ARTICLES

A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model

Christopher I. Bayly,† Piotr Cieplak,‡ Wendy D. Cornell,§ and Peter A. Kollman*

Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, California 94143

Received: July 6, 1993®

We present a new approach to generating electrostatic potential (ESP) derived charges for molecules. The major strength of electrostatic potential derived charges is that they optimally reproduce the intermolecular interaction properties of molecules with a simple two-body additive potential, provided, of course, that a suitably accurate level of quantum mechanical calculation is used to derive the ESP around the molecule. Previously, the major weaknesses of these charges have been that they were not easily transferable between common functional groups in related molecules, they have often been conformationally dependent, and the large charges that frequently occur can be problematic for simulating intramolecular interactions. Introducing restraints in the form of a penalty function into the fitting process considerably reduces the above problems, with only a minor decrease in the quality of the fit to the quantum mechanical ESP. Several other refinements in addition to the restrained electrostatic potential (RESP) fit yield a general and algorithmic charge fitting procedure for generating atom-centered point charges. This approach can thus be recommended for general use in molecular mechanics, molecular dynamics, and free energy calculations for any organic or bioorganic system.

Introduction

The concept of atomic charge is fundamental to all of chemistry. Atomic charges are enormously powerful in understanding chemical reactivity and physical properties. In addition to such qualitative conceptual uses, their values are of quantitative importance in simulating physical properties of condensed-phase matter. Molecular mechanics, molecular dynamics, and Monte Carlo calculations rely on these charges, and the accuracy with which these methods can derive physical properties of solids, liquids, and solutions is critically dependent on the atomic charges used. Unfortunately, there is no "true" charge, given the distributed nature of the electrons around the nucleus in a molecule; thus chemists have no choice but to construct atomic charges that fulfill the qualitative or quantitative needs for this crucial aspect of chemical understanding.

There is considerable debate about which is the best way to determine atomic charges to simulate condensed-phase properties. A number of approaches have been tried. Carrying out a large number of *ab initio* calculations on molecular complexes and fitting these to a potential function that includes electrostatic and van der Waals interactions has been championed by the Clementi group.[1] This method, however, suffers from the magnitude of calculations required and from the two-body nature of the potential. In principle, one can also carry out calculations on clusters of more than two molecules and fit these to many-body effects, but then the computational burden becomes even more extreme.

At the other end of the spectrum is an empirical approach, where the charges are varied to fit the properties of crystals[2] or liquids.[3] The OPLS[3] nonbonded parameters, derived by fitting the enthalpy of vaporization and density of liquids determined by Monte Carlo calculations, have proven to be transferable and powerful in simulating condensed-phase properties of matter.[3] Their weakness is mainly in their empirical nature and the fact that any new fragment requires a Monte Carlo calculation on an appropriate ligand.

Approaches in between these two extremes involve deriving atomic charges for molecules using empirical,[4] semiempirical,[5] or *ab initio* quantum mechanical approaches[6-9] for individual molecules. The empirical or semiempirical approaches are often fast enough that one can generate charges for large databases of molecules. The charges generated are often surprisingly accurate, when compared to those derived from the more accurate *ab initio* method, but there are often specific molecules that are poorly handled. Generating the charges using *ab initio* methods is quite time-consuming but can give the most accurate representation of the charge distribution, provided a suitable atomic basis set is used. This method also has the virtue that the representation of the atomic charges can be successively improved, as the representation of the basis set improves. Although the *ab initio* derived charges fluctuate significantly with small basis sets, after one reaches a basis set of 6-31G* quality,[10] the electrostatic potential is close to convergent with respect to improvements in the basis set.

If one wishes to derive atomic charges using semiempirical or *ab initio* calculations, one can do so by fitting the charges to reproduce the electrostatic potential (ESP) calculated at a large number of grid points around the molecule. This approach was first used by Momany,[6] further refined by Cox and Williams,[7] and then applied to a large number of systems by Singh and Kollman[8] and Weiner *et al.*[11,12] This approach has the virtue that such charges reproduce the quantum mechanically determined multipole moments well and also optimally reproduce the intermolecular interactions with surrounding molecules. This is an essential ingredient for simulations of complex condensed-phase systems. A number of studies have clearly shown the superiority of such charges compared to, for example, those derived from Mulliken population analyses.[8] Even ESP derived charges will, however, be basis set dependent, albeit nowhere near as sensitive to basis set as Mulliken population analysis. Thus, we

† Present address: Merck Frosst Canada Inc., C.P. 1005 Pointe Claire, Dorval, Quebec, Canada H9R 4P8.
‡ Permanent address: Department of Chemistry, University of Warsaw, Pasteur 1, 02-093 Warsaw, Poland.
§ Graduate Group in Biophysics.
® Abstract published in *Advance ACS Abstracts*, September 1, 1993.

0022-3654/93/2097-10269$04.00/0

have argued that, for optimal reproduction of biomolecular properties in aqueous solution using additive potentials, a 6-31G* basis set is an excellent choice. This basis set overestimates the polarity of molecules approximately as much as the dipole is enhanced for a water molecule in the TIP3P[13] or SPC[14] models of water over its gas-phase value. Thus, ESP derived charges using this basis set have the virtue that they can give, in principle, a balanced representation of solvent–solute, solute–solute, and solvent–solute interactions. Of course, this is all within the effective two-body representation of many-body interactions, and there is no necessary reason for the solute to be as polarized as the solvent in aqueous solution. Time will tell whether this approach is the best for deriving effective two-body potentials, but the consistent use of ESP charges for any molecule or fragment, combined with TIP3P,[13] TIP4P,[13] or SPC[14] variants of water potentials offers a most promising approach to biomolecular simulations which is easily generalizable and aesthetically pleasing and consistent.

A major bottleneck in making such an approach has been the fact that ESP based charges have been derived to reproduce intermolecular properties and thus may be less well suited for intramolecular properties and molecular conformational analysis. A good example of this is the tendency of ESP derived charges on carbons in butane or other hydrocarbon fragments to be, in contrast to the Mulliken charges, disturbingly conformation dependent, leading to a number of artifacts in the conformational energetics.[15,16] This compromises the overall model, since these intramolecular interactions are critical for understanding both structure and energetics of complex biomolecular systems.

The reason that the ESP derived charges have these flaws is largely due to the statistical nature of the fitting process. In optimizing the charges to reproduce the ESP around the molecule, the charges on buried atoms can fluctuate wildly in order to yield a miniscule improvement in the statistical "quality-of-fit" to the ESP. Why does this happen? The ESP fitted charges are simply those coefficients which make an optimal least-squares fit of the given model function to the given set of ESP points. As in any fitting process, some coefficients (charges) will be statistically better determined by the data set than others. This is demonstrated in Figure 1, which shows the dependence of the overall quality of the fit in terms of the merit function χ^2 (see Methods below) on each of the charges in methanol. Clearly, χ^2 is relatively little affected when the methyl carbon charge is varied around its "best" value as compared to the other atomic centers. Charges such as this to which the quality of the fit (i.e. χ^2) is relatively insensitive are described here as being "poorly determined"; similarly charges to which χ^2 is sensitive are called "well determined".

The ESP points used in fitting the charges must lie outside the van der Waals surface of the molecule.[8] Hence "buried" charges (e.g. an sp^3 carbon) tend to be poorly determined because even the closest surface points at which the ESP is evaluated are relatively far away and in general closer to (and therefore more strongly dependent on) a different atomic charge. In general, the less solvent-exposed an atom is, the less well determined it is.

A dramatic illustration of the insensitivity of χ^2 to poorly determined centers was given by Singh and Kollman[8] in deriving a charge model with lone pairs on the oxygen of formaldehyde. With the addition of lone pair charge centers, the oxygen atomic center became poorly determined, so that when both the position and magnitude of the lone pair charges were allowed to optimize, the charges became very large and their position ended up nearly on top of the oxygen. Once constrained to be 0.6 Å from the oxygen, the charge on the lone pairs became more reasonable and the fit to the quantum mechanical ESP worsened by only 1%!

In principle, the conformational dependencies resulting from these artifacts can be corrected for by adding torsional potentials

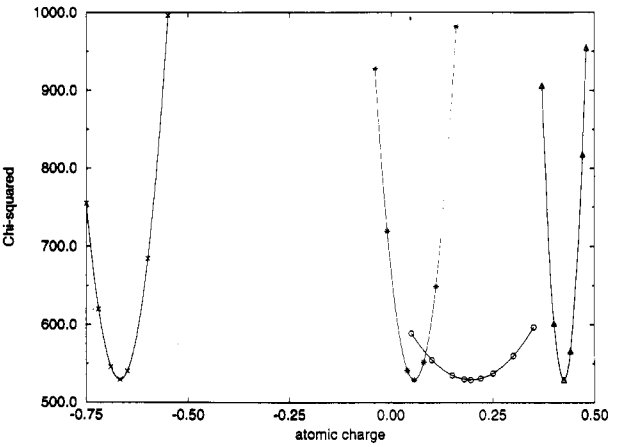

Figure 1. Dependence of the quality of fit (in terms of the merit function χ^2_{esp} (chi-squared)) on each atomic charge in methanol: (×) oxygen; (△) hydroxyl hydrogen; (○) methyl carbon; (*) methyl hydrogen.

to molecular mechanics force fields, but this is only really satisfactory if the charge effects are transferable between related chemical fragments; often they are not. Alternatively, Reynolds *et al.*[16] have tried to address the conformational dependency problem in a more explicit fashion. The ESP for each conformer of a molecule is calculated, and the charge for each center is fitted using all the ESP information at once, with Boltzmann weighting according to the relative energy of each conformer. This approach offers two advantages over fitting a single conformer. The first lies in the original purpose of the authors, which is the determination of a set of optimum consensus charges which best reproduce the ESPs of all conformers represented in the ESP data. This directly resolves much of the conformational dependence for those conformers. The second is that the statistically poorly determined centers for one conformer will often be at least marginally better determined in another. By providing much larger amounts of different ESP data, the statistical problem for the poorly determined centers is decreased, with an associated decrease in the charge fluctuations. The Boltzmann weighting aspect of this approach may be a problem in that the relative energies of different conformers can markedly change between the gas phase, which is used for the weighting, and polar solvated environments, which is a major context for the use of the fitted charges. However, the major drawback of this approach lies in the computational burden of (a) determining the appropriate set of conformers for each molecule and (b) calculating the quantum mechanical ESP, at a suitable level of theory, for each conformer in the set. This burden would increase exponentially with the number of rotatable bonds in the molecule.

