from semiempirical Hamiltonians (or ab initio methods) by preserving the molecular dipole moment determined from the corresponding wave function exactly. Rather than redistributing partial atomic charges completely as was done in ref 29, we use the Mulliken charge as a reference state, and treat only the hybridization components in the dipole calculation for charge redistribution based on atomic electronegativity. The method is applicable both to neutral molecules and charged systems, independent of the origin of the coordinate system. Significantly, not only the qualitative features of the Mulliken population analysis are retained in this procedure, but also the local hybridization dipole components as well as the total molecular dipole moment from the wave function is preserved.¹²⁻¹³

In addition, we incorporate Thole's interaction dipole (TID)³² polarization method into the dipole preserving scheme to ensure that the generally weakly polarized semiempirical wave function can also adequately represent molecular polarization. Together, the present method

preserves molecular dipole moment from the wave function and ensures consistent molecular polarization; the partial atomic charges derived from semiempirical Hamiltonians can be suitable for dynamics simulations of macromolecular systems and modeling of protein-ligand interactions using multilayer QM/QM potentials⁴⁻⁵ or a fully quantal (X-Pol) force field.^{6,8,11}

In this paper we first summarize the dipole-preserving algorithm employed in our implementation, which is a modified approach used by Swart et al.²⁹ Then, we describe the incorporation of a molecular polarization consistent procedure, which is a new contribution of this work. We call this approach the dipole preserving and polarization consistent (DPPC) charge analysis. We choose a set of molecules that were used in the work of Storer et al.¹⁵ in the development of the CM1A charges based on the AM1 Hamiltonian to illustrate the performance and qualitative trends of the DPPC charges in the gas phase because both methods are developed with the aim to produce molecular dipole moments in good agreement with experiment. The Class IV charge models of Cramer and Truhlar are derived based on empirical fit to experimental dipole moments of a database, whereas the present DPPC method is an analytical procedure without empirical parameters except a weighting function for charge distribution. Excluding nitrile compounds, the semiempirical AM1, PM3, and PM6 models, and the corresponding DPPC charges, can yield molecular dipole moments in excellent agreement with experiments with standard errors of about 0.3 D for this database, which is comparable or smaller than the standard errors of the CM1A method (or the latter CM2 model). In a subsequent study of condensed phase systems, including liquids and solutions, the ability to recover correct molecular polarization will be tested in statistical mechanical Monte Carlo and molecular dynamics simulations, employing the explicit polarization (X-Pol)^{8,11} quantal force field.⁷

2. Method

2.1. Background

The Mulliken population charge of atom k is partitioned from the density matrix \mathbf{P} as

$$q_k = Z_k - \sum_{\mu} P_{\mu\mu} S_{\mu k} \quad (1)$$

where Z_k is the nuclear charge, \mathbf{P} is the one-electron density matrix and \mathbf{S} is the overlap matrix of the basis functions. The superscript “o” in eq 1 is to indicate that the Mulliken population charges will be used as the reference state for deriving the DPPC charges. Although the method is general, here, we focus on semiempirical methods based on the NDDO approximation,³⁰ in which the quantum mechanical (QM) molecular dipole moment is determined as a sum of two contributing terms:³³

$$\mathbf{D}_{MP} = \mathbf{D}_{hyb} + \mathbf{D}_{MP} \quad (1)$$

where \mathbf{D}_{MP} is the contribution from Mulliken charges, and \mathbf{D}_{hyb} is the one-center s and p hybridization contribution. The Mulliken population (MP) dipole is simply:



$$\mathbf{P}_{sp}^k = \begin{pmatrix} P_s & & \\ & P_p & \\ & & \dots \end{pmatrix} \quad (2)$$

where N is the number of atoms, and $\{\mathbf{r}_k; k = 1, \dots, N\}$ are the atomic positions. The s - p hybridization component is written as:



$$\mathbf{D}_{MP} = \sum_k \mathbf{R}^k \mathbf{P}_{sp}^k \quad (3)$$

where $(\mathbf{P}_{sp})^k$ is a diagonal matrix with the densities of the s and p orbitals on atom k , and \mathbf{R}^k is the corresponding dipole integral. It is well-known that the one-center hybridization polarization terms make substantial contributions to the total molecular dipole moment.³³ As a result, \mathbf{D}_{MP} is not expected to yield good agreement with experimental dipole.

Our goal is to incorporate additionally distributed charges, called residual charges, to exactly reproduce the hybridization contribution in eq 3 such that the sum of the residual and Mulliken population charges will preserve the local, atomic (hybridization) point dipoles and the total molecular dipole moment from the full density of the molecular wave function. In our approach, the hybrid polarization dipole is defined as the optimization target in the residual charge distributing procedure:¹⁹



$$\mathbf{D} = \sum_k \mathbf{R}^k \mathbf{P}_{sp}^k + \mathbf{D}_{res} \quad (4)$$

It is important to emphasize that the net molecular charge is represented by the Mulliken population charges, while the hybridization term \mathbf{D}_{hyb} has no net charge contributions. Consequently, \mathbf{D} in eq 4 is translationally and rotationally invariant both for neutral and ionic systems such that the charge preserving method describe in this paper is applicable both to neutral molecules and to ions. As shown in eq 3, the total polarization dipole can be written as a sum of atomic contributions; thus, the optimization target can also be expressed in terms of atomic components:



$$\mathbf{D} = \sum_k \mathbf{R}^k \mathbf{P}_{sp}^k + \mathbf{D}_{res} \quad (5)$$

Of course, there are other ways of decomposing the total molecular dipole, or generally, multipole moments, into atomic components. For example, the atomic multipole moments can be obtained from atomic electron densities fitted to reproduce the total molecular electron density (either from the wave function in a QM calculation or from X-ray diffraction).¹ Typically, an auxiliary basis of Slater orbitals is used to represent the atomic density (as in DFT calculations), from which atomic multipole moments are derived. Although this method is very appealing, a very large basis set must be used to obtain a good

fit to the molecular electron density, which is not particularly useful for semiempirical models employing a minimal basis. Further, the fitted auxiliary densities generally do not reproduce the molecular dipole and multipole moments computed directly from the electronic structure method. Consequently, additional auxiliary functions and constraints must be imposed.²⁹

The density fitting procedure was used in the multipole derived charge introduced by Swart et al.,²⁹ who described an algorithm that preserves atomic multipole moments at any order. The approach employed in our study is similar to that method,^{19,29} but the procedure and details are different. In the method of Swart et al., the atomic charges are obtained by fully redistributing the residual charges to all atomic centers with the constraint that the atomic or molecular moments are reproduced. In the present study, the Mulliken population charges are kept as the zeroth order reference, from which the error in the Mulliken population charge, \mathbf{D}_i relative to the exact quantum mechanical moment is defined as the fitting target. Thus, the redistributed charges may be regarded as perturbations to the Mulliken population charge. In the following, we limit our discussion to dipole preservation, although the procedure can be analogously applied to moments at any order.²⁹

2.2. Locally distributed dipole-preserving charges (DPC)

For each atomic (i.e., s - p hybridization polarization) dipole to be preserved, \mathbf{D}_i where $i = 1, \dots, N$, we wish to assign a set of atomic charges, called *residual charges*, distributed to atomic centers in the molecule such that they reproduce the target hybridization dipole exactly:

$$\sum_k q_k \mathbf{r}_{ki} = \mathbf{D}_i \quad (6)$$

where q_k is the residual atomic charge on atom k , associated with the atomic dipole \mathbf{D}_i located on atom i . We seek to obtain a set of residual charges q_k that make the smallest perturbation to the Mulliken population charges under the constraints that the net change is zero, thereby preserving the net molecular charge.

$$\sum_k q_k = 0 \quad (7)$$

The total atomic charge of atom k is thus the sum of the Mulliken population charge q_k^M and the residual charges due to the preservation of all atomic hybridization dipole moments:

$$q_k = q_k^M + \sum_i q_{ki} \quad (8)$$

The residual charges associated with the dipole moment $\mathbf{D}_i (i = 1, \dots, N)$, are obtained by using the Lagrange multiplier method, in which we minimize the charge variation subject to the constraints of eqs 6 and 7, *i.e.*, that the residual dipole, or the error in the Mulliken population dipole is exactly corrected and the total charge variation is zero. Thus, for atom i ,

$$\sum_j \alpha_j \mathbf{D}_j + \beta \sum_k \gamma_{ik} \mathbf{D}_k = \mathbf{D}_i \quad (9)$$

where α and β are the Lagrangian multipliers, and γ_{ik} is a set of weighting functions (*vide infra*) that controls the way that the charge redistribution is accomplished.

