JCTC Journal of Chemical Theory and Computation

Fitting Molecular Electrostatic Potentials from Quantum Mechanical Calculations

Hao Hu, Zhenyu Lu, and Weitao Yang*

Department of Chemistry, Duke University, Durham, North Carolina 27709

Received September 26, 2006

Abstract: We develop here a new method to fit the molecular electrostatic potentials obtained in quantum mechanical calculations to a set of classical electrostatic multipoles, usually point charges located at atomic positions. We define an object function of fitting as an integration of the difference of electrostatic potentials in the entire 3-dimensional physical space. The object function is thus rotationally invariant with respect to the molecular orientation and varies smoothly with respect to molecular geometric fluctuations. Compared with commonly employed methods such as the Merz−Singh−Kollman and CHELPG schemes, this new method, while possessing comparable accuracy, shows greatly improved numerical stability with respect to the molecular positions and geometries. The method can be used in the fitting of electrostatic potentials for the molecular mechanics force fields and also can be applied to the calculation of electrostatic polarizabilites of molecular or atomic systems.

The electrostatic interaction is the key component of the diverse intra- and intermolecule interactions. Accurate treatment of the electrostatic interactions is a fundamental problem in the theoretical study of the structural and dynamic properties of molecular systems.[1] The ultimate method to accurately treat the electrostatic interaction is of course quantum mechanics (QM), in which the spatial and temporal distributions of electrons are determined through the process of solving the Schrödinger or Kohn−Sham equation. However, accurate ab initio quantum mechanical calculations are currently too expensive to be applied to large molecules of general interests in long time scale simulations. Because of the great computational efficiency, molecular mechanics (MM) has become a very effective and popular routine in the simulation of molecular processes. In the molecular mechanics, the complicated interatomic interactions are decomposed in a straightforward fashion into different terms such as the covalent terms (describing bond, bond angle, torsion, and improper dihedral energies), the van der Waals interaction, and the electrostatic interactions. The latter is usually computed as the simple Coulombic interaction between point charges on atoms, a much simplified and of course not so accurate description.

Most terms of the MM force field can be determined quite accurately. The covalent terms can be determined from high-level QM calculations or high-resolution spectroscopic studies. In both cases the parameters are found to be to a large extent transferable between molecules. The repulsive and attractive terms of the van der Waals interaction can be either computed from extremely accurate QM calculations or fitted to the condensed phase properties.[2−6] The electrostatic force, however, is quite difficult to approximate. Right from the start, approximating the complicated quantum mechanical electron−electron interactions by simple Coulombic interactions between point charges naturally leads to the inaccuracy of the calculations. The problem is further complicated by the fact that the atomic charge is a term lacking a rigorous and consistent definition; as a result, assigning the atomic charges is a nontrivial problem. Even though, various schemes have been proposed to determine atomic charges, such as the Mulliken population analysis,[7,8] the natural bond orbital analysis,[9,10] the atoms in molecules analysis,[11] the distributed multipoles analysis (DMA),[1,12,13] the wavefunction mapping Class IV model,[14−17] the electrostatic potential expansion and analysis,[18−20] and, most importantly, the electrostatic potential (ESP) fitting.[21−23] The last method has proven to be very effective and has gained great popularity in the applications of MM simulations. The

---

* Corresponding author e-mail: weitao.yang@duke.edu.

Molecular Volume ESP fitting

J. Chem. Theory Comput., Vol. 3, No. 3, 2007 **1005**

idea of ESP fitting was first proposed by Momany[21] and has subsequently been implemented in different ways such as the Merz−Singh−Kollman scheme,[22−24] the CHELP scheme,[25] and the CHELPG scheme.[26] These schemes share the same basic idea, i.e., least-square fitting the electrostatic potentials at chosen grid points. Because different schemes employed different algorithms to construct and select the grid points, the results from different methods vary and show different numerical dependences to the QM level of theory and basis sets.[27] Nonetheless, it was shown that in general the charges from ESP fitting methods are less varied with respect to the change of molecular geometries, QM level of theory, and/ or basis sets, when compared with other methods based on population analysis.[27] The ESP charges are also consistent with the chemical intuition and reproduce well the electrostatic multipolar properties of the molecules. Therefore, the atomic charges generated from ESP fitting are widely used in the simulations with the MM force fields,[28,29] despite several well-acknowledged deficiencies such as the transferability of the charges in different molecules.[30]

Recently, the applications of the ESP charges have been extended beyond the MM force fields to the QM/MM simulations and the molecular polarizability calculations.[31−33] It was shown that the use of ESP charges can achieve similar accuracy to other methods such as DMA, with much less computational expense and complexity. However, it was also found that conventional ESP methods bear many numerical difficulties in the calculation of atomic polarizabilities. Specifically, ESP charges show a dependence on the orientation and position of the molecule and a strong sensitivity to the molecular geometry. To improve the quality and especially the numerical stability of ESP charges, an effective and robust way of approximating the QM electrostatic interactions by means of fitting the electrostatic potentials is desired.

We report here a new method to fit atomic ESP charges (and other multipoles) from QM calculations. While the fitting process is still carried out as a process of minimizing the difference between the electrostatic potentials from QM calculation and from atomic multipoles, we define the object function in the entire molecular volume space instead of discretely selected grid points surrounding the target molecule. The object function is thus rotationally invariant to the molecular orientation and varies smoothly upon the change of molecular geometry. As a result, with comparable numerical accuracy to existing ESP fitting methods, the new method shows great improvements on the numerical stability of the fitting results. The method will be very useful for future study of molecular polarization interactions, in a pure MM force field or in a combined QM/MM method.

Methods

We formulate the problem here as approximating the interaction between a QM molecule and a classical point charge. Given a molecular system with the electron densities solved as $\rho(\mathbf{r})$, without the consideration of the polarization effects, the electrostatic interaction between this molecule and a point charge q at position $\mathbf{r}_q$ is

$$E_{ele} = \int \frac{\rho(\mathbf{r})q}{|\mathbf{r} - \mathbf{r}_q|} d\mathbf{r} = q \int \frac{\rho(\mathbf{r})}{|\mathbf{r} - \mathbf{r}_q|} d\mathbf{r} = qV_{QM}(\mathbf{r}_q) \quad (1)$$

where $V_{QM}(\mathbf{r}_q)$ is the QM molecular electrostatic potential at the position $\mathbf{r}_q$. One should note that in this equation, the potential $V_{QM}(\mathbf{r}_q)$ is solely determined by the electron densities and thus a property of the QM molecule. From this equation it is also obvious that the exact interaction energy has to be calculated from the quantum mechanically solved electron densities which span the space wherever an electron populates. In principle, this implies that an integration over the entire space is necessary. However, since the electron densities decay very rapidly from atoms, the integration in fact need only be done over a limited space where electron densities possess non-neglectable values. This space is defined as the molecular volume space in the current work.