In this paper and an associated one (ref 15, henceforth referred to as paper II), we address the problems of transferability and conformational dependence by dealing with the ill-behaved charges associated with the statistically poorly determined centers. We have investigated the effect of including a penalty function in the least-squares charge fitting procedure, in the form of restraints on non-hydrogen atomic charges to a target charge. The objective of the restraints is to hold down the ESP derived charges to a lower magnitude with only a minimal decrease in the quality of the fit. Mulliken charges (which exhibit a low conformational dependence) were obvious candidates for the target charge, but these were found to be unsatisfactory, working in some cases but not others. Using a target charge of zero (i.e. restraining the charges according to their magnitude) in conjunction with a nonharmonic restraint function was found to be a satisfactory solution to the problem of transferability and intramolecular electrostatics. With physically reasonable restraint weights that retain well the first two nonvanishing multipole moments of the molecule, we find that for those charges which are well determined by the ESP data set, their magnitude changes

little. On the other hand, those charges in the interior of the molecule, whatever their magnitude, change the most and take on much more intuitively reasonable values.

A different aspect of the charge-fitting protocol which was also re-examined in this work was the use of forced symmetry. In force-field treatments of conformationally labile species, it is of course necessary to have identical charges on nuclei equivalent with respect to the force field in terms of conformational interconversion; e.g. all the hydrogens on a methyl group must bear the same charge because otherwise the three degenerate rotamers of the methyl would give rise to different energies. This is analogous to the conformational averaging of the NMR signal of otherwise inequivalent nuclei. This requires additional symmetry to be forced upon the fitted charges beyond that given by the molecular symmetry. Indeed, the commonly used tactic is to average together the usually quite similar charges for all the centers which are being forced to be symmetric. This was found to have an unexpectedly heavy impact on the quality of the fit to the ESP and on the dipole moment. The basis for this behavior was examined, and an improved approach is presented for this aspect in addition to the charge restraints. Although the solution we propose may not be the final answer, we feel the work here is a major step, following the pioneering studies of Momany[6] and Cox and Williams,[7] in making ESP derived charges a general and useful way to generate atomic charges for simulations of complex systems.

Methods

The charge fitting process begins with having the QM ESP V_i evaluated for each point i of a set of points fixed in space in the solvent-accessible region around the molecule. The points must lie outside the van der Waals radius of the molecule for reasons described elsewhere;[8] here the points were generated using the method given in ref 5 with a shell of points (with a density of 1 point/Å^2) at each of 1.4, 1.6, 1.8, and 2.0 times the van der Waals radii. The QM ESP was evaluated at the 6-31G* level by using GAUSSIAN 90.[17]

The experimental geometry was used for methanol,[18] the AMBER[19] minimized geometry was used for *N*-methylacetamide (using the all-atom force field of Weiner *et al.*[12]), and 6-31G* optimized geometries were used for dimethyl phosphate, butane, propanol, and the tetrahedral intermediate for CH_3O^- attack on *N*-methylacetamide.

A least squares procedure was then used to fit the charge q_j to each atomic center j in the molecule. The calculated ESP $\hat{V}_i$ is given by

$$\hat{V}_i = \sum_j \frac{q_j}{r_{ij}} \tag{1}$$

so the figure-of-merit χ^2_{esp} to be minimized in the least-squares procedure is defined as

$$\chi^2_{esp} = \sum_i (V_i - \hat{V}_i)^2 \tag{2}$$

At the minimum

$$\partial(\chi^2_{esp})/\partial q_j = 0 \quad \text{for all } j \tag{3}$$

where

$$\frac{\partial(\chi^2_{esp})}{\partial q_j} = -2\sum_i \frac{V_i - \hat{V}_i}{r_{ij}} = 0 \tag{4}$$

and thus a system of equations can be formed and solved in matrix form, in conjunction with the constraint that the sum of all charges must equal the total molecular charge. A clear presentation of this stage is given in ref 5.

With the addition of a penalty function to the charge fitting procedure, an additional term is added to χ^2_{esp}, so that the figure-of-merit to be minimized is now

$$\chi^2 = \chi^2_{esp} + \chi^2_{rstr} \tag{5}$$

and the least squares minimum is now defined by

$$\frac{\partial(\chi^2)}{\partial q_j} = \frac{\partial(\chi^2_{esp})}{\partial q_j} + \frac{\partial(\chi^2_{rstr})}{\partial q_j} = 0 \quad \text{for all } j \tag{6}$$

The initial choice for the penalty function was a simple harmonic

$$\chi^2_{rstr} = a\sum_j (q_{0_j} - q_j)^2 \tag{7}$$

where a is the scale factor determining the strength of the restraint and q_0 is the target charge for the restraint. In the first model, a target charge of the Mulliken charge for each center was used, and the second model used a target charge of zero instead. With this type of penalty function, the second term in eq 6 is

$$\partial(\chi^2_{rstr})/\partial q_j = -2a(q_{0_j} - q_j) \tag{8}$$

The third model retained a target charge of zero but modified the penalty function to a hyperbolic form, giving

$$\chi^2_{rstr} = a\sum_j ((q_j^2 + b^2)^{1/2} - b) \tag{9}$$

where a is a scale factor which defines the asymptotic limits of the strength of the restraint and b determines the "tightness" of the hyperbola around its minimum. The second term in eq 6 is now

$$\partial(\chi^2_{rstr})/\partial q_j = aq_j(q_j^2 + b^2)^{-1/2} \tag{10}$$

Modifying the approach in ref 5 for solving the new system of equations defined by eq 6 is straightforward, where solving the matrix equation

$$\mathbf{Aq} = \mathbf{B} \tag{11}$$

for the vector **q** of charges retains the original form for **A** in the off-diagonal ESP-dependent elements

$$A_{jk} = \sum_i \frac{1}{r_{ij}r_{ik}} \tag{12}$$

but the diagonal ESP-dependent elements are now given by

$$A_{jj} = \sum_i \frac{1}{r_{ij}^2} + \frac{\partial(\chi^2_{rstr})}{\partial q_j} \tag{13}$$

and the ESP-dependent elements of **B** are given by

$$B_j = \sum_i \frac{V_i}{r_{ij}} + q_{0_j}\frac{\partial(\chi^2_{rstr})}{\partial q_j} \tag{14}$$

The other difference compared to the unrestrained fit is that using the hyperbolic penalty function (which has nonlinear derivatives in q_j) requires an iterative solution of the equations to self-consistency in q_j.

Charge centers could be made equivalent either for symmetry purposes or for fitting one set of charges to ESP data for multiple conformers of the same molecule. In both cases, preliminary matrices **A** and **B** were generated as if there were no equivalent

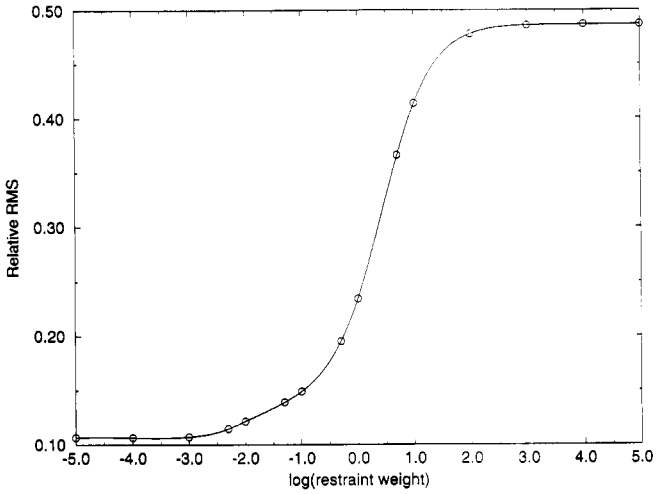

Figure 2. Dependence of the relative RMS (eq 15) on the weight placed on the model 1 restraint function (a in eqs 7 and 8).

charge centers. Then, the rows and columns of **A** (and the rows of **B**) for centers to be fitted to the same charge were simply combined together to form a single row and column of **A** (and a single row of **B**), giving rise to new, smaller versions of matrices **A** and **B**, which were solved as usual.

In accordance with previous work (refs 5–9) comparisons of the quality-of-fit to the QM ESP were based on the "relative root mean square (RMS) error" (RRMS), given by

$$\mathrm{RRMS} = \{\chi^2_{\mathrm{esp}}/\sum_i V_i^2\}^{1/2} \qquad (15)$$

Results and Discussion

Model 1 (harmonic restraints to the Mulliken charges) was used to assess the impact of the penalty function on the charges and quality of fit to the QM ESP of methanol and *trans*-butane. The purpose here was to ascertain what weighting of the penalty function would modify the charges in a desirable way without seriously reducing the quality of fit to the QM ESP. Also, the dipole and quadrupole moments were additional criteria taken into consideration in evaluating various charge sets for use in the force field.