The residual charges α_j should be dominantly distributed to sites closest to the dipole \mathbf{D}_i on atom i that they aim to reproduce. In addition, it is important to be proportionally distributed according to the relative electron-withdrawing ability of different atoms. For example, for acetaldehyde, $\text{CH}_3\text{CH}=\text{O}$, one expects to enhance the partial charges of oxygen and carbon, rather than dominantly on the methyl and hydrogen atoms. The weighting function may be constructed based on the relative values of the off-diagonal matrix elements of the density matrix that contribute to the residual dipole moment, or devised according to the relative Mulliken charges. We choose a weighting function dependent on the relative electronegativity and the interatomic distance as follows:

$$\gamma_{ik} = \lambda \frac{\eta_i \eta_k}{r_{ik}^2} \quad (10)$$

where λ is a constant, chosen to be close to 1.0 \AA^{-2} , η_i and η_k are the Pauling electronegativity of atoms i and k , respectively. A weighting function without the use of electronegativity has been used previously^{19,29} and tested, which distributes charges equally to atomic sites with the same distance from the reference atom i without discriminating their relative electron-withdrawing powers.

Minimization of the Lagrangian (eq 9) by variation with respect to α_j yields

$$\sum_j \alpha_j \mathbf{D}_j^T + \beta \sum_k \gamma_{ik} \mathbf{D}_k^T = \mathbf{D}_i^T \quad (11)$$

where the superscript T indicates a matrix (vector) transpose. The solution of this equation, whose details are given in the Appendix, yields a final expression for the charge increments associated with the atomic hybridization dipole \mathbf{D}_i :

$$\alpha_j = \frac{\mathbf{D}_i^T \cdot \mathbf{D}_j}{\mathbf{D}_j^T \cdot \mathbf{D}_j} \quad (12)$$

where the various terms are defined as follows:



2.3. Molecular polarization consistent charges

The procedure outlined above also offers a convenient way to correct the errors in computed molecular polarization in semiempirical methods, which prevent a good description of hydrogen bonding interactions. Generally, it is necessary to use large basis functions with electron correlations to accurately predict molecular polarizability in electronic structure calculations. Further, higher order polarization terms are ignored in the treatment of intermolecular interactions in condensed phase simulations.

Let α_{SE} be the molecular polarizability tensor of an approximate quantum chemical model such as AM1 or PM6, and α_{target} be the target molecular polarizability tensor from experiment, from a high-level quantum mechanical calculation, or from an empirical dipole polarization model that has been shown to be able to yield a good prediction of molecular polarizability. Our goal is to derive a set of polarized DPPC charges to correct the error due to α_{SE} and to reproduce the molecular polarization described by α_{target} .

In the presence of an external electrostatic field, we write the total molecular dipole moment \mathbf{D}_{SE} as a sum of the permanent dipole \mathbf{D}_0 in the absence of the external field and the induced dipole moment \mathbf{D}_{ind} :

$$\mathbf{D}_{SE} = \mathbf{D}_0 + \mathbf{D}_{\text{ind}} \quad (16)$$

where the induced dipole moment is related to the external electrostatic field $\mathbf{E}(\mathbf{R})$ due to all other molecules, \mathbf{R} , in the condensed phase by

$$\mathbf{D}_{\text{ind}} = \alpha_{SE} \mathbf{E}(\mathbf{R}) \quad (17)$$

Accurate calculation of molecular polarizability requires a large basis set with augmented functions as well as correlation. Typically, the computed molecular polarizabilities from semiempirical methods are too small in comparison with experiments. For example, the

scalar polarizability for water is only 0.5 \AA^3 using AM1 and other semiempirical models, whereas the experimental value is 1.47 \AA^3 . Thus, if the current semiempirical models are used directly to model biomolecular systems or aqueous solution where polarization effects are important, intermolecular interactions due to polarization are also poorly treated.⁷ The error in the computed induced dipole moment due to the use of a small molecular polarizability can be expressed as follows:

$$\mathbf{D}_{ind} = \mathbf{D}_{ext} + \mathbf{D}_{ind}^{(0)} + \mathbf{D}_{ind}^{(1)} + \mathbf{D}_{ind}^{(2)} + \mathbf{D}_{ind}^{(3)} + \dots \quad (18)$$

which can be recast in terms of the computed induced dipole moment:

$$\mathbf{D}_{ind} = \mathbf{D}_{ind}^{(0)} + \mathbf{D}_{ind}^{(1)} + \mathbf{D}_{ind}^{(2)} + \mathbf{D}_{ind}^{(3)} + \dots \quad (19)$$

Here, \mathbf{D}_{ind} is the error due to the use of a weakly polarized model (from semiempirical or ab initio quantum calculations with small basis sets without including electron correlation), which can be included as part of the dipole-preservation target in the charge redistribution scheme described in Section 2.2. In this case, both the intrinsic, gas-phase dipole moment and the correct induced dipole moment due to an external electrostatic field from the environment will be preserved. This approach is called the DPPC method. We further assume (*vide infra*) that the total induced dipole moment can be decomposed into atomic contributions; here, we note that there is no unique way of decomposing molecular polarizability into atomic contributions, but abundant data exist, showing that atomic and group additivity schemes can be used to estimate the total molecular polarizability.^{34–35} Then, the Lagrangian optimization constraint of eq 5 for atomic point dipoles is replaced by

$$\mathbf{D}_{ind} = \mathbf{D}_{ind}^{(0)} + \mathbf{D}_{ind}^{(1)} + \mathbf{D}_{ind}^{(2)} + \mathbf{D}_{ind}^{(3)} + \dots \quad (20)$$

It should be emphasized that both the hybridization and induction polarization in eq 20 are obtained using the polarized wave function in the presence of the instantaneous external field $\mathbf{E}(\mathbf{R})$. The second term, which is the error in the computed molecular induced dipole, becomes zero in the absence of the external electrostatic field. In this case, the DPPC analysis is simply reduced to a dipole preserving charge (DPC) calculation outlined in the previous section. Obviously, if the quantum mechanical method used can accurately describe molecular polarization, the second term in eq 20 is also negligible since α is close to zero. Using the current semiempirical models, which describe molecular polarization rather poorly, we can, with eq 20, obtain a set of atomic partial charges that preserve the molecular dipole moment along with the extra induced dipole moment not adequately represented by the original wave function in condensed phase simulations.

The remaining task is to present a procedure so that only atomic parameters are needed to describe molecular polarization in the presence of an external field. To this end, we adopt another Thole's contribution,³² the interacting dipole (TID) polarization model, which has been adopted, at least conceptually, in many polarizable force fields currently being developed.^{36–41} One remarkable feature of the Thole interacting dipole model is that

isotropic atomic parameters that depend only on atomic numbers are used, but excellent agreement with experimental or high-level quantum mechanical results can be obtained on anisotropic molecular polarizabilities.³² A recent parameter set was reported by van Duijnen and Swart,⁴² and our own investigation confirmed that essentially the same atomic parameters were optimized using a much larger set of compounds in the fitting procedure (unpublished data; J. Pu and J. Gao, 2008). In the TID model,³² the total molecular polarizability can be conveniently decomposed into atomic site contributions,^{38,42} well suited for the present purposes.

Specifically, the TID molecular polarizability is given as follows:^{38,42}

$$\alpha_{TID} = \sum_i \alpha_i \mathbf{e}_i \mathbf{e}_i^T + \sum_{i,j} \mathbf{e}_i \mathbf{e}_j^T \alpha_{ij} \mathbf{e}_j \mathbf{e}_i^T \quad (21)$$

where the Greek letter subscripts specify a Cartesian coordinate (x, y or z), α_{TID} is an element of the molecular polarizability tensor, α_{TID} is a diagonal matrix of isotropic atomic

polarizabilities (the atomic parameters in the TID model),^{32,37,42} α_{ij} is the second-rank atomic interaction tensor (see eq 6 of ref³⁸), and the superscripts i and j run over atoms in the molecule. In the TID model, the molecular polarizability *tensor* is completely determined given a set of *isotropic* atomic polarizability and its geometry. It provides one of the most physical theories on molecular polarization that bridges electronic structure and classical models.

The atomic components of the total molecular polarizability tensor at atom i are defined by summing over the second summation in eq 21,^{34,38,42} which is

$$\alpha_i = \sum_j \mathbf{e}_i \mathbf{e}_j^T \alpha_{ij} \mathbf{e}_j \mathbf{e}_i^T \quad (22)$$

The i th mean atomic polarizability can be approximated as the average value of the three principal polarizability components determined by eq 22,^{34,42} If we assume that the relative atomic contribution to the total molecular polarizability is the same in the semiempirical method as that of the TID model, eq 19 can be simplified as

$$\alpha_i = \frac{1}{3} \text{Tr}(\alpha_i) \mathbf{1} \mathbf{1}^T \quad (23)$$

Eqs 12, 20 and 23 provide a general approach that links a set of atomic polarizabilities (which have proven to be very good for describing molecular polarization)⁴² and a given QM model to yield a set of dipole and polarization preserving charges that may be used in association with the explicit polarization (X-Pol)¹¹ potential to represent interfragment interactions or in multilayer QM/QM calculations.⁵

3. Rationale

The class IV charge models (CM1 through CM4, or simply CMx)^{15,43–44} have been shown to be useful for obtaining partial atomic charges, capable of yielding molecular dipole moments in excellent agreement with experiment. The CMx charges have been used as the fundamental electrostatic input to develop the accompanying solvation models (SMx) for prediction of solvent effects,^{43–44} and the method has been adopted in combined QM/MM simulations of chemical and enzymatic reactions.²⁸ Thus, it provides a good standard for comparison with the partial charges derived from the present DPPC method since the latter is also developed based on the principle to reproduce molecular dipole moments. Unlike the philosophy adopted in the CM1-CM4 charge models,^{15,27} the present DPPC approach does not require a training set of molecular dipole moments to parameterize the model. The DPPC analysis is an optimization procedure to yield partial atomic charges that exactly reproduce the molecular dipole moment from the corresponding quantum mechanical wave function. However, it is important to use a quantum chemical model, either semiempirical, ab initio or density functional theory, that can yield good results on molecular dipole moment in comparison with experiments. Therefore, we use the training set of molecules in the parameterization of the CM1 charges¹⁵ to illustrate that the current semiempirical models,^{12–13,45} consequently, the DPPC charges, can produce molecular dipole moments as accurately as the CMx charges.