In general there are two routes to approximately calculate this electrostatic energy. In the first approach, one may approximate the QM electron densities by classical electrostatic multipoles centered at the atomic positions. In other words, one directly seeks an approximate expression for the electron density, such as the distributed multipole analysis method. In the second approach, one could approximate the QM potential $V_{QM}(\mathbf{r}_q)$ by the potential of a set of classical electrostatic multipoles. That is, given a QM potential in space, one asks the question which set of classical electrostatic multipoles can best mimic this potential. Surely for one point in space there are countless solutions to reproduce the potential at this position; whether the specific solution is useful or not depends on how well this solution also approximates the potentials at other points in space.

Assuming again the electrostatic multipoles are centered at the atomic positions $\mathbf{r}_a$, the question now becomes how to determine a set of multipoles $Q(\mathbf{r}_a)$ so that

$$V(Q(\mathbf{r}_\alpha),\mathbf{r}) \simeq V_{QM}(\mathbf{r}) \quad (2)$$

where $V(Q(\mathbf{r}_\alpha),\mathbf{r})$ is the potential from all the multipoles $Q(\mathbf{r}_a)$. The straightforward solution to this problem would be to solve the least-square minimization problem as

$$\min [\int (V(Q(\mathbf{r}_\alpha),\mathbf{r}) - V_{QM}(\mathbf{r}))^2 d\mathbf{r}] \quad (3)$$

where the integration runs over the molecular volume space where electron densities span.

One can in principle determine a set of multipoles by directly solving the minimization problem defined in eq 3, in which each point in the molecular volume space contributes equally to the object function. Intuitively, however, a modified form of this equation with different weights for different points in space is more suitable owing to two practical considerations. On one hand, for those points near an atom, the magnitudes of the electrostatic potential are usually much larger than those of the points distant from atoms; consequently, the potentials at these points will make the most of the contribution to the minimization process and numerically beat out other points. However, the near-atom space is less important than other space in simulations with classical MM force fields: the short-range interactions are usually represented by covalent interactions which generally

omit electrostatic interactions between near neighbors, while the classical Coulombic term is an important component for long-range interactions. On the other hand, for points very distant from atoms, the values of the electrostatic potential are very small so that they should also be weighted less. That is to say, we would like to give more weight to points within the medium range to atoms as usually defined in classical simulation and give less weight to points too close or too distant to the molecule. In analog to the picture of the structure of liquid, one could imagine we are weighting heavily the points within the first neighboring shell which is of course most important in the simulations.

Similar to the original ESP fitting method, the problem is now defined as the minimization of an object function of

$$\gamma = \int W(\boldsymbol{r})(V(Q(\boldsymbol{r}_\alpha),\boldsymbol{r}) - V_{QM}(\boldsymbol{r}))^2 d\boldsymbol{r} \qquad (4)$$

in which $W(\boldsymbol{r})$ is the weighting function. From the previous discussion, it appears that an ideal weighting function should approach zero at both small and large distances and exhibit a maximum between 3 and 5 Å which is typically 1.4−2.0 times of van der Waals radii; interatomic interaction between this distance is thought to be most important in molecular simulations. The original ESP[23,24] method can also be formulated in the same way, where $W(\boldsymbol{r})$ is in fact defined as a discontinuous function as

$$W(\boldsymbol{r}) = \begin{cases} 0 & |\boldsymbol{r} - \boldsymbol{r}_\alpha| < R_{small} \\ \delta(\boldsymbol{r} - \boldsymbol{r}_{grid}) & R_{small} \le |\boldsymbol{r} - \boldsymbol{r}_\alpha| \le R_{large} \\ 0 & |\boldsymbol{r} - \boldsymbol{r}_\alpha| > R_{large} \end{cases} \qquad (5)$$

in which the two cutoffs are chosen in the MSK method as[23,24]

$$R_{small} = 1.4R_{vdw}$$

$$R_{large} = 2.0R_{vdw} \qquad (6)$$

The discontinuity of the weighting functions in the common ESP fitting methods leads to a well-known problem in the ESP fitting process, namely the numerical instability of the change of the ESP charges with respect to a structural perturbation of the molecule.

To solve this numerical problem, we propose here a new weighting function that not only meets the requirements discussed above but also allows smooth changes with respect to geometries. The new weighting function is defined as

$$W(\boldsymbol{r}) = \exp[-\sigma(\log \tilde{\rho}(\boldsymbol{r}) - \log \rho_{ref})^2] \qquad (7)$$

in which $\tilde{\rho}(\boldsymbol{r})$ is the predefined electron densities, and ρ_{ref} is a reference electron density whose value is chosen in combination with σ to ensure the weighting function behaves as expected. The design of this weighting function is inspired by the fact that the atomic (and molecular) electron densities decay exponentially in space. By adjusting ρ_{ref} and σ, we are able to create a Gaussian-like function $W(\boldsymbol{r})$ which weights heavily on the points in the medium-range of the molecule.

In principle the exact QM electron density $\rho(\boldsymbol{r})$ can be used as $\tilde{\rho}(\boldsymbol{r})$ in the calculation of the weighting function.

However, this is not an optimal solution since $\rho(\boldsymbol{r})$ then depends on both the molecular geometry and the basis sets. The dependence to the basis sets thus complicates the problem if the derivatives of the weighting function with respect to the geometry are sought. To avoid this complication, we decide here to use the sum of the atomic electron densities as

$$\tilde{\rho}(\boldsymbol{r}) = \sum_A \rho_{atom}(\boldsymbol{r}_A,\boldsymbol{r}) \qquad (8)$$

where $\rho_{atom}(\boldsymbol{r}_A,\boldsymbol{r})$ is the atomic density of atom A which will be predetermined. Hence, eq 4 can now be evaluated by 3-D numerical integration which is readily available in many quantum chemistry programs.