For methanol, the results of varying the restraint weight a through a wide range are shown in Figure 2. A weight of less than 0.001 au had a negligible impact on χ^2_{esp}, while using weights higher than 1.0 au caused χ^2_{esp} to increase dramatically and then level off toward its limiting value, which is the χ^2_{esp} derived from the target charges q_0 (in this case the Mulliken charges). Since a good fit to the QM ESP (i.e. low χ^2_{esp}) is important, weights exceeding 0.05 were no longer considered.

Table I gives the charges, RRMS, and moments for methanol with weights spanning the range of 0.001–0.05 au along with the limits of no restraint and complete constraint to the Mulliken charges. The only statistically poorly determined center in this molecule is the methyl carbon; here is where the effects of charge restraints would be the most pronounced. There is a qualitative difference between the unrestrained ESP charge (0.19) and the Mulliken charge (–0.17) for the carbon. As the restraint weight increased, the carbon charge gradually changed from the unrestrained ESP value toward the Mulliken value. As the restraint weight increased, the carbon charge gradually changed from the unrestrained ESP value toward the Mulliken value. At the highest restraint weight in the table (0.05 au), it was already quite close to the Mulliken charge, but the RRMS increased by less than 31% compared to the 357% increase corresponding to the Mulliken charge set. The dipole moment was still very close to the unrestrained ESP fit value, but the quadrupole moments deteriorated somewhat, falling between the unrestrained and

TABLE I: Fitted Charges and Dipole and Quadrupole Moments for Methanol for Different Weights of the Charge Restraint

	value for given charge restraint weight[a]					
	none	0.001	0.005	0.01	0.05	Mulliken[b]
			Charges			
O	–0.6680	–0.6559	–0.6314	–0.6200	–0.6169	–0.7261
H	0.4233	0.4235	0.4249	0.4269	0.4404	0.4383
C	0.1955	0.1402	0.0221	–0.0411	–0.1403	–0.1747
H(t)	0.0568	0.0701	0.0988	0.1148	0.1449	0.1732
H(g)	–0.0039	0.0110	0.0428	0.0597	0.0860	0.1447
			RRMS			
	0.1065	0.1074	0.1147	0.1216	0.1391	0.4864
			Dipole and Quadrupole Moments			
μ^c	1.8972	1.8967	1.8975	1.9008	1.9336	2.7741
$Q_{xx}{}^d$	1.3147	1.4189	1.6445	1.7697	2.0024	2.4351
$Q_{yy}{}^d$	1.8025	1.7304	1.5834	1.5139	1.4742	0.9874
$Q_{zz}{}^d$	–3.1171	–3.1493	–3.2279	–3.2836	–3.4766	–3.4225

[a] Restraint weight in atomic units (a in eqs 7 and 8). [b] Mulliken charges evaluated against the QM ESP. [c] Dipole moment in debye. [d] Quadrupole moment along the principal axes in debye angstroms.

TABLE II: Fitted Charges and Dipole and Quadrupole Moments for *trans*-Butane for Different Weights of the Charge Restraint

	value for given charge restraint weight[a]					
	none	0.001	0.005	0.01	0.05	Mulliken[b]
			Charges			
C1	–0.3419	–0.3464	–0.3812	–0.4091	–0.4620	–0.4786
H1(t)	0.0854	0.0884	0.1016	0.1114	0.1300	0.1609
H1(g)	0.0731	0.0791	0.0984	0.1116	0.1355	0.1561
C2	0.1584	0.0977	–0.0332	–0.1079	–0.2335	–0.3001
H2	–0.0241	0.0011	0.0580	0.0912	0.1473	0.1528
			RRMS			
	0.7473	0.7674	0.9466	1.1121	1.4483	2.2144
			Dipole and Quadrupole Moments			
μ^c	0.0001	0.0001	0.0001	0.0001	0.0002	0.0000
$Q_{xx}{}^d$	–1.2265	–1.3221	–1.5407	–1.6608	–1.7438	3.8969
$Q_{yy}{}^d$	0.7813	1.0047	1.5302	1.8392	2.3019	–0.6263
$Q_{zz}{}^d$	0.4452	0.3173	0.0105	–0.1784	–0.5581	–3.2705

[a] Restraint weight in atomic units (a in eqs 7 and 8). [b] Mulliken charges evaluated against the QM ESP. [c] Dipole moment in debye. [d] Quadrupole moment along the principal axes in debye angstroms.

Mulliken values. The restraint weights of 0.01 and 0.005 au appeared to be optimal for methanol, yielding charges of low magnitude on the carbon, very slight increases of 14% and 8%, respectively, in the RRMS over the unrestrained case, excellent agreement of the dipole moment, and good agreement in the quadrupole moments.

The results for *trans*-butane, given in Table II, were similar to those of methanol, indicating that the charge restraints behaved similarly in both the polar and nonpolar case. The optimal restraint weights fell in the range 0.001–0.01 au, which influenced the carbon charges while having a small impact on the quadrupole moment (the dipole moment is zero by symmetry; the very low moments appearing in Table II are due to the asymmetry of the ESP points around the molecule). The restrained charge sets given still show a relatively small increase in RRMS compared to the Mulliken charges. The RRMS numbers are unusually high compared to the other molecules tested (cf. paper II); this arises from the low polarity resulting in a very low denominator in evaluating the RRMS (eq 15).

Clearly evident from Tables I and II is that the Mulliken charges perform much worse in reproducing the QM ESP than the ESP fit charges. This was found to be true in general for the series of molecules examined (cf. paper II), with the Mulliken charges tending to have RRMS values around two to four times as bad as the ESP-fit charges (a notable exception was *N*-methylacetamide (NMA) for which the Mulliken charges gave only a slight

TABLE III: Comparison of Charge Fitting for Methanol Using Models 2 and 3 to the Unrestrained ESP Fit and to the Mulliken Charges

			model 2 at given weight[b]		model 3 at given weight[b]	
	ESP fit	Mulliken[a]	0.005	0.01	0.0005	0.001
			Charges			
O	−0.6680	−0.7261	−0.6224	−0.5971	−0.6498	−0.6369
H	0.4233	0.4383	0.4125	0.4026	0.4215	0.4198
C	0.1955	−0.1747	0.0569	0.0025	0.1252	0.0781
H(t)	0.0568	0.1732	0.0841	0.0918	0.0726	0.0828
H(g)	−0.0038	0.1447	0.0345	0.0501	0.0153	0.0282
			RRMS			
	0.1065	0.4864	0.1134	0.1223	0.1079	0.1105
			Dipole and Quadrupole Moments			
μ^c	1.8972	2.7741	1.8748	1.8562	1.8923	1.8881
$Q_{xx}{}^d$	1.3147	2.4351	1.5392	1.6096	1.4404	1.5228
$Q_{yy}{}^d$	1.8025	0.9874	1.5347	1.3863	1.6953	1.6192
$Q_{zz}{}^d$	−3.1171	−3.4225	−3.0739	−2.9959	−3.1357	−3.1420

[a] Mulliken charges evaluated against the QM ESP. [b] Restraint weight in atomic units (*a* in eqs 7 and 8 for model 2, eqs 9 and 10 for model 3). [c] Dipole moment in debye. [d] Quadrupole moment along the principal axes in debye angstroms.

TABLE IV: Comparison of Charge Fitting for *trans*-Butane Using Models 2 and 3 to the Unrestrained ESP Fit and to the Mulliken Charges

			model 2 at given weight[b]		model 3 at given weight[b]	
	ESP fit	Mulliken[a]	0.005	0.01	0.0005	0.001
			Charges			
C1	−0.3419	−0.4786	−0.1208	−0.0742	−0.1716	−0.0918
H1(t)	0.0854	0.1609	0.0283	0.0161	0.0417	0.0208
H1(g)	0.0731	0.1561	0.0243	0.0138	0.0363	0.0181
C2	0.1584	−0.3001	0.0456	0.0249	0.0614	0.0287
H2	−0.0241	0.1528	−0.0008	0.0028	−0.0021	0.0031
			RRMS			
	0.7473	2.2144	0.8384	0.8768	0.8051	0.8627
			Dipole and Quadrupole Moments			
μ^c	0.0001	0.0000	0.0001	0.0001	0.0001	0.0001
$Q_{xx}{}^d$	−1.2265	3.8969	−1.2096	−1.2017	−1.2274	−1.2103
$Q_{yy}{}^d$	0.7813	−0.6263	0.7523	0.7353	0.7937	0.7557
$Q_{zz}{}^d$	0.4452	−3.2705	0.4573	0.4664	0.4337	0.4546

[a] Mulliken charges evaluated against the QM ESP. [b] Restraint weight in atomic units (*a* in eqs 7 and 8 for model 2, eqs 9 and 10 for model 3). [c] Dipole moment in debye. [d] Quadrupole moment along the principal axes in debye angstroms.

worse fit to the QM ESP than the ESP-fit charges themselves). Further limiting the usefulness of the Mulliken charges is their high magnitude, especially on alkyl carbons, often exceeding that of the corresponding ESP-fit charges (cf. C1 in Table II). Having charges of high magnitude as the target charges for the penalty function was inconsistent with the objective of reducing the magnitude of the charges where possible. Also, the negative Mulliken charge on the carbon in methanol (cf. Table I) is an example of a frequently occurring situation where the Mulliken charge is in even qualitative disagreement with both chemical intuition and the ESP fit charges from the same wavefunction.