The CM1 database consisting of 185 neutral molecules were selected from several compilations of experimental dipole moments with relative precision within ± 0.02 D.¹⁵ Storer et al. described the construction of this database as “mainly monofunctional organic molecules that have been chosen to isolate possible systematic deficiencies in the AM1 and PM3 wave functions.”¹⁵ For H, C, N and O containing compounds, the database includes 12 alcohols, 8 esters and lactones, 16 carbonyl compounds, 9 acids, 10 ethers, 13 amines, 3 amides, 7 imines and aromatics, and 24 mono and multifunctional nitrile compounds. In addition, 83 compounds are included to cover halogen, silicon and sulfur containing compounds. The root mean square (RMS) errors for CM1-AM1 charges from ref¹⁵ are duplicated in Table 1, along with that of the recently developed PM6 model. Overall, AM1 and PM3, respectively, have RMS errors of ± 0.45 and ± 0.47 D for the 185 compounds, while the RMS errors in the dipole moments computed using the CM1 and CM2 charges are ± 0.25 and ± 0.22 D. AM1 and PM3 perform very poorly for the nitriles with average errors greater than 1.0 D and 0.8 D, respectively. If the 24 cyano compounds are excluded, the errors in the semiempirical models are reduced to ± 0.34 D for AM1, compared with the CM1A charge model of ± 0.24 D. If we only consider the H, C, N and O containing compounds, which are the most relevant to biomolecular systems, excluding the nitriles, the AM1 error in dipole moment is ± 0.24 D, even smaller than that of ± 0.29 D from the CM1A charge model and ± 0.32 D from CM2 charges. If the Mulliken charges are directly used to estimate molecular dipole moments, the errors are very large, in the range of ± 0.8 to ± 0.9 D.¹⁵ Overall, the comparison suggests that the current semiempirical models^{12–13,45} are excellent for estimating molecular dipole moments despite their inconsistent performance on relative energies.

4. Computational Details

Molecular geometries for the H, C, O, and N subset of the CM1-database have been optimized using the corresponding semiempirical model to determine the DPPC charges. We have also implemented and recomputed the CM2 charges, from which the CM2 dipole moments were estimated. Note that the CM2 parameters were optimized using geometries optimized with the HF/MIDI! method, thereby the small difference between the present CM2 dipoles and those reported in ref ⁴⁶.

The DPPC method is not without parameters since a weighting function is used to redistribute the residual partial charges associated the constrained dipole target. The parameters include the distance damping factor λ in eq 6, which is chosen to be unity, and the Pauling electronegativity for each element (H: 2.20; C: 2.55; N: 3.04; and O: 3.44). To preserve the desired molecular polarizability using the Thole interaction dipole model, isotropic atomic polarizability parameters will be needed; these parameters have been optimized and reported by van Duijnen and Swart.⁴² In addition, the DPPC method involves the inversion of a 3×3 geometrical matrix δ^i , which could be singular in certain situations such as a planer molecule. In this case, we adopted the eigenvalue smoothing approach to ensure numerical stability. Thus, the three eigenvalues of δ^i are perturbed with a small increment of $\theta = 10^{-5}$ to avoid singularity:¹⁹


$$\omega = \omega_{\max} + \theta \quad (24)$$

where ω is one of the three eigenvalues of δ^i and ω_{\max} is the largest eigenvalue.

5. Illustrative examples

We first present the dipole preserving charges (DPC) in the gas phase for the compounds in the CM1 database in Sections 5.1 through 5.6. Then, the effect of including polarization consistency in the DPPC method is discussed next.

5.1. Water and alcohols

The Mulliken population, DPC and CM2 atomic charges for water and methanol determined using AM1, PM3 and PM6 models (note that there are no CM2 parameters for the PM6 method) are shown in Figure 1. The absolute values of standard Mulliken population charges are generally too small in comparison with the typical values used in the force fields for biomolecular simulations and that needed in the generalized Born-base solvation models. The dipole moments computed using the Mulliken population charges are substantially smaller than the results from the corresponding wave function and experiments. The DPC charge for oxygen of water is enhanced from $-0.38 e$ to $-0.65 e$, which is slightly smaller than the CM2 charge of -0.71 . The latter over-estimates the molecular dipole moment proportionally. For alcohols, the oxygen partial charge is increased by about 0.2 a.u. similar to that from the empirical fit in the CM2 model. As in the CM2 model, the partial charges on atoms of the alkyl group are not significantly affected by the Lagrangian minimization procedure, and it is of interest to point out that the hydrogen atoms attached to a carbon can have different partial charges, depending on their conformational orientation, whereas in

they are typically assigned with an identical partial charge in force fields such as OPLS⁴⁷ and CHARMM,⁴⁸ independent on the instantaneous environment. Overall, the mean unsigned errors in the computed dipole moments for the 12 alcohol compounds are 0.13 D for the AM1-DPC charges, 0.33 D for PM3, 0.36 D for PM6, and 0.20 D for CM2, while the RMS errors are 0.17, 0.61, 0.43, and 0.25 D, respectively.

5.2. Esters and lactones

Eight compounds are included for the esters and lactones functional groups in the CM1 database. The RMS errors in the computed dipole moments for the DPC model are 0.25, 0.30, 0.38 D from the AM1, PM3 and PM6 models, compared to the CM2 error of 0.33 D using the AM1 wave function. To illustrate the qualitative trend, the partial atomic charges are depicted in Figure 2 for methyl acetate. The positive charge on the carbonyl carbon and the negative charge on the ester and carbonyl oxygen atoms are significantly enhanced in the DPC model and the results are in excellent accord with the CM2 method.

5.3. Carbonyl compounds

A total of sixteen carbonyl compounds, ranging from aldehyde and ketones to cyclic and conjugated carbonyls are included, which have RMS errors of 0.25, 0.42 and 0.43 D using AM1, PM3, and PM6, respectively. The CM2 error using AM1 is 0.32 D, somewhat larger than that of the original CM1A charges (0.21 D). The partial atomic charges from the DPC and CM2 models are shown in Figure 3 for 2-butanone. Although the carbonyl oxygen and carbon net charges are increase by about 0.1 a.u., they are still smaller than that used in the CHARMM force field,⁴⁸ which are -0.51 and $+0.51$ a.u. for oxygen and carbon, respectively. The force field charges are designed for condensed-phase simulations that include effective polarization effects. It would be interesting to examine the DPC performance in solution and biomolecular simulations. The CM2 charges are slightly more enhanced than the DPC charges for the carbonyl group, and the general patterns of the partial atomic charges from both models (DPC and CM2) are very similar.

5.4. Carboxylic acids and ethers

Although the average errors for the nine carboxylic acids appear to be large from all methods, this is somewhat skewed by the two acrylic acid configurations. Nevertheless, the results are reasonable with RMS errors of 0.31, 0.23, and 0.41 D for AM1, PM3, and PM6, respectively, and it is 0.58 D from the CM2 charges. The ether compounds are very well described by all models with RMS errors less than 0.2 D, except PM6, which has a relatively large RMS deviation of 0.45 D.

5.5. Amines and ammonia

Although the molecular dipole moments are very well reproduced by the AM1 and PM3 models, with an average error of 0.25 D in comparison with experiments, it is well known that the charge separation between nitrogen and attached (polar) hydrogen atoms are not well described. The later PM6 model performs much worse overall, with an RMS error of 0.64 D. The main problem is that the negative charge on the amino nitrogen atom is small in comparison with that optimized for condensed phase simulations. The redistribution of the


hybridization polarization charges in the DPC model enhances the nitrogen negative charges of alkyl amines and ammonia significantly, in good accord with the corresponding CM2 charges. This is illustrated for methylamine and ammonia in Figure 4. However, the planer aniline group is still under estimated in the DPC charges, suggesting that further improvements may be needed to circumvent the singularity issue in the geometrical matrix inversion.