With the object function and weighting function available to carry out the minimization process, we consider the simplest case where there is only one point charge on each atom. The potential is

$$V(Q(\boldsymbol{r}_\alpha),\boldsymbol{r}) = \sum_{\alpha=1}^{N} \frac{q_\alpha}{|\boldsymbol{r} - \boldsymbol{r}_\alpha|} \qquad (9)$$

in which q_α stands for the point charge on atom α. We also apply a constraint to ensure the charge of the molecule is constant

$$g = q_{tot} - \sum_{j=1}^{N} q_j = 0 \qquad (10)$$

With a Lagrange Multiplier for the constraint, the final object function is

$$z = \gamma + \lambda g = \int W(\boldsymbol{r})(V(Q(\boldsymbol{r}_\alpha),\boldsymbol{r}) - V_{QM}(\boldsymbol{r}))^2 d\boldsymbol{r} + \lambda(q_{tot} - \sum_{j=1}^{N} q_j) = 0 \quad (11)$$

The solution is determined by solving the $n+1$ linear equations

$$0 = \frac{\partial z}{\partial q_k} = -2 \int W(\boldsymbol{r})V_{QM}(\boldsymbol{r}) \frac{1}{|\boldsymbol{r} - \boldsymbol{r}_k|} d\boldsymbol{r} + 2 \int W(\boldsymbol{r}) \frac{1}{|\boldsymbol{r} - \boldsymbol{r}_k|} \sum_{j=1}^{N} \frac{q_j}{|\boldsymbol{r} - \boldsymbol{r}_j|} d\boldsymbol{r} - \lambda \quad (12)$$

and

$$0 = \frac{\partial z}{\partial \lambda} = q_{tot} - \sum_{j=1}^{N} q_j \qquad (13)$$

This task can be accomplished by many well-established techniques such as the Single Value Decomposition (SVD).

Obviously the results of this ESP fitting scheme critically depend on the choices of two parameters in eq 7: σ and ρ_{ref}. Values of the two parameters cannot be determined by simply minimizing the object function defined in eq 4 since one could easily generate small values for the object function with an ill-defined weighting function. Alternately, we determined here to use an indirect criterion to evaluate the

Molecular Volume ESP fitting

J. Chem. Theory Comput., Vol. 3, No. 3, 2007 **1007**

quality of the fitting. That is, given a specific conformation of a target molecule, we randomly place a probe charge around this molecule at reasonable distances (e.g., 2−7 Å) and compare the electrostatic energies calculated with eq 1 and with ESP charges. The former is regarded as the exact result, where the latter is an approximated result. The criterion is defined as

$$E_{RRMSD} = \sqrt{\frac{\sum_{i=1}^{N}\left(\frac{E_{ele,i} - E_{ESP,i}}{E_{ESP,i}}\right)^2}{N}} \quad (14)$$

where $E_{ele,i}$ is the exact electrostatic energy defined in eq 1, $E_{ESP,i}$ is the approximated electrostatic energy calculated with ESP charges, and N is the number of trials to place the probe charges. The relative-root-mean-square-deviation (RRMSD) of energies defined in this equation thus allows us to practically measure the quality of the electrostatic interactions computed with the ESP fitted charges, which is the essence of ESP fitting.

Of course, the best-fits of two parameters σ and ρ_{ref} will vary with specific molecules. To ensure the two parameters are transferable between different molecules, the variation of fitting results with respect to the changes of σ and ρ_{ref} has to be evaluated; the variations of the fitting results in different molecules also have to be examined to identify the range of the two parameters so that they are suitable for different molecules. After that, σ and ρ_{ref} can be used in the future study of other molecules.

Computation Details

The program Gaussian03 was employed for this study.[34] We modified the link l602 which was designed to carry out population analysis. We first constructed an integration grid with standard 3-D integration method used in Density Functional calculations[35] and then computed electrostatic potentials on the grid points with the converged electron densities. The fitting procedure was accomplished by solving eq 11. Unless specifically stated, we did not apply the constraints to reproduce the molecular dipole in the fitting process.

To determine the atomic electron densities for calculating eq 8, we carried out a UHF/aug-cc-pVQZ calculation for different atoms in the first three rows in the periodic table. Based on the well-known fact that the atomic electron densities are exponentially decaying in a piecewise manner,[36,37] we fitted the atomic electron densities to

$$\rho_{atom}(r) = \sum_{i=1}^{N} A_i \exp(-B_i r) \quad (15)$$

where N is the number of components decaying exponentially in each atom: $N = 1$ for H and He, 2 for Li to Ne, 3 for Na to Ar, and 4 for K to Kr. After the coefficients A_i and B_i were determined for each element, they were tabulated and stored in the program for the calculation of the weighting function. It should be noted here that the exact solution of
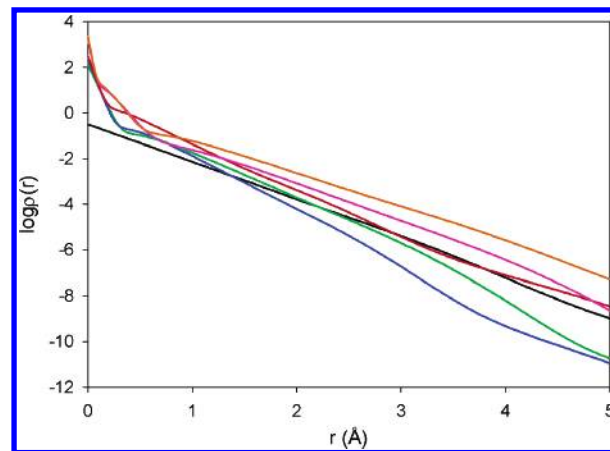

Figure 1. Atomic electron densities (in atomic units) of the elements H (black), C (green), N (blue), O (red), P (pink), and S (orange). The densities were computed by the UHF/aug-cc-pVQZ method.

atomic electron densities is not required because the uncertainties will be absorbed late in the process of determining σ and ρ_{ref}.

All QM calculations for the electrostatic potentials and electron densities were carried out at the level of B3LYP/6-31G*.[38,39] All molecular geometries were optimized at the same level of theory prior to the ESP analysis and fitting. In current work, the standard pruned (75,302) grid implemented in Gaussian was chosen for the purpose of 3-D integration. The Merz−Singh−Kollman ESP fitting method implemented in Gaussian was used with a default setup, whenever it was used for the purpose of comparison.