Using target charges of zero (model 2) restrains all charges to a lower value; the extent to which this happens again depends upon the restraint weight. Tables III–VI give results for using model 2 on methanol, *trans*-butane, *trans*-NMA, and *trans,trans*-dimethyl phosphate (tDMP), respectively. As with the Mulliken target charges, a restraint weight range of 0.005–0.01 au was optimal for decreasing the charges without seriously degrading the RRMS or the electric moments. With this model the decision was made to leave hydrogens unrestrained since they are virtually always well solvent-exposed (i.e. well determined statistically)

TABLE V: Comparison of Charge Fitting for *trans*-*N*-Methylacetamide Using Models 2 and 3 to the Unrestrained ESP Fit and to the Mulliken Charges

			model 2 at given weight[b]		model 3 at given weight[b]	
	ESP fit	Mulliken[a]	0.005	0.01	0.0005	0.001
			Charges			
C1	−0.4902	−0.4941	−0.0402	0.0155	−0.2361	−0.0779
H1(t)	0.1552	0.1490	0.0564	0.0463	0.0965	0.0614
H1(g)	0.1206	0.1240	0.0204	0.0128	0.0608	0.0248
C	0.7611	0.7766	0.3604	0.2488	0.5869	0.4579
O	−0.6261	−0.6074	−0.5333	−0.4981	−0.5912	−0.5633
N	−0.5372	−0.5805	−0.2302	−0.1430	−0.4192	−0.3234
H	0.3208	0.3245	0.2246	0.1969	0.2823	0.2514
C2	−0.0489	−0.0066	−0.0672	−0.0493	−0.0418	−0.0395
H2(t)	0.1117	0.0962	0.1042	0.0922	0.1073	0.1035
H2(g)	0.0564	0.0474	0.0423	0.0326	0.0470	0.0403
			RRMS			
	0.0680	0.0817	0.0954	0.1107	0.0742	0.0850
			Dipole and Quadrupole Moments			
μ^c	4.1396	3.9764	4.1089	4.0971	4.1279	4.1189
$Q_{xx}{}^d$	10.2871	9.8162	10.1473	9.9764	10.3146	10.3014
$Q_{yy}{}^d$	−7.9054	−7.4989	−7.4254	−7.1045	−7.7881	−7.6770
$Q_{zz}{}^d$	−2.3817	−2.3173	−2.7219	−2.8719	−2.5265	−2.6244

[a] Mulliken charges evaluated against the QM ESP. [b] Restraint weight in atomic units (*a* in eqs 7 and 8 for model 2, eqs 9 and 10 for model 3). [c] Dipole moment in debye. [d] Quadrupole moment along the principal axes in debye angstroms.

TABLE VI: Comparison of Charge Fitting for *trans,trans*-Dimethyl Phosphate Using Models 2 and 3 to the Unrestrained ESP Fit and to the Mulliken Charges

			model 2 at given weight[b]		model 3 at given weight[b]	
	ESP fit	Mulliken[a]	0.005	0.01	0.0005	0.001
			Charges			
C	0.1617	−0.1466	0.0791	0.0460	0.0816	0.0479
H(t)	0.0009	0.1203	−0.0050	−0.0065	0.0164	0.0199
H(g)	0.0183	0.1480	0.0564	0.0708	0.0422	0.0541
–O–	−0.5092	−0.7004	−0.3448	−0.2810	−0.4664	−0.4322
P	1.2223	1.5057	0.5650	0.3096	1.1102	0.9910
O=	−0.8011	−0.8223	−0.6246	−0.5550	−0.7711	−0.7392
			RRMS			
	0.0128	0.0834	0.0199	0.0248	0.0132	0.0143
			Dipole and Quadrupole Moments			
μ^c	2.3116	0.3121	2.2688	2.2503	2.3088	2.3031
$Q_{xx}{}^d$	33.4563	48.9791	31.7463	30.9764	33.1604	32.8441
$Q_{yy}{}^d$	−13.6141	−23.8970	−14.1467	−14.2867	−13.6831	−13.7722
$Q_{zz}{}^d$	−19.8422	−25.0821	−17.5996	−16.6897	−19.4773	−19.0719

[a] Mulliken charges evaluated against the QM ESP. [b] Restraint weight in atomic units (*a* in eqs 7 and 8 for model 2, eqs 9 and 10 for model 3). [c] Dipole moment in debye, relative to center of mass. [d] Quadrupole moment along the principal axes in debye angstroms.

and there are many of them in a molecule; restraining them was found to do little more than reduce the quality-of-fit of the overall system.

For methanol roughly the same results were obtained between model 1 and model 2 for the weak restraint (weight = 0.005 au); for the stronger restraint (weight = 0.01 au), model 2 gave lower magnitude charges in general, a slightly worse dipole moment, but did markedly better with the quadrupole moment. For *trans*-butane model 2 performed much better than model 1, giving greatly decreased charges on the methyl carbons, improved RRMS values, and much better conservation of the quadrupole compared to the unrestrained ESP, although the dipole moment was slightly nonzero.

A problem with model 2 became clear when it was applied to the very polar tNMA and the negatively charged tDMP. Because the restraining force from the harmonic penalty function is linear with charge (cf. eq 8), the polar centers experience a much stronger

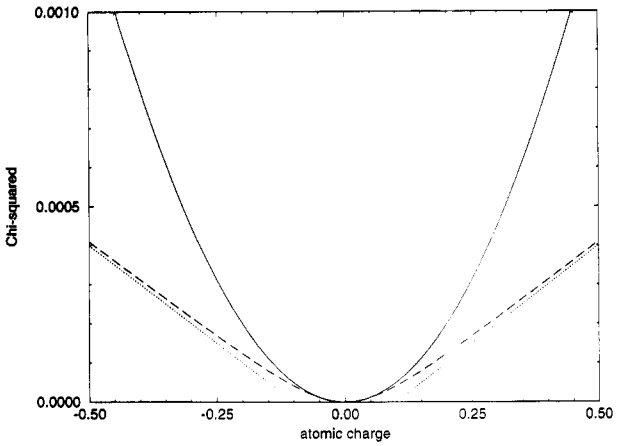

Figure 3. Comparison of the harmonic penalty function of model 2 (solid line) with the hyperbolic penalty function of model 3 (dashed line). Note the asymptotic behavior of model 3 at higher magnitudes of the charge; the asymptotes are shown as dotted lines.

restraining force. The polar centers in the examples given here, namely, the OH in methanol, the amide in tNMA, and the phosphorous and the two free oxygens in tDMP, were therefore restrained with a much larger force than the other centers, resulting in fitted charges that systematically underestimated their inherent polarity. Thus, while the overall fit was still quite good in terms of RRMS, most of the error was found to lie in the charged regions of the ESP (very important for solvation) in the form of significant underestimates of the magnitude of the ESP. Model 2 therefore showed that while a target charge of zero could successfully be used to decrease the charges in a general way, the quadratic form of the penalty function overrestrained centers associated with polar regions. One way of dealing with the overrestraining problem would be to designate before the fitting process some centers as being polar, and therefore left free of restraints. This tactic was rejected for two reasons: first, the designating process would be subjective and questionable in the case of centers of intermediate polarity, and second, a buried polar center (such as the phosphorus in tDMP) would still be statistically poor determined and hence a (reasonable) restraint would still be warranted.

Model 3 differs from model 2 in that it seeks to correct the overrestraining of polar centers by modifying the form of the penalty function. A hyperbolic function of the form given in eq 9 possesses the same desirable characteristics for a restraint as a quadratic function, i.e. a single minimum about which the function is symmetric, and it is well-defined at all points and easily differentiable. Figure 3 shows that unlike the quadratic restraint the force from the hyperbolic restraint does not continually increase with increasing charge but rather asymptotically reaches a limiting value determined by the restraint weighting factor a. The second parameter b required by the hyperbolic function defines the "tightness" of the hyperbola around the minimum; i.e. at how low a magnitude of the charge does the restraint adopt asymptotic behavior. A value of 0.1 electrons for b made the restraint appropriately tight for our purposes; a value much higher did not introduce enough restraining force for charges of intermediate magnitude, and smaller values would unnecessarily restrain charges of low magnitude and could begin to delay convergence of the fitting procedure.

Tables III–VI compare models 2 and 3 for the same four molecules. Due to the change in form of the penalty function for model 3, the weight does not act in the same way as with the previous two models; the optimal weak and strong weights were an order of magnitude smaller. For methanol and *trans*-butane model 3 gave yet another improvement in the RRMS and electric moments over model 2, with charge decreases on the methyl carbons comparable to model 2. The model 3 charges on the polar OH centers in methanol retained higher magnitudes than for model 2, especially when the strong restraint weight charge sets are compared. The most outstanding improvement in using the hyperbolic restraints is seen for tNMA and tDMP (Tables V and VI). Model 2 lowers the tNMA amide charge magnitudes enough to substantially decrease its aqueous solvation energy compared to the unrestrained charges (11.3 kcal/mol (weak restraint) versus 12.4 kcal/mol (unrestrained) for the aqueous tNMA-to-methane free energy perturbation; cf. paper II for a complete description of the methodology). This overrestraint is reflected in the 9 and 63% increase in the RRMS for the weak and strong restraint, respectively. Model 3 lowers the amide charge magnitudes much less, increasing the RRMS by only 9% (weak) and 25% (strong), and decreasing the solvation energy by only a small amount (12.1 kcal/mol (weak restraint) for aqueous tNMA-to-methane; cf. paper II). The phosphorus charge in tDMP is dramatically decreased by model 2, adopting counter-intuitively low values. This is easily rationalized on the basis of it being both statistically poorly determined and having an exaggerated restraint presented by the harmonic penalty function. The oxygen charges are also low for an anionic species, and the RRMS and electric moments show some deterioration. With the better behaved restraining force offered by model 3, the phosphorus and oxygen charges remain appropriately high, and the RRMS and electric moments remain close to those of the unrestrained fit.