5.6. Amides and imines

The atomic charges for acetamide and N-methylformamide are shown in Figure 5. Atomic charges on the nitrogen atom in amides and imines from the DPC model are significantly smaller than those mapped with the CM2 model. In the latter case, more than -0.4 a.u. of negative charge is placed on the nitrogen atom, along with proportional increase in the positive charges on the hydrogen atoms attached to the amide and imine nitrogen atoms and the carbonyl or imine carbon atoms. The partial charges on the amide nitrogen atom used in the OPLS⁴⁷ and CHARMM⁴⁸ force fields are -0.57 and -0.47 a.u., respectively. Although the force field charges take into account average polarization effects in aqueous solution, the magnitude is significantly smaller than the CM2 charges. It is expected that the DPC charges will be enhanced in aqueous solution due to polarization effects. Interestingly, despite the large range of negative charge on the amide nitrogen atom from different models, the overall molecular dipole moments are not significantly affected, reflecting the arbitrariness of partial charge assignment.

5.7. Polarization

Although the current semiempirical models can yield an excellent estimate of molecular dipole moment in the gas phase, polarization effects are severely underestimated, preventing these methods to be applied to model biomolecular interactions such as solvation and ligand-protein binding. The DPPC method is designed to reproduce the correct electronic polarization with a better description of the induced dipole moment in the presence of an external electric field. To properly test this approach, statistical mechanical Monte Carlo and molecular dynamics simulations of solutions are required, which is beyond the scope of this paper, but will be carried out subsequently. Here, we examine the polarization preservation capability of the DPPC model by considering the interaction profile of two water molecules along the minimum interaction energy path (Figure 6), and the bimolecular complexes of one water with a fluoride ion and with a chloride ion.

We use the X-Pol potential with the AM1 model in the present calculation,⁶ in which the molecular orbitals are block-localized within each monomer space. The system is represented by a two fragments,⁶⁻¹⁰ in which charge transfer effects are not included.⁴⁹ First, the induced dipole moment for each water is determined by subtracting the “permanent” (i.e., gas-phase) dipole moment of water from that of the polarized monomer wave functions optimized from the X-Pol method. Then, the errors for the atomic induced dipoles,  in each water monomer from the AM1 Hamiltonian are estimated according to eq 23, using the Thole interaction dipole model.⁴² The induced dipole error

 together with the hybridization polarization  is used to derive the polarized DPPC charges, which are included in SCF optimization of the X-Pol wave function.

The X-Pol interaction energies from the AM1 dipole preserving model (DPC) only and from the full DPPC charges are compared in Figure 6. The Mulliken population charges are too small to adequately represent the Coulomb potential from the original (AM1) wave function. These charges are significantly enhanced to preserve the dipole moment of the individual water molecules that are polarized in the presence of the electrostatic field of the other monomer. Thus, the binding energy is lowered by 1.6 kcal/mol (from -2.2 to -3.8 kcal/mol) using the DPC charges in the X-Pol(AM1) potential. The DPPC model yields stronger hydrogen bonding interactions (-4.1 kcal/mol) than the DPC model alone, due to correction of dipole polarization effects. At the interaction minimum, the binding energy from DPPC model is 6% greater than that with polarization preservation contributions. The partial atomic charges at the optimal geometry are shown in Figure 7. The Thole interacting dipole model yields a molecular polarizability identical to the experimental value for water; the partial charge on the acceptor oxygen atom is enhanced by 0.04 e units in going from the AM1 model (DPC charges) to the correctly polarized model (DPPC). Note that the X-Pol potential constructed from monomer block-localized molecular orbitals does not include charge transfer energy between the two monomers; thus, the total interaction energy is smaller than high-level ab initio results. In addition, the present calculations did not include the long-range dispersion interactions modeled by Lennard-Jones terms.⁶⁻⁷ Charge transfer and exchange repulsion can be explicitly incorporated into the X-Pol theory.⁴⁹

The polarization effect is further examined by placing a negative charge at distances of 2.049 and 1.550 Å from one of the hydrogen atoms of water at an angle of about 152° from oxygen (Figure 8), which mimics bimolecular hydrogen bonding complexes with a chloride ion and with a fluoride ion, respectively. There is no net polarization effect on the anion in the present minimum basis set representation of the full shell halogen ions and there is no charge transfer between water and the ion in this analysis, so only the wave function of the water molecule is polarized. Clearly, the dipole preserving feature (DPC) is important in charge estimation in comparison with the Mulliken population analysis (Figure 8).

Importantly, the inclusion of polarization consistency in the DPPC procedure to mimic the Thole interaction dipole polarizability, which is identical to that of the experimental value, further enhances the atomic charge separation by as much as 0.13 and 0.17 |e| units on the donor hydrogen atom for the chloride and fluoride ion complexes, respectively (Figure 8). Concomitantly, the interaction energies are enhanced from -13.9 to -15.4 kcal/mol for the chloride ion complex, and from -20.2 to -20.7 kcal/mol for the fluoride ion complex. Apparently, there are some saturation effects in dipole polarization since the probe ion is very close to bonding distance where the neglect of charge transfer effects is no longer a good approximation in the X-Pol model. In the latter case, if the water structure is allowed to relax, the AM1 model yields a hydroxide and hydrogen fluoride complex instead.

6. Conclusions

In the present study, we describe a method for estimating atomic partial charges, in which dipole moment from the corresponding quantum mechanical calculation is exactly preserved and the molecular polarizability is consistently incorporated. The method is called the dipole preserving and polarization consistent (DPPC) charge method. The dipole preservation approach has an origin from the work of Thole and van Duijnen;¹⁹ the new features of the DPPC model include charge redistribution based on relative atomic electronegativity and the treatment of molecular dipole polarization, consistent with the experimental or theoretical molecular polarizability. The latter is typically underestimated in semiempirical methods and in ab initio molecular orbital theory without inclusion of electron correlation. It was emphasized that the semiempirical models such as AM1, PM3 and PM6 can yield excellent results on molecular dipole moments in comparison with experiment, making these methods well suited for the DPPC analysis. The standard errors on the computed dipole moments from these semiempirical methods are similar to that determined using the well-parameterized Class IV charge models that have been successfully used in a variety of applications.

Using the database designed in the optimization of the CM1 model, excluding nitrile compounds that are not adequately treated by the semiempirical models, we show that the trends of the DPPC partial atomic charges are in good accord with those obtained from the CM2 model (using AM1). Furthermore, it was illustrated that the DPPC method can effectively introduce enhanced electronic polarization to mimic the correct molecular polarizability in a water dimer test case as well as in water-halogen ion bimolecular complexes using the X-Pol potential with AM1 Hamiltonian. One main difference between the present DPPC method and other charge models is that the DPPC partial charges reproduce the local (hybridization) and molecular dipole moments exactly, whereas partial atomic charges derived from electrostatic potential fitting procedure are known to have difficulties for atoms in the interior of a large molecule such as a protein. The DPPC analysis can also be coupled with combined QM/MM or QM/QM simulations to derive polarization consistent charges for use in empirical force fields.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by the National Institutes of Health (RC1 GM091445) and the National Science Foundation (CHE09-57162).

Appendix. Derivation of Equation 12

We start with for the charge distribution over all N atoms in the molecule,



that preserves the atomic dipole \mathbf{D}_i ,

 (A1)

The total charge variation, by definition, is

 (A2)


where the total weight W^j and the weighted average coordinate $\langle r \rangle_j$ associated with the dipole \mathbf{D}_j have been defined in eqs 11 and 12, respectively. Similarly, the total residual (dipole-preserving and polarization-correction) dipole is

 (A3)

Multiply both sides of eq A1 by $\langle r \rangle_j$ from the right side, we obtain

 (A4)

Subtract eq A4 from A3 to yield

 (A5)

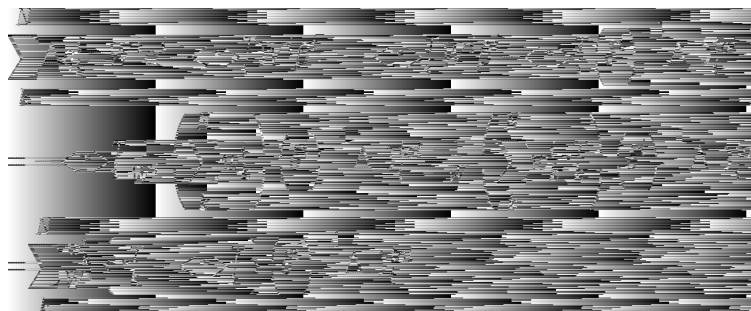
where we have used the definition for the matrix δ^j (eq 10). Multiplying both sides by the inverse matrix of δ^j we obtain

 (A6)

Obviously,

 (A2)

Since



(A7)

we separate the result term by term and obtain



(A8)