To evaluate the quality of the fitting and the values of the two parameters, σ and ρ_{ref}, in the weighting function, we calculated the RRMS energy deviation defined in eq 14. In the calculation, we placed the molecule in the center and randomly placed a probe charge in the neighborhood of the molecule. For obvious reasons the probe change would not be included in the calculation if it was too close or too far to the molecule. In general, we made selections to the position of the probe charge so that it was within 7 Å to at least one atom of the molecule and was not closer than r_{vdW} + 0.2 Å to any atom of the molecule. This selection ensures us to have enough samples in the space region that are most important in normal simulation studies. The standard van der Waals radii were used. Specifically, they are 1.2, 1.70, 1.55, 1.52, and 1.80 Å for the elements H, C, N, O, and S, respectively.[40] For each set of the calculations, 40,000 effective random positions were generated for the probe charge to ensure the convergence of the RRMS energy deviation (data not shown).

Results

Exponentially Decaying of Atomic Density. In the current work, the atomic densities were computed for the elements H, Li−F, and Na−Ar. Some other elements were not included in this calculation because the chosen basis sets were not available for them. The computed atomic densities for the elements H, C, N, O, P, and S are shown in Figure 1, which clearly demonstrates the exponential decaying

***Table 1.*** Multiple Exponential Components of Atomic Electron Densities

element	A	B
H ($N = 1$)	0.384137961	3.90762643
C ($N = 2$)	166.591448	29.0603279
	3.23010126	5.01709331
N ($N = 2$)	256.609200	31.2114908
	2.58989432	5.45471548
O ($N = 2$)	243.630909	26.3836036
	2.53736474	4.29335839
P ($N = 3$)	2282.83071	73.7103367
	155.142338	15.6986998
	1.82194667	3.38628928
S ($N = 3$)	2736.19302	78.9192252
	206.867393	17.4500522
	2.78312612	3.51974385

behavior of the atomic electron densities. This is in agreement with the electronic shell structure where the decay also exhibits piecewise feature. As a result, the atomic electron densities can be fitted into multiple components of decaying exponentials. The fitting results for H, C, N, O, P, and S are reported in Table 1. One more important observation in Figure 1 is that the atomic electron densities of different elements tend to have similar values at the distances near van der Waals radii, even though they could be significantly different at short and long distances. This observation, plus the exponentially decaying feature of electron densities, ensures that a weighting function can be satisfactorily constructed according to eqs 7 and 8 with large weights for regions close to the vdW isosurface of a molecule.

Determination of Parameters σ and ρ_{ref}. The two parameters, σ and ρ_{ref}, in the weighting function, have to be determined in such a way that they are suitable for fitting electrostatic potentials even for different chemical molecules. For this purpose, a 2-dimensional scanning was performed for the two parameters in broad ranges, for different molecules including ethanol (CH_3CH_2OH), methylammonium ion ($CH_3NH_3^+$), methylamine (CH_3NH_2), CH_3COO^-, CH_3COOH, DMSO ($CH_3S(O)CH_3$), and N-methylethanamide (NMA) ($CH_3CONHCH_3$).

Without loss of generality, we show the 2-D scanning results for ethanol and DMSO in Figure 2. It is clear when σ is too small or ρ_{ref} is too large, the quality of fitting is rather poor. The best fit was obtained when the value of ρ_{ref} is about $10^{-4}-10^{-5}$ ($-10 \le \ln\rho_{ref} \le -8$) with reasonable values of σ. In fact, the quality of the ESP fitting is not significantly changing over a broad range of σ and ρ_{ref}. This observation becomes a quite useful result. As long as the values of σ and ρ_{ref} are within this region, they could be used in the fitting of different molecules without the loss of accuracy.

Based on these observations, we chose the values of σ and $\ln\rho_{ref}$ to be 0.8 and -9, respectively, for the rest of the calculations. For these chosen values, we also computed E_{RRMSD} defined in eq 14 for different molecules as a test of the accuracy of the current method. The results are shown in Table 2. In general, the current method performs as well as the popular Merz–Singh–Kollman method. It must be
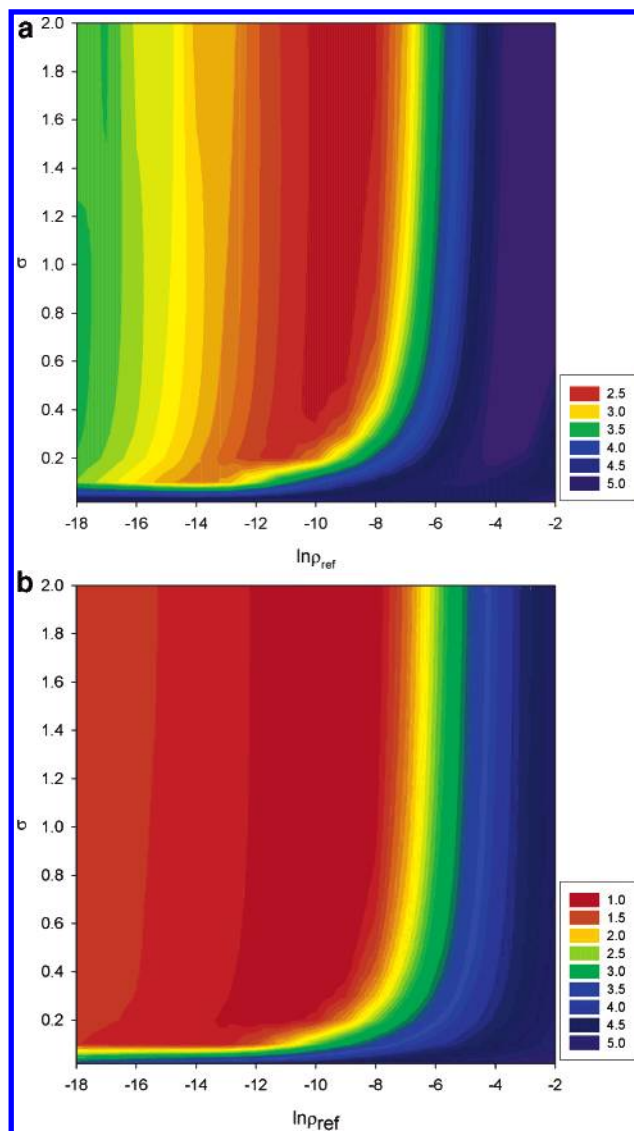
Figure 2. Variation of E_{RRMSD} with the change of $\ln\rho_{ref}$ and σ: (a) molecule ETOH and (b) molecule NMA. ρ_{ref} is in atomic units.