Model 3 thus showed itself to be the best approach for lowering the magnitudes of the fitted charges, especially on poorly determined centers, while retaining virtually all of the characteristics of the unrestrained ESP fitted charges in terms of quality-of-fit (χ^2_{esp}) and the electric moments. On the basis of the molecules examined in this work and in paper II, using restraint weights of either 0.0005 au for weak restraint and 0.001 au for stronger restraint gave the best results. Charge sets arising from the unrestrained, weakly restrained, or strongly restrained fitting will be referred to as un, wk, or st, respectively.

The effects of forcing symmetry beyond the molecular symmetry (in order to accommodate conformational interconversion) was examined for all the test molecules mentioned in this work and in paper II, but for purposes of brevity methanol will be used as a simple example presented in more detail. The forced symmetry necessary for a force-field treatment of methanol is only to require all the methyl hydrogens to bear the same charge. Methanol has one methyl hydrogen *trans* to the OH bond, lying in the C_s plane, and the other two methyl hydrogens *gauche*, symmetrically on either side of the C_s plane. The *trans* hydrogen is electronically different from the *gauche* hydrogens and the QM ESP derived charges are therefore different for the *trans* and *gauche* hydrogens (cf. Table III). Forcing symmetry in the methyl hydrogens can be carried out either by averaging the three methyl hydrogen charges *a posteriori* (i.e. after the fitting) or by equivalencing them in the fitting process, requiring all three to be fitted to the same charge.

A notation is now defined which refers to the forced symmetry on methyl and methylene hydrogens: fr ("free") denotes no forced symmetry, ap ("a posteriori") denotes charge averaging after the fitting process, and eq ("equivalenced") denotes equivalencing of the symmetric charges during the fit. To describe the fitting process as a whole, the restraint model (un, wk, or st) is followed by a period as a delimiter and then the forced-symmetry model, e.g. wk.ap denotes a weakly restrained fit (no forced symmetry in the fit) followed by *a posteriori* averaging to force the symmetry, whereas un.eq denotes an unrestrained fit with forced symmetry by equivalencing centers in the fit.

Table VII compares the ap and eq with the fr results for methanol. Forcing the extra symmetry on the methyl hydrogens had a surprisingly large impact on all aspects of the fit, more than any of the restraint models alone. The RRMS increased by more than 61% for un.ap and more than 39% for un.eq compared to

TABLE VII: Comparison of *a Posteriori* Averaged, Equivalenced, and Free Charges for Methanol Using Both Unrestrained and Weakly Restrained (Model 3) Fitting Schemes (Results of Two-Stage Approach Also Given)

forced symmetry[a]	unrestrained (un)			weakly restrained (wk)			two-stage
	fr	ap	eq	fr	ap	eq	wk.fr/st.eq
			Charges				
O	−0.6680	−0.6680	−0.6027	−0.6498	−0.6498	−0.5921	−0.6498
H	0.4233	0.4233	0.3862	0.4215	0.4215	0.3861	0.4215
C	0.1955	0.1955	0.0996	0.1252	0.1252	0.0546	0.1166
H(t)	0.0568	0.0163	0.0389	0.0725	0.0343	0.0504	0.0372
H(g)	−0.0039	0.0163	0.0389	0.0151	0.0343	0.0504	0.0372
			RRMS				
	0.1064	0.1719	0.1481	0.1079	0.1670	0.1485	0.1672
			Dipole and Quadrupole Moments				
μ^b	1.8972	2.1510	2.0013	1.8923	2.1316	1.9924	2.1392
Q_{xx}^c	1.3146	1.3604	1.4187	1.4403	1.4836	1.5005	1.5052
Q_{yy}^c	1.8024	1.3218	1.1032	1.6952	1.2413	1.0551	1.2289
Q_{zz}^c	−3.1171	−2.6823	−2.5220	−3.1356	−2.7249	−2.5556	−2.7340

[a] fr, free; ap, *a posteriori* averaged; eq, equivalenced; wk, weak restraints; st, strong restraints. (See text for explanation). [b] Dipole moment in debye. [c] Quadrupole moment along the principal axes in debye angstroms.

TABLE VIII: Comparison of Hydrogen Bonding Energies for the Methanol–Water Dimer Using Several Charge Fitting Schemes

charge fitting scheme[b]	hydrogen bonding energy[a] (kcal/mol)	
	$H_2O\cdots HOCH_3$	$HOH\cdots OCH_3$
un.fr	−6.8	−6.0
un.ap	−7.0	−6.3
un.eq	−6.2	−5.9

[a] Methodology given in detail in the associated paper (paper II). [b] un, no restraints; fr, free; ap, *a posteriori* averaged; eq, equivalence. (See text for explanation).

un.fr; these are much greater increases than with even the strong restraints of models 1–3. The dipole moment also increased substantially, by >13 and >5% for un.ap and un.eq, respectively, where little change had resulted from restraining the charges as presented in Tables III–VI. The weakly restrained charge sets in Table VII give very similar results to the unrestrained charge sets; wk.ap is marginally better than un.ap and wk.eq is slightly worse than un.eq in terms of RRMS and the electric moments.

There were also pronounced effects on the intermolecular interactions from forced symmetry in the methanol charges, as shown for the unrestrained charge sets in Table VIII. Accompanying the enhanced dipole moment in the ap cases came an increase in the methanol–water hydrogen bond energies compared to un.fr. Interestingly, in the un.eq case the hydrogen bond energies decreased compared to the corresponding fr model, even though the dipole moments were still enhanced; this also held true for the wk.eq case (−6.1 and −5.8 kcal/mol with water as proton acceptor and donor, respectively).

In requiring all three methyl hydrogens to bear the same charge, the degrees of freedom of the methanol system (i.e. the number of non-equivalent charge centers, minus one for the "fixed total charge" constraint) is lowered from 4 to 3. With the ap treatment, the system is not allowed to relax with respect to the adjustment in charge distribution, and the quality of fit markedly decreases as a result. The enhancement of the dipole moment reflects an overall enhancement of the polar regions of the ESP. This is apparent upon comparison of the un.ap and un.fr ESP in Figure 4. Each ESP point is represented by an ellipse proportional in size to the magnitude of the residual (i.e. the error) between the QM ESP and the calculated (based on the atomic charges) ESP as given in eq 2. The color of the ellipse is associated with the value of the ESP at that point, with blue representing a neutral or nonpolar region (i.e., an ESP of zero), gradually changing through cyan to green for negative ESP or gradually changing through magenta to red for positive ESP. The minor axis of the ellipse is directly proportional to the residual, so a short minor axis (giving a prolate ellipse) indicates an over-/underestimate of a positive/negative ESP and a long minor axis (oblate ellipse) indicates an under-/overestimate. Comparing the un.ap ESP points shown in Figure 4b with the un.fr ESP points in Figure 4a, the increase in the error with un. ap is quite apparent. The large ellipses (i.e. large errors) in the polar regions are prolate in the positive (magenta) region around the hydroxyl hydrogen, indicating positive overestimates, and oblate in the negative (cyan to green) region around the hydroxyl oxygen, i.e. negative overestimates. These large errors occur in the region bordering the hydroxyl and the newly-adjusted methyl; the methyl itself is well fitted. The most polar areas, which cap the hydroxyl oxygen and hydrogen are also well fitted (small ellipses); this is the key region for hydrogen bonding and solvation interactions; over-/underestimates here have a much stronger impact than in the less polar areas.

In the eq approach, the atomic charges are allowed to redistribute themselves around the system which improves the fit over the ap result, but with one fewer degree of freedom the fit cannot be as good as with the fr case. Also, the true non-equivalence of the *trans* versus *gauche* hydrogens with respect to the QM ESP is removed, which will by necessity introduce error into the best possible fit. How will this error be distributed over the molecule? The least squares fitting procedure will tend to spread the error approximately evenly over the whole set of ESP data, which amounts to fitting the lowest electric moment (i.e. the dipole) the best. The higher moments, corresponding to more localized variations in the ESP, will be worse approximated. Thus the un.eq results show an improvement in the dipole moment over un.ap and a relatively small deterioration over un.fr (cf. Table VII), but the charges themselves decrease considerably. The un.eq ESP points in Figure 4c reflect the improvement in the fit over un.ap (Figure 4b), but notice that the most polar areas clearly contain more error than in the un.ap case, in keeping with the more even distribution of the error over all the ESP points. This explains the sizable decrease in hydrogen bonding energies resulting from the eq charge set, as shown in Table VIII. As a consequence of the lowered hydrogen bonding energies, the solvation free energy of aqueous methanol using the eq charge set was also significantly less than that of the ap charge set (cf. paper II).