References

1. Coppens, P. X-ray charge densities and chemical bonding. Chester, England: Oxford University Press; 1997.
2. Bachrach, S. Rev. Comput. Chem. Lipkowitz, KB.; Boyd, DB., editors. Vol. Vol. 5. Weinheim: VCH; 1994.
3. Gao J, Xia X. Science. 1992; 258:631. [PubMed: 1411573]
4. Hratchian HP, Parandekar PV, Raghavachari K, Frisch MJ, Vreven T. J. Chem. Phys. 2008; 128:034107. [PubMed: 18205488]
5. Parandekar PV, Hratchian HP, Raghavachari K. J. Chem. Phys. 2008; 129:145101. [PubMed: 19045166]
6. Gao J. J. Phys. Chem. B. 1997; 101:657.
7. Gao J. J. Chem. Phys. 1998; 109:2346.
8. Xie W, Gao J. J. Chem. Theory Comput. 2007; 3:1890. [PubMed: 18985172]
9. Xie W, Song L, Truhlar DG, Gao J. J. Chem. Phys. 2008; 128 234108/1.
10. Song L, Han J, Lin YL, Xie W, Gao J. J. Phys. Chem. A. 2009; 113:11656. [PubMed: 19618944]
11. Xie W, Orozco M, Truhlar DG, Gao J. J. Chem. Theory Comput. 2009; 5:459. [PubMed: 20490369]
12. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP. J. Am. Chem. Soc. 1985; 107:3902.
13. Stewart JJP. J. Mol. Model. 2007; 13:1173. [PubMed: 17828561]
14. Tuttle T, Thiel W. Phys. Chem. Chem. Phys. 2008; 10:2159. [PubMed: 18404221]
15. Storer JW, Giesen DJ, Cramer CJ, Truhlar DG. J Comput.-Aided Mol. Des. 1995; 9:87. [PubMed: 7751872]
16. Lowdin P-O. J. Chem. Phys. 1950; 18:365.
17. Mulliken RS. J. Chem. Phys. 1964; 61:20.
18. Hall D, Williams DE. Acta Crystallogr. 1975; A 31:56.
19. Thole BT, Vanduijnen PT. Theor. Chim. Acta. 1983; 63:209.
20. Sternberg U, Koch FT, Mollhoff M. J. Comput. Chem. 1994; 15:524.
21. Singh UC, Kollman PA. J. Comput. Chem. 1984; 5:129.
22. Luque FJ, Illas F, Orozco M. J. Comput. Chem. 1990; 11:416.
23. Momany FA. J. Phys. Cem. 1978; 82:592.
24. Gao J, Luque FJ, Orozco M. J. Chem. Phys. 1993; 98:2975.
25. Li J, Zhu T, Cramer CJ, Truhlar DG. J. Phys. Chem. A. 1998; 102:1820.

26. Coppens P. *Angew. Chem. Int. Ed.* 2005; 44:6810.
27. Kalinowski JA, Lesyng B, Thompson JD, Cramer CJ, Truhlar DG. *J. Phys. Chem. A.* 2004; 108:2545.
28. Acevedo O, Jorgensen WL. *Acc. Chem. Res.* 2010; 43:142. [PubMed: 19728702]
29. Swart M, Van Duijnen PT, Snijders JG. *J. Comput. Chem.* 2001; 22:79.
30. Pople JA, Santry DP, Segal GA. *J. Chem. Phys.* 1965; 43:S129.
31. Stewart JJP. *J. Comput.-Aided Mol. Des.* 1990; 4:1. [PubMed: 2197373]
32. Thole BT. *Chem. Phys.* 1981; 59:341.
33. Pople JA, Segal GA. *J. Chem. Phys.* 1965; 43:S136.
34. Applequist J, Carl JR, Fung K-K. *J. Am. Chem. Soc.* 1972; 94:2952.
35. Miller KJ. *J. Am. Chem. Soc.* 1990; 112:8533.
36. Gao J, Habibollahzadeh D, Shao L. *J. Phys. Chem.* 1995; 99:16460.
37. Xie W, Pu J, MacKerell AD Jr. Gao J. *J. Chem. Theory Comput.* 2007; 3:1878. [PubMed: 18958290]
38. Xie W, Pu J, Gao J. *J. Phys. Chem. A.* 2009; 113:2109. [PubMed: 19123850]
39. Ren P, Ponder JW. *J. Comput. Chem.* 2002; 23:1497. [PubMed: 12395419]
40. Ren P, Ponder JW. *J. Phys. Chem. B.* 2003; 107:5933.
41. Harder E, Anisimov VM, Vorobyov IV, Lopes PEM, Noskov SY, MacKerell AD Jr. Roux B. *J. Chem. Theory Comput.* 2006; 2:1587.
42. van Duijnen PT, Swart M. *J. Phys. Chem. A.* 1998; 102:2399.
43. Li J, Zhu T, Cramer CJ, Truhlar DG. *J. Phys. Chem. A.* 2000; 104:2178.
44. Chamberlin AC, Cramer CJ, Truhlar DG. *J. Phys. Chem. B.* 2006; 110:5665. [PubMed: 16539512]
45. Stewart JJP. *J. Comp. Chem.* 1989; 10:209.
46. Zhu T, Li J, Hawkins GD, Cramer CJ, Truhlar DG. *J. Chem. Phys.* 1998; 109:9117.
47. Jorgensen WL, Tirado-Rives J. *J. Am. Chem. Soc.* 1988; 110:1657.
48. MacKerell AD Jr. Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE III, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. *J. Phys. Chem. B.* 1998; 102:3586. [PubMed: 24889800]
49. Gao J, Cembran A, Mo Y. *J. Chem. Theory Comput.* 2010; 6:2402.
50. Gao, J. *Rev. Comput. Chem.* Lipkowitz, KB.; Boyd, DB., editors. Vol. Vol. 7. New York: VCH; 1995. p. 119

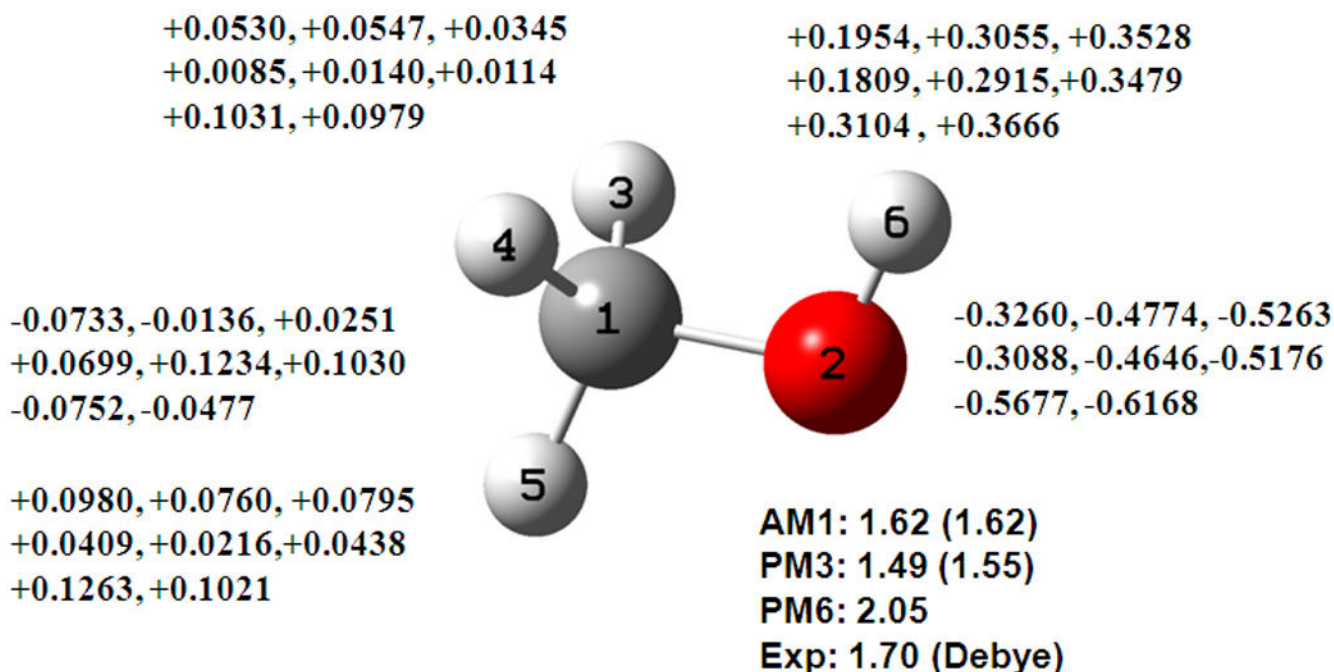
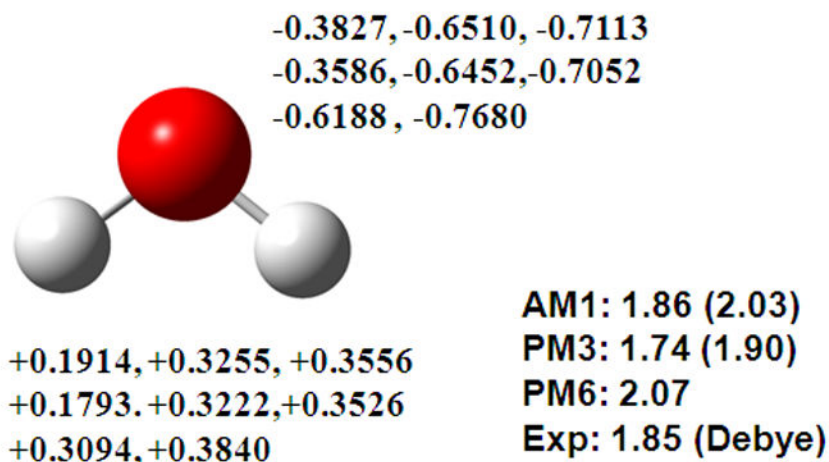


Figure 1.