Table 2. E_{RRMSD} Calculated for Various Molecules

molecule	MSK	CHELPG	current method[a]
CH_3CH_2OH	13.7	21.4	10.6 (4.16)
CH_3NH_2	19.3	19.4	18.7 (1.08)
$CH_3NH_3^+$	.00673	0.00672	0.00677 (0.00146)
CH_3COO^-	.00519	0.00482	0.00513 (0.00333)
CH_3COOH	2.20	5.93	2.19 (1.08)
DMSO	11.3	7.47	6.46 (1.59)
NMA	5.00	3.99	2.48 (1.02)

[a] Data in the parentheses are E_{RRMSD} calculated including also a point dipole on each atom of the molecule. In this case, the values of the two parameters, ρ_{ref} and σ, are the same as those of the point charge only without further optimization.

reminded since E_{RRMSD} is only one criterion of many possibilities, it should never be used to draw qualitative conclusions.

The basis set dependence of the current method was also examined here. In Table 3 we show the atomic charges of the methanol molecule fitted to ESP calculated with B3LYP and different basis sets. It is shown that the charges tended

Molecular Volume ESP fitting

J. Chem. Theory Comput., Vol. 3, No. 3, 2007 **1009**

Table 3. Atomic Charges of the Methanol Molecule Determined with B3LYP and Different Basis Sets

	C	O	H[a]	H[b]	H[c]
STO-3G	0.0882	−0.4509	0.3062	0.0193	0.0179
3-21G	0.2076	−0.6367	0.4116	0.0121	−0.006605
6-31G	0.3262	−0.7258	0.4448	−0.0066	−0.0320
6-31G*	0.1809	−0.6007	0.3902	0.0193	−0.008944
6-311G*	0.1856	−0.6376	0.4154	0.0225	−0.008504
cc-pVDZ	0.2061	−0.5644	0.3642	0.00777	−0.0214
cc-pVTZ	0.1601	−0.5711	0.3695	0.0252	−0.009039
cc-pVQZ	0.1792	−0.5809	0.3702	0.0227	−0.0140
cc-pV5Z	0.1954	−0.5881	0.3707	0.0202	−0.0185
aug-cc-pVDZ	0.1985	−0.5856	0.3672	0.0194	−0.0188
aug-cc-pVTZ	0.2009	−0.5865	0.3684	0.0189	−0.0205
aug-cc-pVQZ	0.1995	−0.5862	0.3683	0.0192	−0.0200
aug-cc-pV5Z	0.2000	−0.5863	0.3683	0.0191	−0.0201

[a] Hydrogen of the hydroxyl group. [b] Methyl hydrogen atom trans to the hydroxyl group. [c] Methyl hydrogen atom gauche to the hydroxyl group.

to converge to similar values when a large basis set such as cc-pVXZ or aug-cc-pVXZ is used. A similar observation has been reported before.[27]
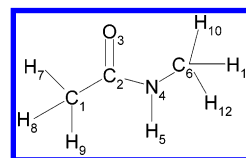
Molecular Orientation Dependence of the Fitting Results. One important advantage of the current charge fitting scheme is that the object function is defined in molecular volume space, which is not changed by the global movements of the molecule. As a result, the fitting results will not be changed by the rotation or translation of the molecule, within the limits of numerical accuracy. By contrast, conventional charge fitting schemes, such as the Merz−Singh−Kollman and the CHELPG methods, always bear such a difficulty that the fitting grid points are changing upon the rotation of the molecules. Consequently, the results of those methods are numerical unstable and fluctuate significantly. This deficiency has limited broader applications of such methods to problems which require stable fitting results.

To compare the dependences of the results on the global rotation of the molecule, we randomly translated and rotated the molecule N-methylethanamide (NMA) ($CH_3CONHCH_3$) 1000 times and for each orientation calculated the charges by different methods. The root-mean-square-fluctuations (RMSF) for each atomic charge were calculated and reported in Table 4. As expected, the RMSF of the charges from the CHELPG method are about the half of those in the MSK method, indicating that the CHELPG method has slightly weaker dependence on the molecular orientation and thus better numerical stability.[26] Importantly, the RMSF of the charges in the current method is on average about 20−30 times smaller than those in the MSK method. Therefore, the current method has the best numerical stability of the methods compared here. It should be noted that increasing the number of grid points will improve the numerical stability of both methods. However, the improvements are relatively moderate in the MSK and CHELPG methods but quite significant in the current method. In the current method, switching to a fine (99,590) grid for numerical integration will reduce the RMSF by more than an order of magnitude with the number of points only increased to 2.5 times. In

Table 4. rms Fluctuations of the Atomic Charges Determined by Different Methods

atom	MSK[a]	CHELPG	current method[b]
C1	0.02374 (0.00878)	0.0127	0.000935 (0.000114)
C2	0.01095 (0.00446)	0.00501	0.000336 (0.000049)
O3	0.00260 (0.00113)	0.00102	0.0000983 (0.000016)
N4	0.00805 (0.00389)	0.00424	0.000546 (0.000052)
H5	0.00274 (0.00125)	0.00100	0.000210 (0.000014)
C6	0.01899 (0.00851)	0.0101	0.000914 (0.000095)
H7	0.00625 (0.00236)	0.00332	0.000265 (0.000031)
H8	0.00586 (0.00217)	0.00332	0.000254 (0.000031)
H9	0.00589 (0.00220)	0.00316	0.000254 (0.000031)
H10	0.00495 (0.00219)	0.00265	0.000254 (0.000025)
H11	0.00520 (0.00233)	0.00265	0.000252 (0.000024)
H12	0.00524 (0.00230)	0.00265	0.000252 (0.000024)

[a] ESP charges were fitted with default grids in Gaussian03. The data in the parentheses were fitted with a grid setup of 6 layers and 2 points per unit area, which yields a number of grid points approximately 3 times the default setup. [b] ESP charges were fitted with (75 302) a DFT grid. The data in the parentheses were fitted with (99 590) a DFT grid whose number of grids is approximately 2.5 times the former.

Figure 3. Structure of the molecule N-methylethanamide (NMA).

contrast, the RMSF of the MSK method is only reduced to half when the number of points is increased 3 times (Table 4).

Molecular Geometry Dependence of the Fitting Results. It is well-known that the molecular electron distribution changes with the varying of the molecular geometry. The direct consequence of this phenomenon is that the atomic charges (dipoles and quadrupoles too), calculated by fitting the electrostatic potential or by discrete multipole analysis,[41] also change with a perturbation of the molecular geometries. The correlation between the atomic charges and the molecular geometries in fact provides an opportunity for designing of a polarizable force field which allows charge transfer within a molecule. For this purpose, the computed charges must be differentiable with respect to the molecular geometry, no matter whether the differentiation is carried out numerically or analytically. When analytical differentiation is difficult to achieve, the numerical method becomes the only feasible way. In such a case, the numerical uncertainty of the charges must be small enough so that its effect can be neglected in the process of numerical differentiation, otherwise, the numerical results cannot be trusted.