This behavior for methanol was found for *trans*-NMA (cf. hydrogen bonding and solvation results in paper II), and this prompted a re-evaluation of one of the underlying assumptions behind atomic charge fitting, i.e. that the best set of charges to use for modeling Coulombic interactions would result from the "best fit" to the QM ESP in terms of the χ^2_{esp} and electric moments. The above results suggest that since hydrogen bonding and solvation energies are much more sensitive to the polar parts of the ESP than the less polar parts, the polar regions need to be

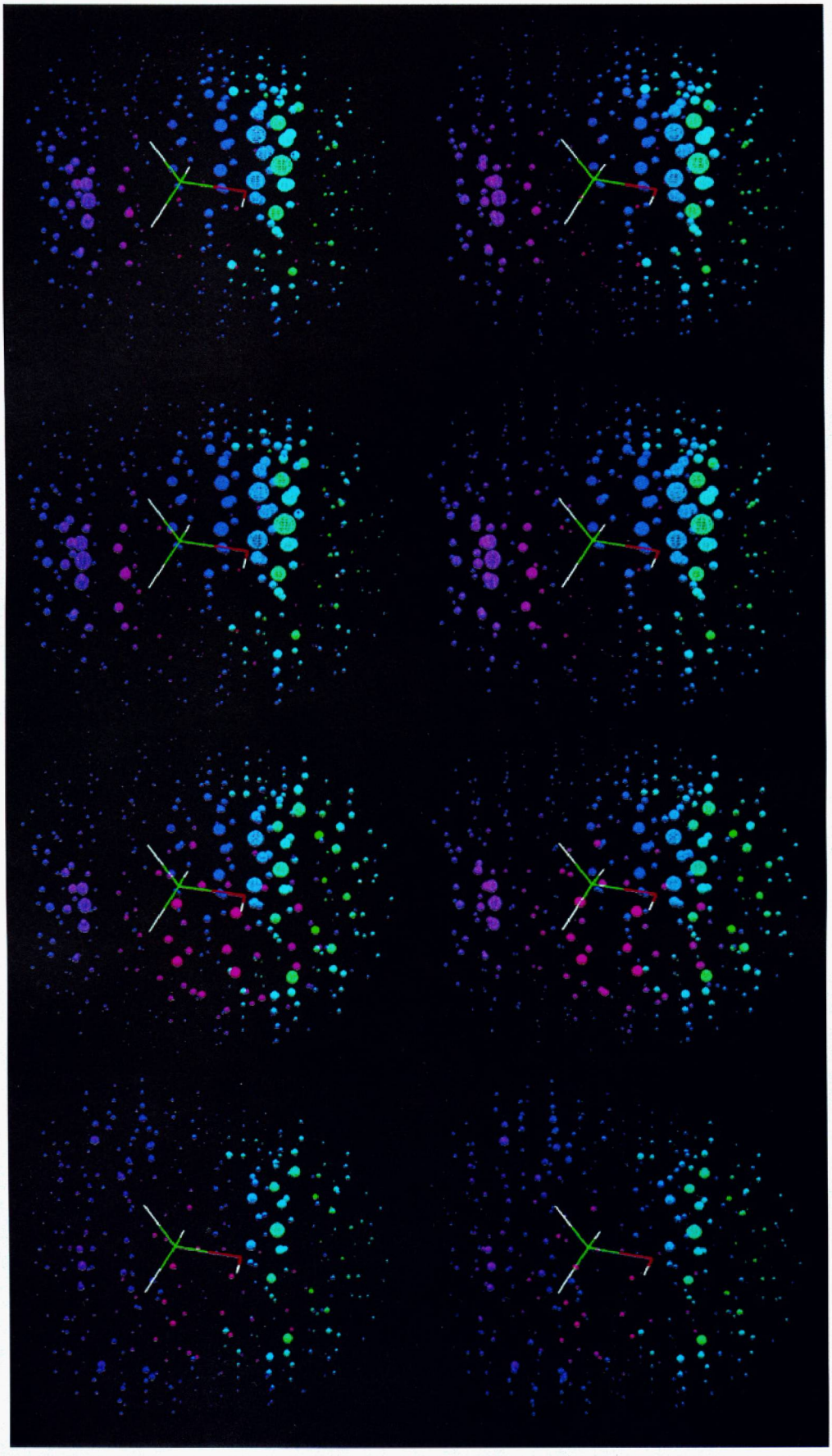

Figure 4. Visual comparison of the QM and calculated ESP for several charge sets fitted to methanol (refer to the text for a description of the comparison). The charge sets shown are (a, top) un.fr, (b, second from top) un.ap, (c, second from bottom) un.eq, and (d, bottom) wk.fr/st.eq.

the better fitted. Given the limited degrees of freedom offered by using only atomic monopoles and further limitations presented by the necessity of forcing symmetry, a set of atomic charges will only be able to model the QM ESP to a limited extent even in the best of cases. Recognizing that the error represented by χ^2_{esp} cannot be removed (although it should still be minimized), the

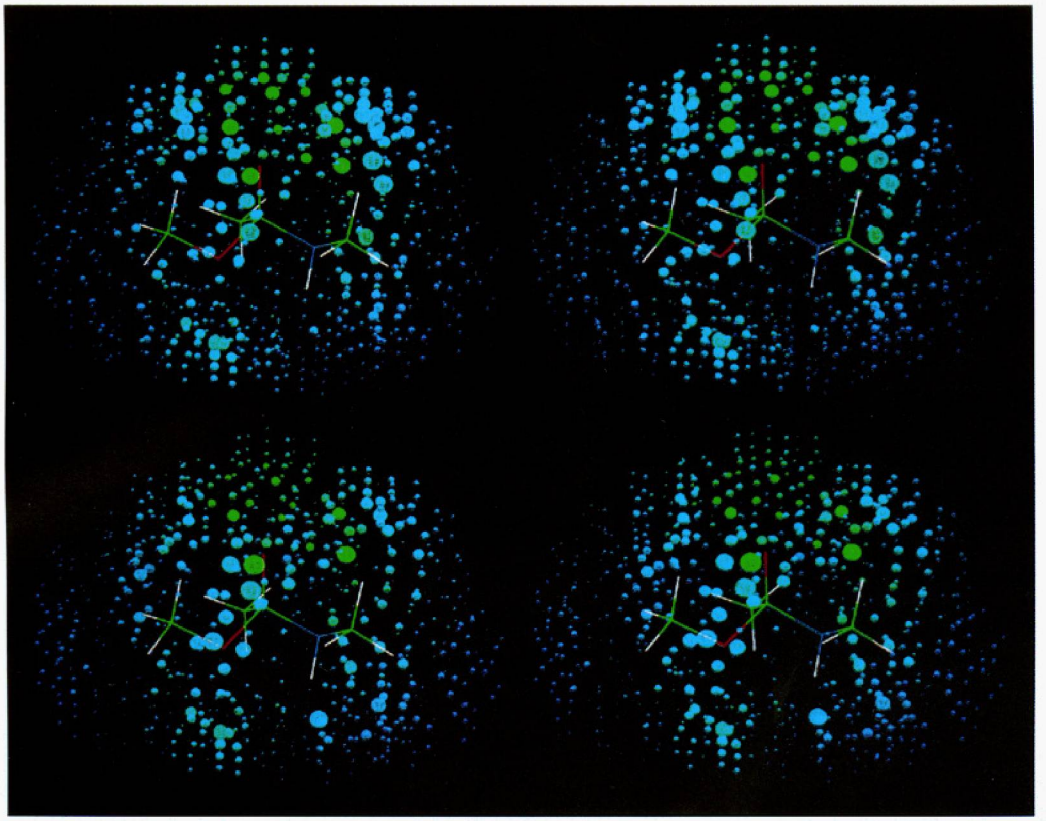

Figure 5. Visual comparison of the QM and calculated ESP for (a, top) the un.ap charges, and (b, bottom) the two-stage wk.fr/st.eq charges for the tetrahedral intermediate resulting form the attack of methoxide on *N*-methylacetamide (refer to the text for a description of the comparison).

problem is then one of keeping that error away from the sensitive polar areas, moving it to the less polar areas as much as possible. It would thus be better to achieve a good fit of the polar areas even at some expense to the χ^2_{esp} and the electric moments than to have a best fit containing significant error in the polar areas.

There are several options for selectively fitting the polar regions well. An obvious first candidate is to accord a greater weight during the fit to ESP values of higher magnitude. Several initial attempts were made using weighting schemes based on either the square of the QM ESP or its absolute value. While all such schemes succeeded in improving the fit of the ESPs of highest magnitude, the majority of the ESPs in the polar regions are of intermediate magnitude, where no significant improvement was obtained compared to the eq fit. This approach was therefore abandoned.

The ap averaging discussed above was another possibility for selective fitting. Because the ap averaging is almost always carried out over the hydrogens of nonpolar CH_2 and CH_3 groups, the charges characterizing the polar regions (which were originally fitted at the fr level, having the maximum degrees of freedom available for the fit) are kept fixed at the original (fr values). In many cases (e.g. methanol and tNMA) this suffices to maintain well fitted polar regions (cf. Figure 4b), because the polar region is far enough away from the region involving the averaged charges that it is only impacted where it borders the "averaged" region. There remain, however, the disadvantages of enhanced ESPs (in the border region) and the marked degradation in the quality of the overall fit. A more serious problem with this approach is that in a molecular environment where a polar region is not well solvent-exposed but rather is flanked by alkyl groups subject to ap averaging (e.g. a sterically hindered polar group), the averaging can very adversely affect the fit to the polar region. An example of this is the tetrahedral intermediate resulting from the attack of methoxide on tNMA, giving $CH_3C(OHC_3(O^-)NHCH_3$. In Figure 5a are shown the un.ap ESP points; as with the methanol case the regions of large error are associated principally with the regions bordering the averaged methyls. The most polar area of this molecule is obviously the anionic oxygen, which is flanked by all three methyls such that it has relatively little solvent-exposed area of its own. The combined effect of the averaging serves to make this important area also the area with the highest error, again in the form of overestimates of the ESP. The ap approach is thus shown to be somewhat "hit or miss" in terms of getting well fitted polar regions; this, added to the other disadvantages mentioned above, stimulated the search for an alternative approach.