Computed dipole moments and partial atomic charges for water and methanol from Mulliken population (first), the DPC method (second), and the CM2 model (third), using the AM1 (top line), PM3 (middle line), and PM6 (lower line) models. Partial charges are determined by Mulliken population analysis given first, followed by the DPPC charges and CM2 charges. The first row for each atom lists charges from the AM1 method, the second row shows charges from the PM3 model, and the third row depicts charges from the PM6 Hamiltonian. All charges are given in electron charge unit. Computed dipole moments from each model are given in Debye followed by values determined with the CM2 charges, and

the experimental dipole value is listed last. This convention is used throughout Figures 1 to 5.

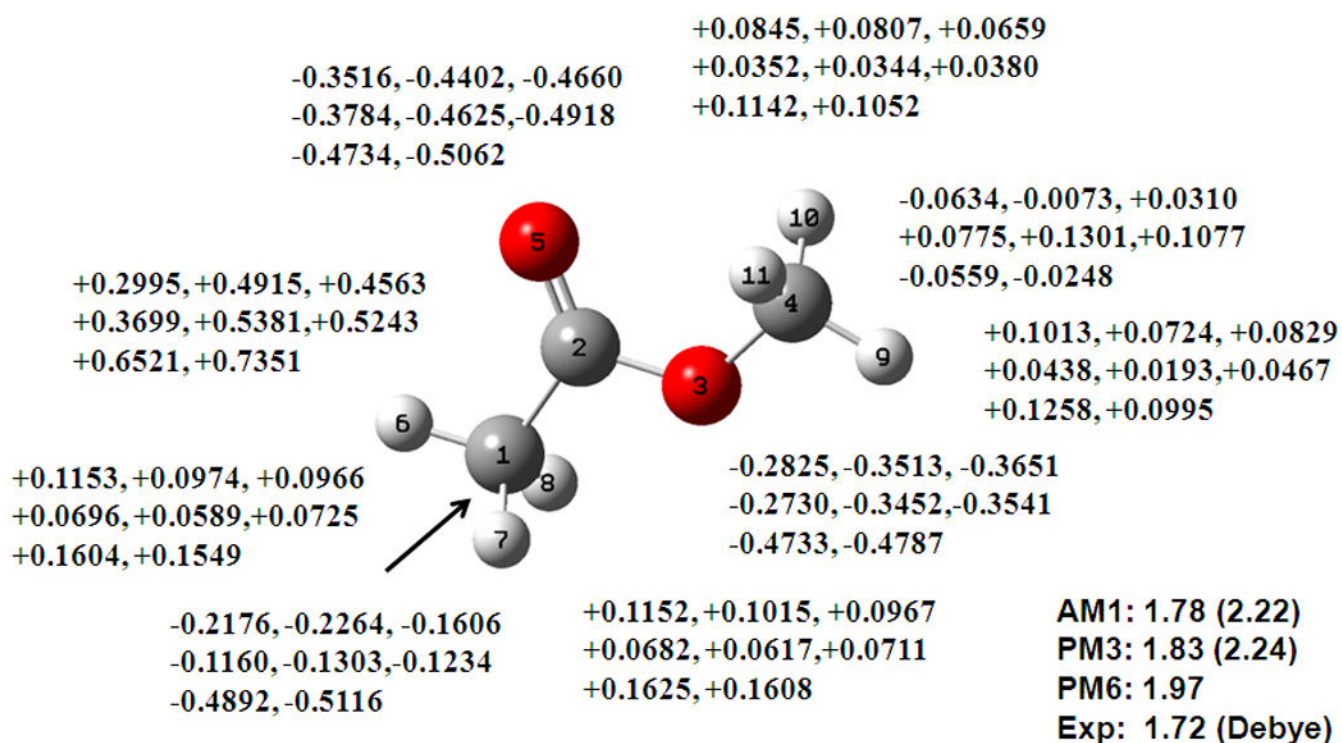


Figure 2. Computed dipole moment and partial atomic charges for methyl acetate from Mulliken population (first), the DPC method (second), and the CM2 model (third), using the AM1 (top line), PM3 (middle line), and PM6 (lower line) models. See also Figure 1 caption.

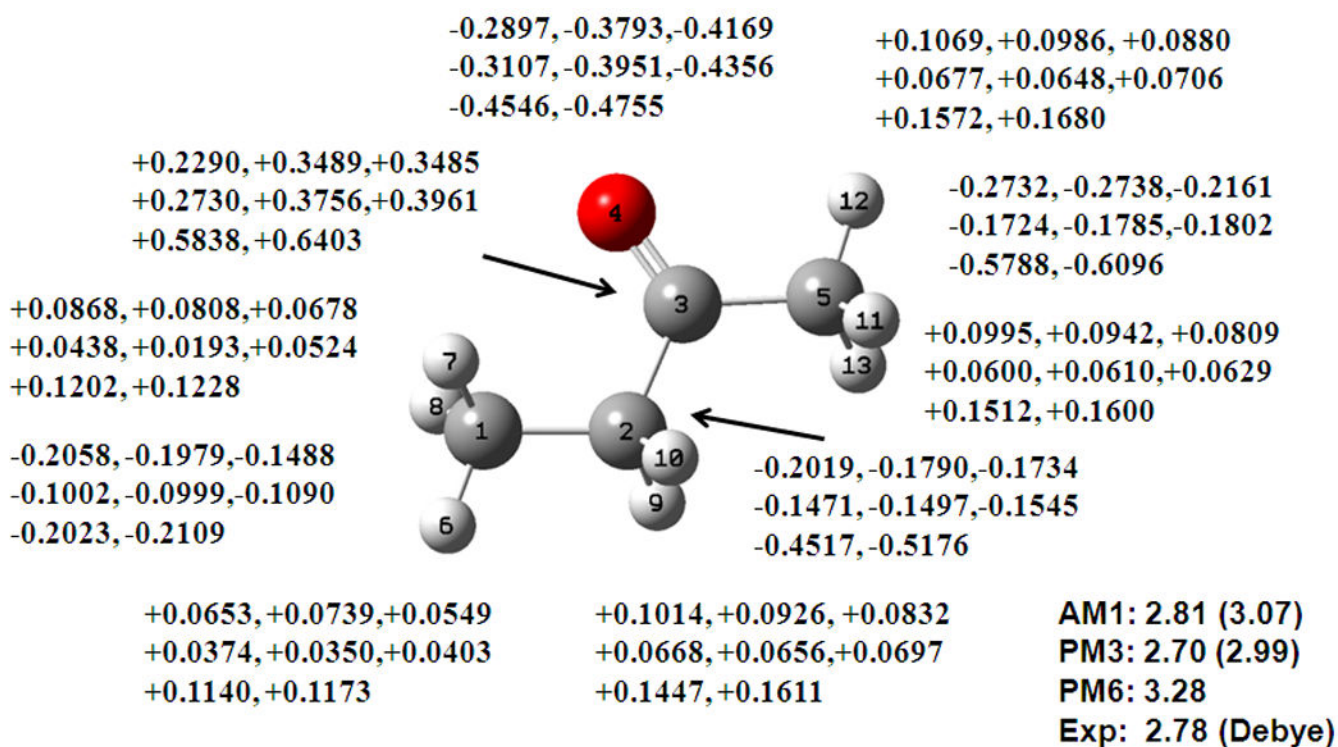
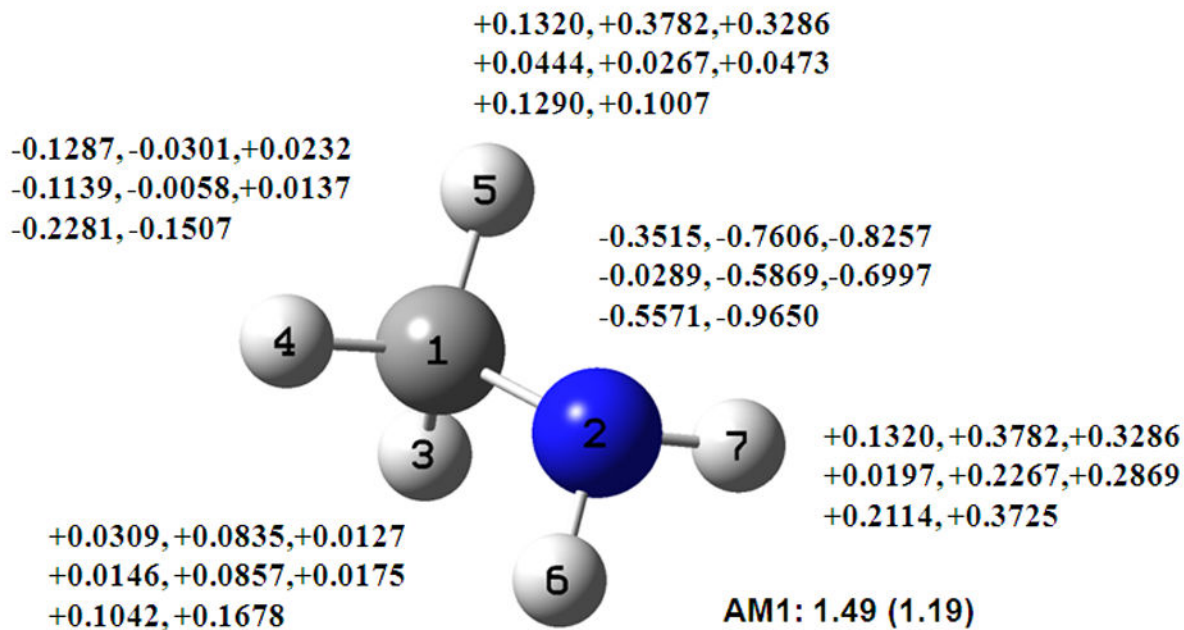


Figure 3. Computed dipole moment and partial atomic charges for 2-butanone from Mulliken population (first), the DPC calculation (second), and the CM2 model (third), using the AM1 (top line), PM3 (middle line), and PM6 (lower line) models. See also Figure 1 caption.