We carried out analysis again on the NMA molecule (Figure 3). The bond C_2−N_4 is defined as the pivot for intramolecular rotation with the relative geometries of the left and right sides of the molecule fixed. The dihedral angle O_3−C_2−N_4−H_5 was rotated for 360 degrees with 4 degrees per step. The atomic charges were computed with the MSK
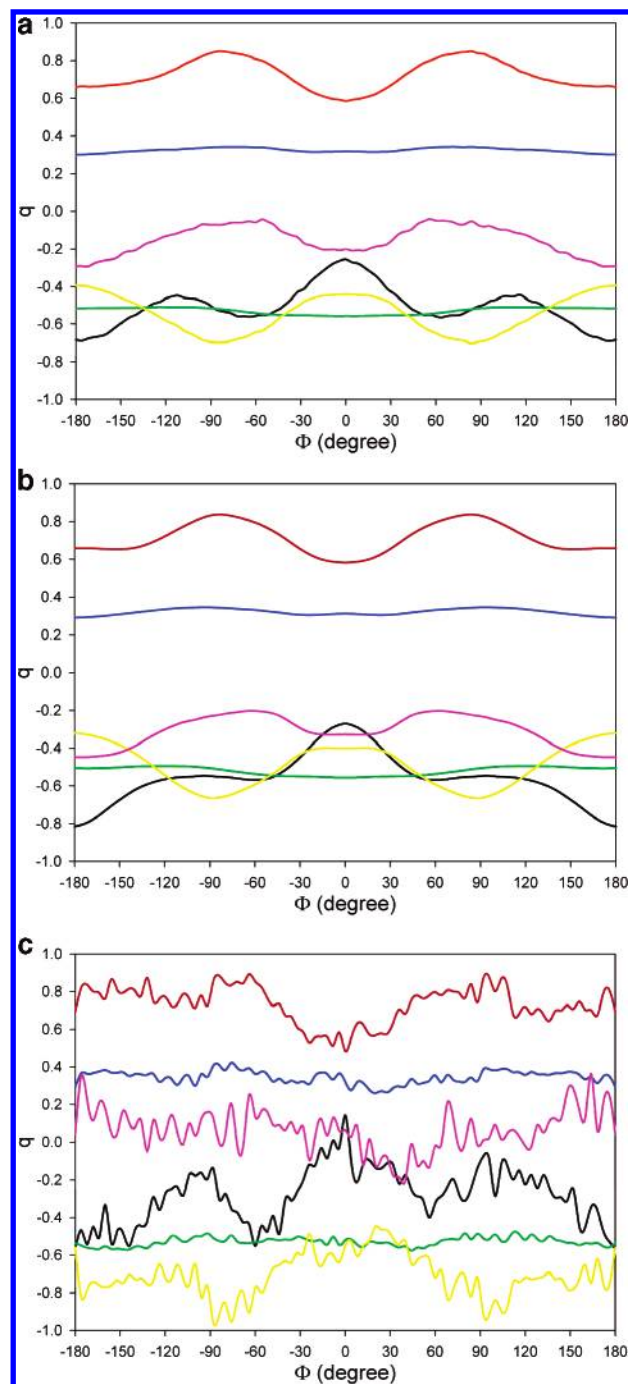
Figure 4. Variations of the ESP charges of the molecule NMA with respect to the twisting of the dihedral angle O3−C2-N4−H5. For the clarity of illustration, the charges of the hydrogen atoms of the methyl groups are not shown: (a) the Merz−Singh−Kollman method; (b) the current method; and (c) the current method with the harmonically restrained charge model (weight 0.005). The fitted charges are shown for atoms C1 (black), C2 (red), O3 (blue), N4 (yellow), H5 (green), and C6 (pink).

method and current method for each rotated conformation. The results are shown in Figure 4.

Not surprisingly, both methods consistently show the same behaviors for the changes of the atomic charges with respect to the rotation of the dihedral angle. The shape and the magnitudes of both methods share a great resemblance. This

agreement demonstrates the validity of current method. More importantly, however, the variations of the atomic charges show great discontinuity in the MSK method, while they are smoothly changing in the current method. This difference marks the most important advantage of the current method in the calculation of the atomic polarizablities since it allows numerical differentiation with respect to geometrical perturbations to be carried out with high numerical precision.

Discussion

The motivation to utilize ESP charges in the molecular simulations is to approximate the electrostatic interaction in an atomic pairwise manner, instead of explicitly dealing with electron densities which is too complicated and too expensive. If one only cares about the quality of simulation with nonpolarizable MM force fields, the only quantity that matters here is how well the ESP charges mimic the QM electrostatic potentials. For this purpose, many methods have been developed on the basis of fitting ESP at discrete grid points in selected neighborhood regions of the target molecule. The successes of the MM simulations using fitted atomic charges strongly demonstrate the validity and usefulness of those ESP fitting methods.

When derivatives of the ESP charges are sought for the calculation of the atomic/molecular polarizability, those methods displayed numerical instability due to the nonsmooth and noncontinuous nature of the object function of fitting. Apparently, this discontinuity arises from the algorithm to construct and select the grid points for fitting. To overcome this problem, the grid points must be rotationally invariant with respect to the molecular orientation and continuously change with respect to geometrical perturbation.

From this perspective, the current method completely overcomes this problem by implementing a new algorithm to generate rotationally invariant grids, specifically the 3-D integration grids used in quantum chemistry calculations. Furthermore, there is one additional advantage for utilizing a complete set of grids in the entire molecular volume space. That is, this improves the quality of fitting when higher order multipoles are added to atomic sites. When the order of the electrostatic multipoles increases, the magnitudes of the electrostatic potential decay rapidly with a decreased order to the distance. For example, the ESP is of order r^{-1} for a monopole, r^{-2} for a point dipole, and r^{-3} for a point quadrupole. This fact suggests that for multipoles it is important to include grid points near atoms; otherwise, the results will not be as reliable as monopoles.