Introducing additional charges in addition to the atomic centers, e.g. "lone pairs" charges, appears attractive in adding additional degrees of freedom to the system (which can only improve the fit). However, other than that advantage, the problems outlined above for the eq and ap approaches remain. Also, systems containing lone pair charges will tend to have higher charges on the lone pair atoms,[8] exacerbating the potential problems presented to the rest of the force field (cf. the Introduction). Compounding this effect, polar atoms such as alcohol and ether oxygens that are well solvent-exposed, and hence well determined statistically for charge fitting purposes, would become statistically much more poorly determined "buried centers" similar to methyl carbons if lone pair charges were added to them. Finally, the simplicity of the atom-centered point charge model would be sacrificed in considering how to treat lone pair charges from both a theoretical and practical standpoint[8] (e.g. where to locate them in space and with what flexibility). Again, an alternative approach seemed desirable.

To this end, a two-stage fitting procedure was developed based upon the fr and eq methods. The first stage is a wk.fr fit; this has the maximum degrees of freedom in order to get well fitted polar areas and uses weak hyperbolic restraints to decrease the overall magnitude of the charges. The second stage is where symmetry forcing is carried out as necessary; the goal of this stage is to hold fixed the charges characterizing the polar areas

TABLE IX: Overall Comparison of Several Charge Fitting Schemes for the Tetrahedral Intermediate Resulting from Methoxide Attack on *trans*-*N*-Methylacetamide

fitting scheme[a]	un.fr	un.ap	wk.fr	wk.ap	wk.fr/ st.eq
			Charges		
8	−0.6403	−0.6403	−0.5455	−0.5455	−0.5455
6	1.2696	1.2696	1.0242	1.0242	1.0242
8	−0.9589	−0.9589	−0.9065	−0.9065	−0.9065
7	−0.8839	−0.8839	−0.7904	−0.7904	−0.7904
1	0.3292	0.3292	0.3131	0.3131	0.3131
6	−0.8634	−0.8634	−0.6101	−0.6101	−0.1562
1	0.2046	0.1647	0.1483	0.1082	−0.0219
1	0.1527	0.1647	0.1004	0.1082	−0.0219
1	0.1369	0.1647	0.0759	0.1082	−0.0219
6	0.1121	0.1121	0.0166	0.0166	−0.2248
1	0.0408	0.0031	0.0665	0.0275	0.0950
1	−0.0116	0.0031	0.0200	0.0275	0.0950
1	−0.0200	0.0031	−0.0039	0.0275	0.0950
6	0.2752	0.2752	0.0986	0.0986	−0.0748
1	−0.0125	−0.0476	0.0364	−0.0025	0.0472
1	−0.0807	−0.0476	−0.0310	−0.0025	0.0472
1	−0.0497	−0.0476	−0.0128	−0.0025	0.0472
			RRMS		
	0.0218	0.0305	0.0225	0.0330	0.0293
			Dipole and Quadrupole Moments		
μ^b	3.2544	3.8041	3.2476	3.8908	3.6063
$Q_{xx}{}^c$	2.7973	3.7808	2.6930	4.1805	4.9403
$Q_{yy}{}^c$	5.4703	5.3303	5.5428	5.0570	4.1157
$Q_{zz}{}^c$	−8.2676	−9.1110	−8.2358	−9.2375	−9.0560

[a] fr, free; ap, *a posteriori* averaged; eq, equivalenced; wk, weak restraints; st, strong restraints. (See text for explanation). [b] Dipole moment in debye, relative to center of mass. [c] Quadrupole moment along the principal axes in debye angstroms.

to keep the optimal wk.fr description, while allowing the other charges (which describe nonpolar regions) to readjust to the forced symmetry using the eq approach. Thus, the only centers allowed to readjust in this stage are methyl groups (N.B.: in all the other single-stage fits described by "eq", all charges including polar centers are allowed to adjust). Using the stronger st restraint (0.001 au) in this second stage is beneficial in further lowering the alkyl carbon charges. The two-stage fit is therefore denoted by wk.fr/st.eq.

The comparison of wk.fr/st.eq with the ap and fr fitting protocols for methanol, the tetrahedral intermediate, and for *trans*-NMA is given in Tables VII, IX, and X, respectively; visualization of the ESP points for comparison of the fits is shown for each molecule in Figures 4–6, respectively. As expected, the charge sets for *trans*-NMA and the tetrahedral intermediate behaved as in the case of methanol in that the restrained (wk) charge sets paralleled the unrestrained (un) sets in terms of the RRMS and electric moments, but the restrained charges were lower. Also, comparing the ap with the fr charges shows that the ap treatment always raised the dipole moment by 10–20% as well as significantly increasing the RRMS. The two-stage fit either reduced these increases or, in the case of methanol (which is very restricted in the second stage fit by having 1 degree of freedom), gave about the same results as wk.ap.

For methanol, comparing the ESP visualizations containing forced symmetry, the two-stage fit (Figure 4d) clearly gives a much improved fit of the polar areas compared to un.eq (Figure 4c). However, it appears almost identical to that of un.ap (Figure 4b), with very little error until the border region between the polar and nonpolar areas. Thus, for methanol, the two-stage fit offers little advantage in terms of reproducing the quantum mechanical ESP over the un.ap approach widely used prior to this work. Note, however, that it has achieved the same quality of fit with charges of lower magnitude overall (cf. Table VII), in particular with a markedly smaller charge on the methyl carbon.

For the tetrahedral intermediate, the wk.fr/st.eq fit (Figure 5b) gave a big improvement in the fit to the polar regions of the

TABLE X: Overall Comparison of Several Charge Fitting Schemes for *trans*-*N*-Methylacetamide

fitting scheme[a]	un.fr	un.ap	wk.fr	wk.ap	wk.fr/ st.eq
			Charges		
C1	−0.4902	−0.4902	−0.2361	−0.2361	−0.0411
H1(t)	0.1552	0.1320	0.0965	0.0727	0.0173
H1(g)	0.1206	0.1320	0.0608	0.0727	0.0173
C	0.7611	0.7611	0.5869	0.5869	0.5869
O	−0.6261	−0.6261	−0.5912	−0.5912	−0.5912
N	−0.5372	−0.5372	−0.4192	−0.4192	−0.4192
H	0.3208	0.3208	0.2823	0.2823	0.2823
C2	−0.0489	−0.0489	−0.0418	−0.0418	−0.2078
H2(t)	0.1117	0.0748	0.1073	0.0670	0.1127
H2(g)	0.0564	0.0748	0.0470	0.0670	0.1127
			RRMS		
	0.0680	0.1216	0.0742	0.1309	0.1168
			Dipole and Quadrupole Moments		
μ^b	4.1396	4.5722	4.1279	4.5890	4.4236
$Q_{xx}{}^c$	10.2871	10.2551	10.3146	10.3385	10.2170
$Q_{yy}{}^c$	−7.9054	−8.1862	−7.7881	−8.1311	−8.2343
$Q_{zz}{}^c$	−2.3817	−2.0689	−2.5265	−2.2074	−1.9827

[a] fr, free; ap, *a posteriori* averaged; eq, equivalenced; wk, weak restraints; st, strong restraints. (See text for explanation). [b] Dipole moment in debye. [c] Quadrupole moment along the principal axes in debye angstroms.

ESP compared to the un.ap (Figure 5a) and wk.ap charges (not shown). For *trans*-NMA, in comparing the un. ap charges (Figure 6a) with the wk.fr/st.eq charges (Figure 6b), while both give a good fit of the most exposed part of the polar areas, only the two-stage fit retains this good description into the regions bordering the nonpolar regions. Interestingly, with both *trans*-NMA and the tetrahedral intermediate the methyl charges underwent a sizable redistribution in the second stage fit. While the charges increased in magnitude on one methyl group, they decreased in magnitude on the other. In both cases this is accompanied by a marked improvement in the dipole moment and the RRMS compared to the ap fits. This is consistent with the effects of the restraints to lower charge magnitudes because the methyl carbon charges refitted in the second stage still come out with a lower overall magnitude, if their charge magnitudes are summed. This points to a distinct advantage of the second stage fit as opposed to the *a posteriori* averaging: charge can redistribute between methyls to accommodate the lowered degrees of freedom of the system in an optimal fashion. While the resulting charge sets look quite different, they all fit the quantum mechanical ESP well, with the two-stage fit offering the best set of charges from the standpoint of both (a) reproducing the QM ESP while having the necessary forced symmetry and (b) having charges of comparitively low magnitude to minimize the potential problems in other areas of the force field (cf. Introduction).

In the above, we have presented examples of the usefulness of the two-stage RESP model for molecules involving polar regions and methyl groups. How can we generalize this for any organic/biological molecule? Regarding the polar centers, by implication oxygen, nitrogen, the halogens, sulfur, and phosphorus, along with any hydrogens that may be attached to them, are to have their charges frozen in the second stage. We suggest that sp^2 and sp carbons should in general also be frozen in the second stage because their charges are likely to be much more well determined in the fit, in contrast to the more buried sp^3 carbons. Also by implication, methylene groups should be treated like methyl groups, having their charges re-optimized in the second stage of the fit for the same reasons outlined above concerning forcing symmetry on the methyl hydrogen charges.