AM1: 1.85 (1.60)
PM3: 1.55 (1.39)
PM6: 2.24
Exp: 1.47 (Debye)



AM1: 1.49 (1.19)
PM3: 1.40 (1.16)
PM6: 2.03
Exp: 1.31 (Debye)

Figure 4. Computed dipole moments and partial atomic charges for ammonia and methylamine from Mulliken population (first), the DPC analysis (second), and the CM2 model (third), using the AM1 (top line), PM3 (middle line), and PM6 (lower line) models. See also Figure 1 caption.

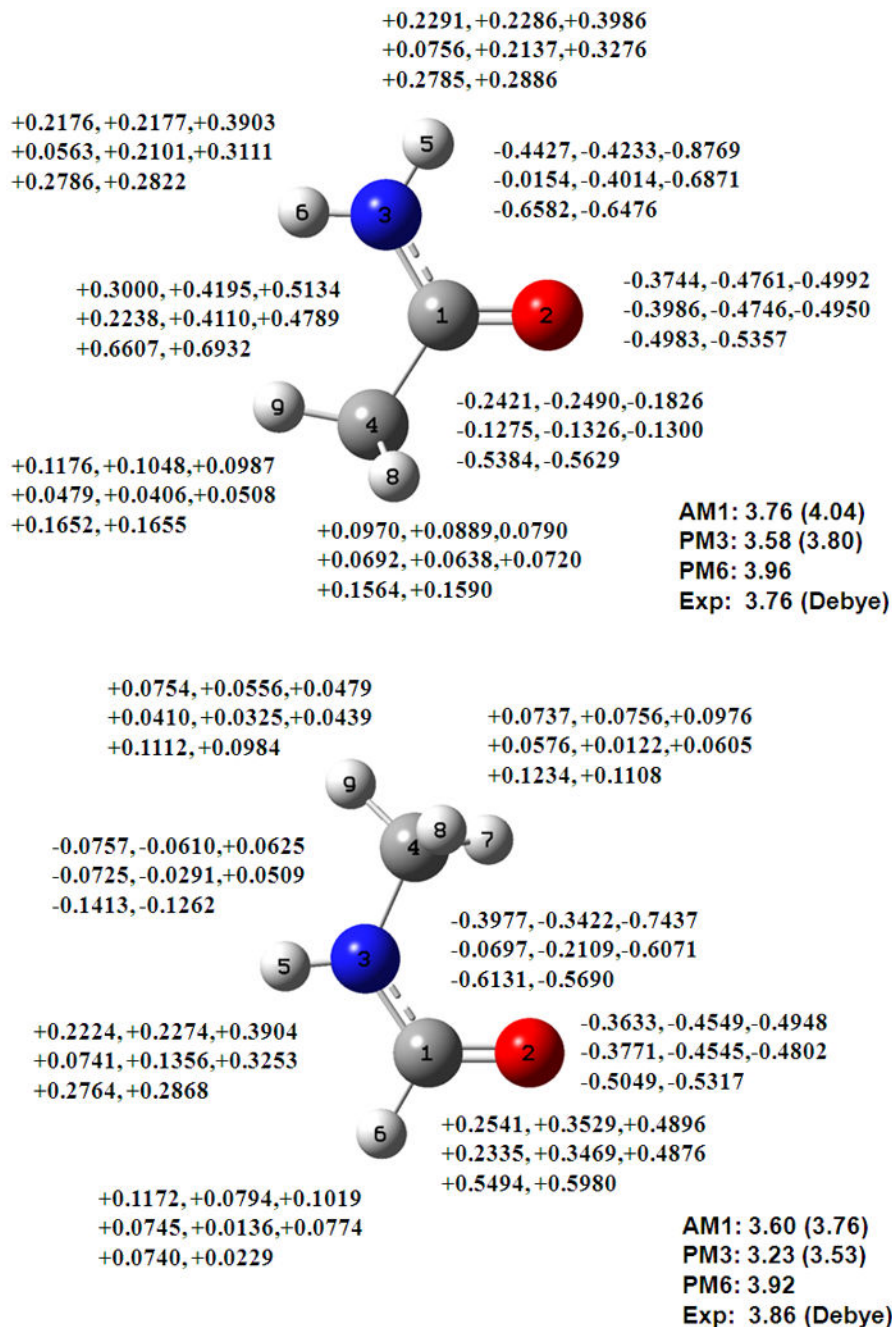


Figure 5. Computed dipole moments and partial atomic charges for acetamide and N-methylformamide from Mulliken population (first), the DPC method (second), and the CM2 model (third), using the AM1 (top line), PM3 (middle line), and PM6 (lower line) models. See also Figure 1 caption.

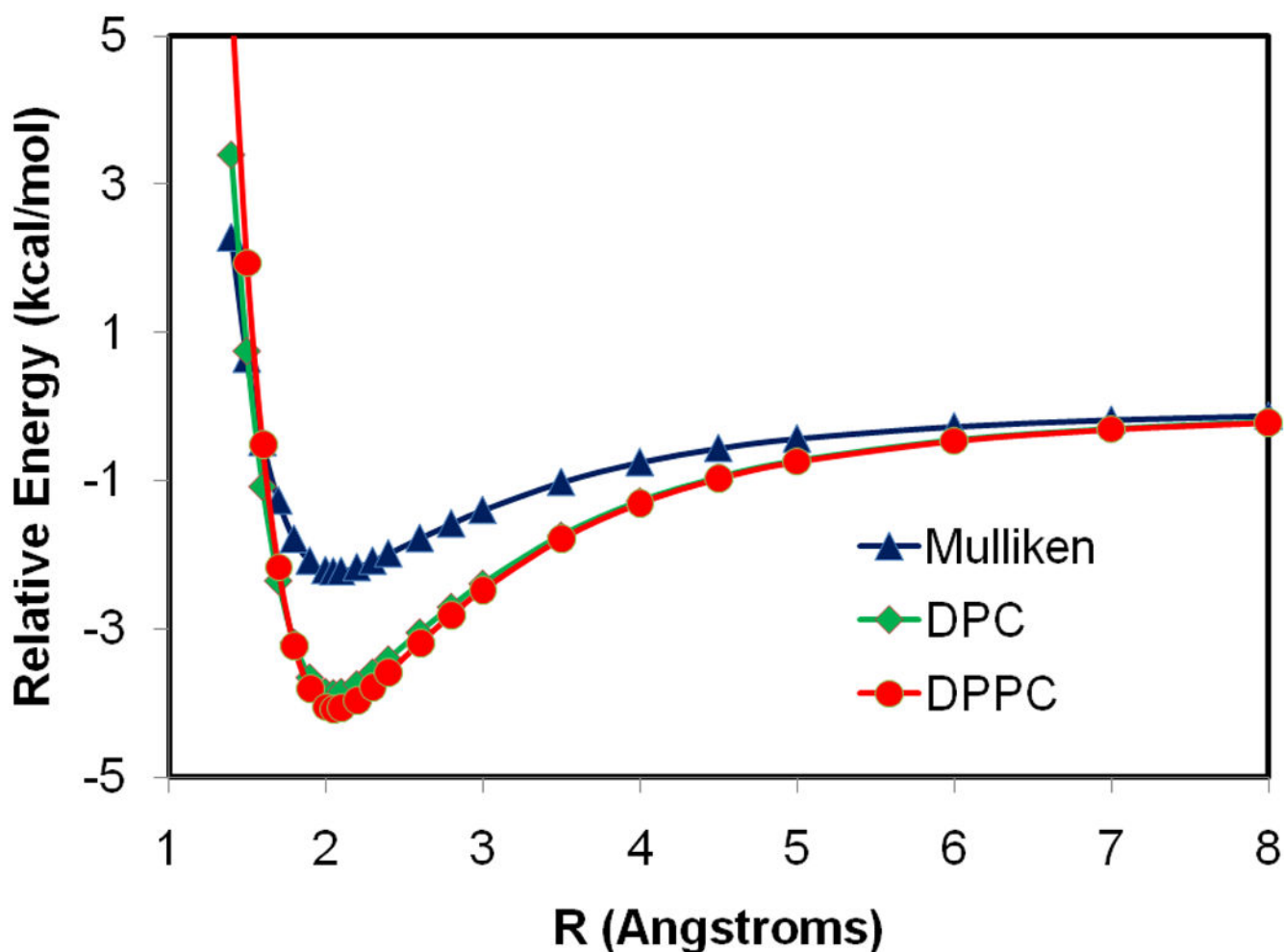
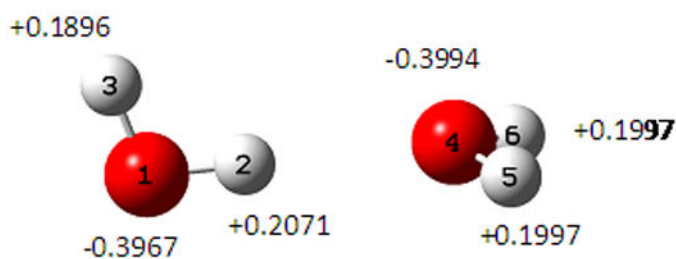
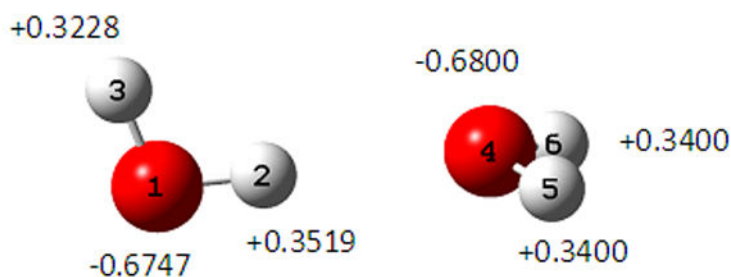