As suggested by the results of E_{RRMSD} (Table 2), the fitting qualities of several ESP fitting algorithms, including the current one, are comparable to each other. They all suffer the same imperfection because the implementation of the electrostatic multipoles is limited to monopole. As shown in Table 2, the deviations of electrostatic energy calculation are rather large for many molecules. Detailed inspection of the results indicates the large deviation is mostly contributed by the region closely surrounding the target molecule. In those regions, it is well-known that atomic point charges reproduce poorly the quantum mechanical ESP for two reasons. One is the complicated electrostatic potential from

Molecular Volume ESP fitting

J. Chem. Theory Comput., Vol. 3, No. 3, 2007 **1011**

Table 5. Comparison of Atomic Charges of NMA Molecule Determined with and without Restraints

atom	free	restraint[a]
C1	−0.7218	−0.0864
C2	0.6912	0.2193
O3	−0.5207	−0.4040
N4	−0.4359	−0.1971
H5	0.3120	0.2338
C6	−0.2672	−0.1590
H7	0.1602	0.0120
H8	0.2049	0.0648
H9	0.2048	0.0647
H10	0.1246	0.0769
H11	0.1239	0.0875
H12	0.1239	0.0875

[a] The harmonic restraint proposed in the original paper[42] was employed with a weight of 0.005.

the complex electron distribution in the close neighborhood of the molecule. Another reason is the charge penetration effect. To improve the fitting quality, higher order multipoles could be added to the atomic positions. As an example, the inclusion of atom-centered point dipoles will substantially improve the quality of fitting results for the current method (Table 2). Of course, the addition of a dipolar term will increase the computational cost; such a cost might be still affordable for MM force field based simulations, with the steady improvement of computers' speed.

Like many other methods, the current algorithm also bears many problems common to the ESP fitting approach. Obviously, fitting a set of electrostatic multipoles to the electrostatic potential from accurate quantum mechanical calculation is inherently a poorly posed question. Despite the subtle differences, all ESP fitting methods rely on the definition of an object function, often the rmsd of the electrostatic potentials on selected grid points. Given the complexity of the ESP surface and the drastic variation of the values on the ESP surface, it is not surprising that the fitting results are sensitive to many factors such as the choice of the grid points, the basis sets of the QM calculations, and the geometry and orientation of the molecules.

One such well-known example of the deficiencies of the ESP fitting approach is the large charges on the buried atoms of a molecule. In this case, the neighborhood spaces of those atoms are often occupied by other atoms too, and the ESP in those regions make little contribution to the object function of the fitting process. The immediate consequence is that the fitting quality is insensitive to the charges of those buried atoms, and it leads to large and numerically unstable charges on those atoms. To overcome such a difficulty, an "ad hoc" yet very effective approach has been proposed as to restrain the magnitude of the charges in the fitting process.[42,43] This restrained minimization approach can also be applied to our method which generated a set of charges closer to common intuitions (Table 5).

One particular motivation for the development of this new method is to improve the numerical stability of the ESP fitting process, specifically in the calculation of atomic polarizabilities. The results shown in Figure 4 clearly indicated that our new method is indeed providing better

results. The dependence of the charges on the molecular geometries, in this case the dihedral angle of the peptide plane, was discovered before and was proposed as a new standpoint for designing polarizable force fields.[41] Some may argue that such a geometrical dependence could be an artifact from the burying of some atoms. Especially given the magnitude of the fluctuation of the charges, one may question if such a fluctuating charge can be used in the MM force fields. Answering those questions requires a comprehensive analysis on the ESP fitting approach. Certainly as we discussed before, the burying of some atoms will influence the fitting process and ultimately will affect the fitted atomic charges. In the case of the twisting of the peptide plane here, however, the attempt to remove this effect by means of the RESP approach is ineffective. By comparing the fitting results with and without harmonic restraints (Figure 4b,c), one clearly sees that the inclusion of restraint indeed lowers the magnitude of the charge on the carbon atoms (C1 and C6) for the dihedral angle close to zero which is the normal stable conformation of this molecule. Nonetheless, the charge dependence on the dihedral angle is retained regardless of the presence of the restraints. This result suggests that the variation of the buried extent of heavy atoms is not the major source of the geometrical dependence of the fitted charges. On the other hand, even though the charges of the carbon atoms are closer to "chemical intuition" when the molecule is at the stable conformations, the overall bumpy fluctuations of the fitted RESP charges reminds us that great caution should be taken when the RESP type of "ad hoc" approximations are used. The increased roughness implies that the restraint might alter the numerical behavior of the object function in an unpredicted and unstable way so that a mathematically more elegant solution is desired (T. Heaton-Burgess, to be published). Furthermore, from our point of view, even though the fitting results without restraints may be sometimes beyond our "chemical intuitions", as long as the overall electrostatic potential of the target molecule is well approximated, the result will be accurate and will capture the essence of the electrostatic interactions.

In summary, with much improved numerical stability, the current method will make an important contribution to the calculation of the polarizability type of properties and to the development of the next generation of molecular mechanics force fields.

Instructions on how to modify the Gaussian code to implement this new algorithm can be found at http://www.chem.duke.edu/∼yang/software.htm.

References

(1) Stone, A. J. *The Theory of Intermolecular Forces*; Clarendon Press: Oxford, U.K., 1996.

(2) Williams, D. E. Nonbonded potential parameters derived from crystalline hydrocarbons. *J. Chem. Phys.* **1967**, *47*, 4680−4684.

Hu et al.

(3) Scheraga, H. A. Calculations of conformations of polypeptides. *Adv. Phys. Org. Chem.* **1968**, *6*, 103−184.

(4) Ferro, D.; Hermans, J. Semiempirical energy calculations on model compounds of polypeptides. Crystal structures of DL-acetylleucine N-methylamide and DL-acetyl-n-butyric acid N-methylamide. *Biopolymers* **1972**, *11*, 105−117.

(5) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem.* **1998**, *B 102,* 3586−3616.

(6) Jorgensen, W. L. OPLS Force Fields. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Ed.; Wiley: New York, 1998; Vol. 3, pp 1986−1989.

(7) Mulliken, R. S. Electronic population analysis on LCAO-MO [linear combination of atomic orbital-molecular orbital] molecular wave functions. I. *J. Chem. Phys.* **1955**, *23*, 1833−1840.

(8) Mulliken, R. S. Criteria for the construction of good self-consistent-field molecular orbital wave functions, and the significance of L.C.A.O.M.O. population analysis. *J. Chem. Phys.* **1962**, *36*, 3428−3439.