The situation is less clear for methine groups, but allowing them to vary in the second stage has two advantages: (1) Increasing by 2 per methine the degrees of freedom available to the second stage fit can improve the overall fit to the electrostatic

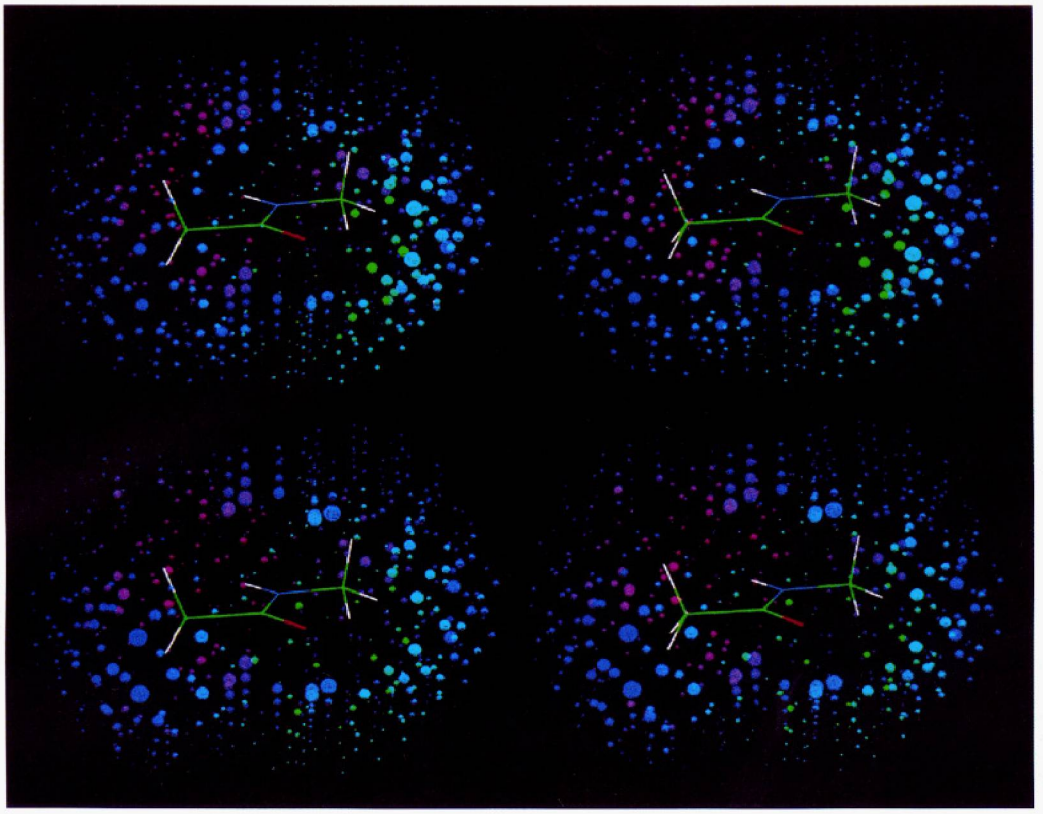

Figure 6. Visual comparison of the QM and calculated ESP for (a, top) the un. ap charges, and (b, bottom) the two-stage wk. fr/st. eq charges for *N*-methylacetamide (refer to the text for a description of the comparison).

potential, and (2) it is more logical from a chemical point of view to treat methines analogously to methyls and methylenes. Alternatively, one could keep methines frozen in the second stage, only refitting those charges for which it is necessary to force the symmetry (methyl and methylene groups). There may be benefits to the latter approach; for example, if methines are not frozen in the second stage, each methine carbon center will be restrained at twice the strength it experienced in the first stage. Also, by refitting methines in the presence of the penalty function with the stronger restraint, the charges on those atoms are sometimes reduced so as to minimize the penalty function at the expense of fitting the ESP. It remains to be seen whether or not there is a clear benefit to allowing the methine groups to vary in the second stage of the fit.

Conclusions

We have developed a model to restrain electrostatic potential derived charges using a simple penalty function. This model of charges (RESP) reduces the overall magnitude of the fitted charges, most often the statistically poorly determined charges, without seriously impairing the quality of the fit of the resulting electrostatic potential compared to that determined quantum mechanically. We have found that a hyperbolic restraint function is better than a harmonic, because in such a function the larger but better determined charges are not unduly penalized based on their magnitude.

We have also addressed the issue of how best to develop the forced charge symmetry for those atoms not equivalent by symmetry, but which exchange rapidly on a molecular dynamics time scale. Simply constraining the atoms to be equivalent during the fit has the undesirable side effect of reducing the magnitude of the ESP in the regions of the important polar atoms. On the other hand, averaging them after the fit gives a model with an enhanced dipole moment compared to that calculated quantum mechanically and a somewhat less accurate fit. As a compromise, a two-stage approach was developed in which the first stage fit to the potential is carried out with weak restraints but without forced symmetry on methyl groups, followed by a second stage refit of only those groups but now using forced symmetry and strong restraints.

In conjunction with quantum mechanical ESP data based on a 6-31G* wave function, this approach maintains the degree of dipolar enhancement needed to "balance" that of TIP3P and SPC water, a compromise necessary to adquately model polar interactions within a two-body additive intermolecular force field. As shown in the associated paper (ref 15), the RESP model behaves well in the calculation of both solvation free energies and intramolecular conformational energies.

The RESP partial charges not only perform well for the quantitative aspects of intra-/intermolecular interaction, but they are also suitable for qualitative (i.e. interpretive) uses in a way the unrestrained ESP charges have not been. The magnitude of the charges exhibit much smaller fluctuations between related functional groups, and the charges are consistent with chemical intuition while still reflecting the variations arising from the local chemical environment in a realistic and non–arbitrary way. The RESP approach developed here thus represents a powerful, general, and algorithmic method to derive atomic charges, especially for force field purposes.

Acknowledgment. This research was supported by Grants GM-29072 and NSF-CHE-91-13472 to P.A.K. and by a NSERC postdoctoral fellowship to C.I.B. P.C. was supported by DARPA (Grant MDA-91-J-1013) and partially by the Polish Committee for Scientific Research, KBN Grant No. 2 0556 91 01. Use of the Computer Graphics Lab (supported by NIH National Center for Research Resources RR-1081 to R. Langridge) is gratefully acknowledged. We are thankful for the quantum mechanical ESP-generating protocol developed by Ian Gould. The color images in Figures 4–6 were generated using the new "delegate" facility of the MidasPlus molecular modeling system.[20] We thank

Conrad Huang of the Computer Graphics Lab for suggesting the use of ellipsoids for illustrating the ESP residuals and for developing the new "discern" delegate module to MidasPlus used in generating these graphic representations.

References and Notes

(1) *MOTECC-90, Modern Techniques in Computational Chemistry*; E. Clementi, Ed.; ESCOM: Leiden, The Netherlands, 1990.
(2) Hagler, A.; Euler, E.; Lifson, S. *J. Am. Chem. Soc.* **1974**, *106*, 765.
(3) Tirado-Rives, J.; Jorgensen, W. *J. Am. Chem. Soc.* **1988**, *110*, 1657.
(4) Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219.
(5) Besler, B. H.; Merz, K. M.; Kollman, P. A. *J. Comput. Chem.* **1990**, *4*, 431.
(6) Momany, F. *J. Phys. Chem.* **1978**, *82*, 592.
(7) Cox, S. R.; Williams, D. E. *J. Comput. Chem.* **1981**, *2*, 304.
(8) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129.
(9) Faerman, C.; Price, S. *J. Am. Chem. Soc.* **1990**, *112*, 4915.
(10) Hariharan, P. C.; Pople, J. A. *Chem. Phys. Lett.* **1972**, *66*, 217.
(11) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.
(12) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230.
(13) Jorgensen, W. L.; Chandresekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(14) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. In *Intermolecular Forces*; Pullman, B., Ed.; Elsevier: Dordrecht, The Netherlands, 1981; pp 331–342.
(15) Cornell, W.; Cieplak, P.; Bayly, C. I.; Kollman, P. *J. Am. Chem. Soc.*, in press.
(16) Reynolds, C. A.; Essex, J. W.; Richards, W. G. *J. Am. Chem. Soc.* **1992**, *114*, 9075.
(17) Frisch, M. J.; Head-Gordon, M.; Tracks, G. W.; Foresman, J. B.; Schlegel, H. B.; Raghavachari, K.; Binkley, J. S.; Gonzales, C.; DeFrees, D. J.; Fox, D. J.; Whiteside, R. A.; Seeger, R.; Melius, C. F.; Baker, J.; Kahn, L. R.; Stewart, J. J. P.; Fluder, E. M.; Topiol, S.; Pople, J. A. *Gaussian 90*; Gaussian, Inc.: Pittsburgh, PA, 1990.
(18) Lees, R. M.; Baker, J. G. *J. Chem. Phys.* **1968**, *48*, 5299.
(19) Pearlman, D.; Case, D. A.; Caldwell, J.; Seibel, G.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. AMBER 4.0; University of California: San Francisco, 1990.
(20) Ferrin, T.; Huang, C.; Jarvis, L.; Langridge, R. The MIDAS Display System. *J. Mol. Graph.* **1988**, *6* (1), 13–27, 36–37. The MidasPlus software package is licensed by the Regents of the University of California and can be obtained by writing to the Computer Graphics Laboratory, University of California, San Francisco, CA 94143–0446.