Figure 6. Interaction energy profiles for a water dimer complex as a function of the donor hydrogen and acceptor oxygen distance. The structural arrangement of the dimer complex is depicted by the inset in the figure. The X-Pol potential with the AM1 Hamiltonian is used with each water molecule treated as a separate fragment. In the present calculation, a value of 2.0 was used for the semiempirical α parameter in QM/MM type integrals,⁵⁰ and the long-range dispersion interactions modeled by a Lennard-Jones term was not included. The Coulomb interactions between the two water fragments are determined by the X-Pol theory using the self-consistently optimized Mulliken population charges in blue, the dipole preserving charges (DPC) in green, and the dipole preserving and polarization consistent (DPPC) charges in red.

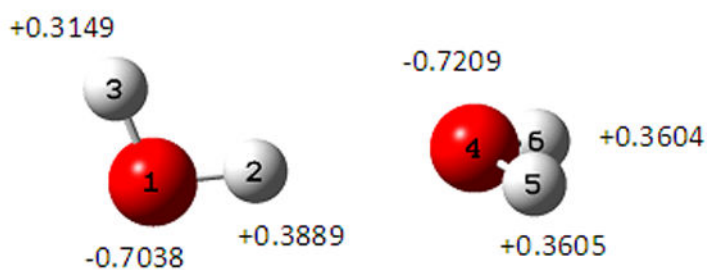
(a). Mulliken



(b). DPC.

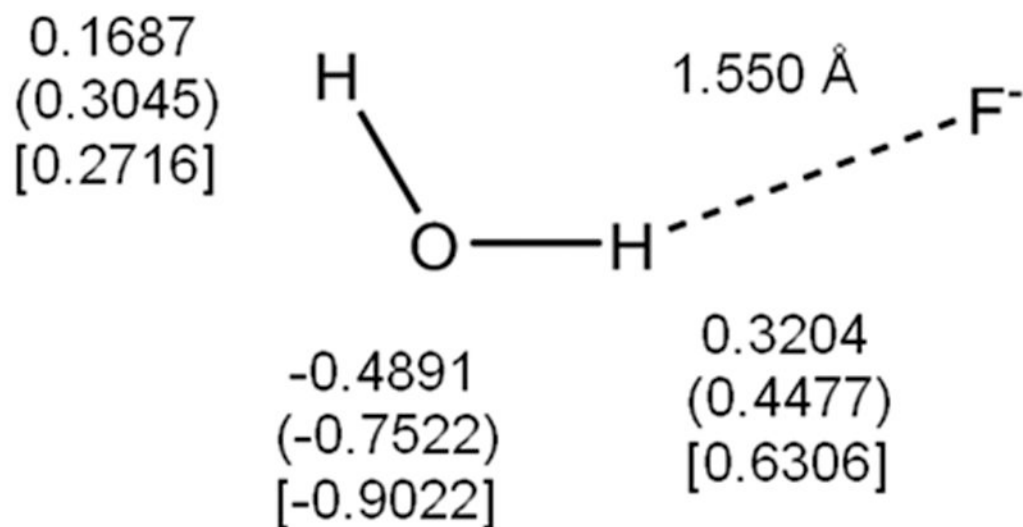


(c). DPPC

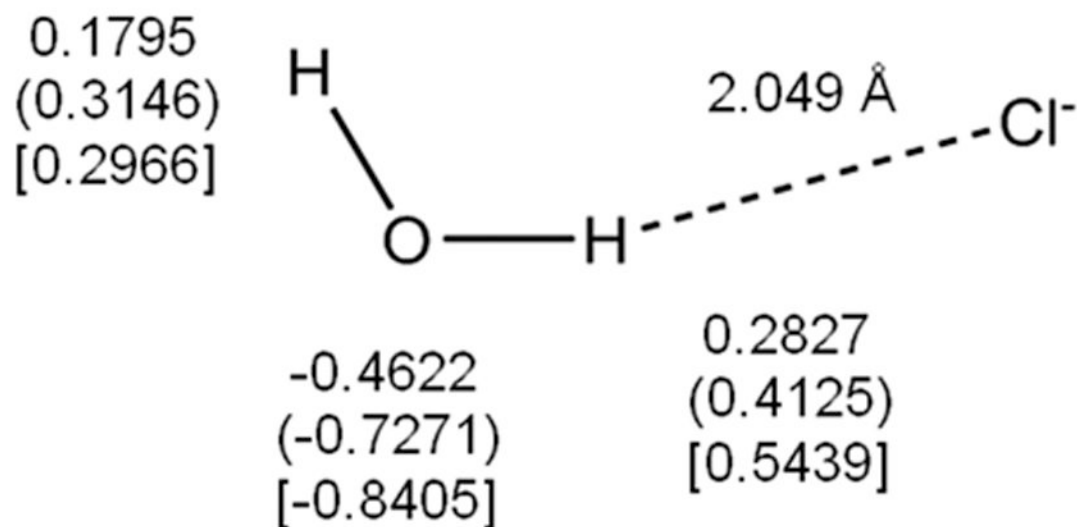
**Figure 7.**

Equilibrium partial charges at the optimized dimer configuration, determined from the fully polarized monomer wave functions by the electrostatic field of the other water molecule using (a) Mulliken population analysis, (b) dipole preserving charge only, and (c) the full dipole preserving and polarization consistent analysis.

(A)



(B)

**Figure 8.**

Computed partial atomic charges (a.u.) on the water molecule in (A) HOH...Cl⁻ and (B) HOH...F⁻ bimolecular complexes, in which the water geometry is held as that in the gas phase from AM1 optimization. The Mulliken population charges are given first, followed by the dipole preserving charge (DPC) in parentheses and the dipole preserving and polarization consistent (DPPC) charges in square brackets. The ion charges are negative one in the present minimum basis set representation using the AM1 model.

Table 1

Root mean square errors in the computed dipole moment (debye) from experimental data for compounds used in the parameterization of the class IV charge model (CM1A).

compounds	No.	PM3	PM6	AMI	CM1A	CM2(AMI) ^a
Alcohols	12	0.61	0.43	0.17	0.21	0.25 (0.17)
Esters, lactones	8	0.30	0.38	0.25	0.24	0.33 (0.18)
Carbonyls	16	0.42	0.43	0.25	0.18	0.32 (0.17)
Acids	9	0.23	0.41	0.31	0.41	0.58 (0.16)
Ethers	10	0.19	0.45	0.17	0.18	0.14 (0.20)
Amines	13	0.25	0.64	0.25	0.17	0.14 (0.17)
Amides	3	0.52	0.26	0.15	0.69	0.23 (0.23)
Imines	7	0.35	0.38	0.28	0.43	0.34 (0.31)
Nitriles	24	0.78	0.30	1.00	0.16	0.29 (0.21)
H, C, N, O without nitriles	78	0.38	0.46	0.24	0.29	0.32
All H, C, N, O	102	0.47	0.43	0.47	0.29	0.31 (0.21)
Other compounds	83	0.47		0.40	0.19	
All compounds	185	0.47	0.37 ^b	0.45	0.25	(0.22)

^aValues in parentheses were obtained using the HF/MIDI1 geometry in ref. ²⁵.

^bPM6 average unsigned error for 131 compounds containing H, C, N, O, F, S, Cl, Br, and I atoms. The PM6 average error for 313 main group compounds is 0.60 D. ¹³ The corresponding AM1 average unsigned errors are 0.38 and 0.65 D for that data set. ¹³