(9) Reed, A. E.; Curtiss, L. A.; Weinhold, F. Intermolecular interactions from a natural bond orbital, donor-acceptor viewpoint. *Chem. Rev.* **1988**, *88*, 899−926.

(10) Reed, A. E.; Weinstock, R. B.; Weinhold, F. Natural population analysis. *J. Chem. Phys.* **1985**, *83*, 735−746.

(11) Bader, R. F. W. *Atoms in Molecules A Quantum Theory*; Oxford University Press: New York, 1990.

(12) Stone, A. J. Distributed multipole analysis, or how to describe a molecular charge distribution. *Chem. Phys. Lett.* **1981**, *83*, 233−239.

(13) Stone, A. J.; Alderton, M. Distributed multipole analysis methods and applications. *Mol. Phys.* **1985**, *56*, 1047−1064.

(14) Winget, P.; Thompson, J. A.; Xidos, J. A.; Cramer, C. J.; Truhlar, D. G. Charge Model 3: A class IV charge model based on hybrid density functional theory with variable exchange. *J. Phys. Chem. A* **2002**, *106*, 10707−10717.

(15) Li, J. B.; Williams, B.; Cramer, C. J.; Truhlar, D. G. A class IV charge model for molecular excited states. *J. Chem. Phys.* **1999**, *110*, 724−733.

(16) Li, J. B.; Zhu, T. H.; Cramer, C. J.; Truhlar, D. G. New class IV charge model for extracting accurate partial charges from wave functions. *J. Phys. Chem. A* **1998**, *102*, 1820−1831.

(17) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. Class-Iv Charge Models - a New Semiempirical Approach in Quantum-Chemistry. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 87−110.

(18) Simmonett, A. C.; Gilbert, A. T. B.; Gill, P. M. W. An optimal point-charge model for molecular electrostatic potentials. *Mol. Phys.* **2005**, *103*, 2789−2793.

(19) Sigfridsson, E.; Ryde, U. Comparison of methods for deriving atomic charges from the electrostatic potential and moments. *J. Comput. Chem.* **1998**, *19*, 377−395.

(20) Lee, T.-S.; York, D. M.; Yang, W. A new definition of atomic charges based on a variational principle for the electrostatic potential energy. *J. Chem. Phys.* **1995**, *102*, 7549−7556.

(21) Momany, F. A. Determination of Partial Atomic Charges from Ab Initio Molecular Electrostatic Potentials. Application to Formamide, Methanol, and Formic Acid. *J. Phys. Chem.* **1978**, *82*, 592−601.

(22) Cox, S. R.; Williams, D. E. Representation of the molecular electrostatic potential by a net atomic charge model. *J. Comput. Chem.* **1981**, *2*, 304−323.

(23) Singh, U. C.; Kollman, P. A. An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* **1984**, *5*, 129−145.

(24) Besler, B. H.; Merz, K. M., Jr.; Kollman, P. A. Atomic charges derived from semiempirical methods. *J. Comput. Chem.* **1990**, *11*, 431−439.

(25) Chirlian, L. E.; Francl, M. M. Atomic charges derived from electrostatic potentials: A detailed study. *J. Comput. Chem.* **1987**, *8*, 894−905.

(26) Breneman, C. M.; Wiberg, K. B. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J. Comput. Chem.* **1990**, *11*, 361−373.

(27) Martin, F.; Zipse, H. Charge distribution in the water molecule - A comparison of methods. *J. Comput. Chem.* **2005**, *26*, 97−105.

(28) Cornell, W. D.; Cieplak, P.; Bayly, C.; Gould, I. R.; Merz, K. M. J.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins and nucleic acids. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(29) MacKerell, J. A. D.; Brooks, B.; Brooks, I. C. L.; Nilsson, L. R. B.; Won, Y.; Karplus, M. CHARMM: The energy function and its parameterization with an overview of the program. In *The Encyclopedia of Computational Chemistry*; P. v. R. S. e., Ed.; John Wiley & Sons: Chichester, 1998; Vol. 1, pp 271−277.

(30) Chipot, C.; Angyan, J. G.; Ferenczy, G. G.; Scheraga, H. A. Transferable net atomic charges from a distributed multipole analysis for the description of electrostatic properties: a case study of saturated hydrocarbons. *J. Phys. Chem.* **1993**, *97*, 6628−6636.

(31) Morita, A.; Kato, S. Ab Initio Molecular Orbital Theory on Intramolecular Charge Polarization: Effect of Hydrogen Abstraction on the Charge Sensitivity of Aromatic and Nonaromatic Species. *J. Am. Chem. Soc.* **1997**, *119*, 4021−4032.

(32) Lu, Z.; Yang, W. Reaction path potential for complex systems derived from combined ab initio quantum mechanical and molecular mechanical calculations. *J. Chem. Phys.* **2004**, *121*, 89−100.

(33) Anisimov, V. M.; Lamoureux, G.; Vorobyov, I. V.; Huang, N.; Roux, B.; Alexander, D.; MacKerell, J. Determination of Electrostatic Parameters for a Polarizable Force Field Based on the Classical Drude Oscillator. *J. Chem. Theory Comput.* **2004**, *1*, 153−168.

Molecular Volume ESP fitting

J. Chem. Theory Comput., Vol. 3, No. 3, 2007 **1013**

(34) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; 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.; Gonzalez, C.; Pople, J. A. *Gaussian 03, C.02*; Gaussian, Inc.: Wallingford, CT, 2004.

(35) Becke, A. D. A multicenter numerical integration scheme for polyatomic molecules. *J. Chem. Phys.* **1988**, *88*, 2547–2553.

(36) Wang, W.-P.; Parr, R. G. Statistical atomic models with piecewise exponentially decaying electron densities. *Phys. Rev. A* **1977**, *16*, 891–902.

(37) Weinstein, H.; Politzer, P.; Srebrenik, S. A misconception concerning the electronic density distribution of an atom. *Theor. Chim. Acta* **1975**, *38*, 159–163.

(38) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, 37, 785-.

(39) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(40) Bondi, A. van der Waals Volumes and Radii. *J. Phys. Chem.* **1964**, *68*, 441–451.

(41) Koch, U.; Stone, A. J. Conformational dependence of the molecular charge distribution and its influence on intermolecular interactions. *J. Chem. Soc., Faraday Trans.* **1996**, *92*, 1701–1708.

(42) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.

(43) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. Application of RESP Charges To Calculate Conformational Energies Hydrogen Bond Energies, and Free Energies of Solvation. *J. Am. Chem. Soc.* **1993**, *115*, 9620–9631.