Fast Assignment of Accurate Partial Atomic Charges: An Electronegativity Equalization Method that Accounts for Alternate Resonance Forms

Michael K. Gilson,*,†,‡ Hillary S. R. Gilson,† and Michael J. Potter†

VeraChem LLC, 20010 Century Boulevard, Suite 102, Germantown, Maryland 20874,
and Center for Advanced Research in Biotechnology, University of Maryland Biotechnology Institute,
9600 Gudelsky Drive, Rockville, Maryland 20850

Received July 21, 2003

A fast, accurate method of assigning partial atomic charges is described. The method is based upon the concept of electronegativity equalization and is parametrized to fit electrostatic potentials obtained from ab initio quantum calculations. A novel algorithm for identifying alternate resonance forms is used to ensure that chemically equivalent atoms are assigned equal charges. The resulting charges are independent of conformation, yield good agreement with ab initio electrostatic potentials, and are similar to standard force field charges for common biochemical components. The method is broadly parametrized and generates charges for a drug-like compound in about 0.45 s on a 2.26 GHz Pentium 4 PC. It should thus be useful in a range of applications, such as molecular design and QSAR. The resonance algorithm is expected to have additional applications, such as in atom-typing and detection of molecular symmetry.

1. INTRODUCTION

Most molecular modeling techniques rely upon an energy model, or force field, that yields the energy of a molecule as a function of its conformation. The energy contributions in a force field can be separated into bonded terms that account for bond-stretches, angle-bends, and torsional bond-rotations and nonbonded terms, which are typically treated as a sum of Lennard-Jones terms and Coulombic interactions among partial atomic charges. The Coulombic interactions account not only for general electrostatic interactions but also for hydrogen bonding. These nonbonded terms are of critical importance, playing a central role in determining molecular conformations and intermolecular binding affinities. However, although tested parameters (e.g., refs 1−8) are available for common biomolecular components, such as amino acids and nucleic acids, it can be difficult to obtain adequate parameters for the wide variety of chemistries found in synthetic compounds, such as drug candidates and synthetic hosts. This is especially true for atomic charges, which depend on the precise bonding environment in which each atom occurs. In fact, assigning accurate partial atomic charges has long been a significant challenge in the general use of modeling methods. Recent papers (e.g., refs 9−12) include excellent reviews of this topic, so only a focused review is offered here.

One of the most widely accepted approaches to generating physically reasonable partial atomic charges for new compounds involves fitting the charges to electrostatic potentials (ESPs) computed with ab initio quantum mechanics at sampling points around the molecule. These ESP methods are general, are well-founded theoretically, and have been shown to generate useful results.[13−17] However, they require minutes to hours of computer time for each molecule, so they are cumbersome for interactive molecular design and too slow for processing catalogs and databases containing thousands or millions of compounds. Also, ESP charges depend on the conformation of the molecule used in fitting the charges, partly because of real changes in the distribution of electrons, and partly because changes in conformation alter the distribution of sampling points. Of particular concern is that the ESP method frequently assigns different charges to chemically equivalent atoms, such as the hydrogens of a methyl group, while force-field based modeling requires that equivalent atoms be assigned equal charges. The Restrained Electrostatic Potential (RESP) method[9,10] addresses these problems by averaging atomic charges over chemically equivalent atoms and over representative molecular conformations. However, the additional steps in the RESP procedure impose further computational demands and pose the nontrivial problems of identifying chemically equivalent atoms and selecting representative conformations.

An alternative is to use rule-based methods, which classify atoms into types based upon their bonding environments and assign charges accordingly.[18] This approach is fast and accurate when sufficiently specific atom types are available in the parameter set. However, complex atom-typing algorithms may be required, and, more importantly, it is difficult to establish a set of types that is large enough to accommodate the diverse chemistries that are of interest. As a consequence, these methods often do not have specifically parametrized charges for all atoms in a molecule and are thus forced to drop back to general or default atom types that are not well parametrized. Also, there is no guarantee that the charges these methods assign to the atoms in a molecule will sum to the correct total charge. This problem has been addressed[19] by determining the difference between the net assigned charge of the molecule and the correct net charge and repairing the assigned charges by spreading the required difference charge uniformly over some or all of the

* Corresponding author phone: (240)686-0565; fax: (240)686-0564; e-mail: gilson@verachem.com.
† VeraChem LLC.
‡ University of Maryland Biotechnology Institute.

10.1021/ci034148o CCC: $25.00
Published on Web 11/04/2003

atoms. This approach formally solves the problem but is not physically motivated and may introduce error.

A closely related approach is to use bond-charge increments, which place net-neutral dipoles across bonds based upon the types of the atoms involved.[20,21] Such methods are fast but again require complex atom-typing algorithms and drop back to generalized default parameters when atoms or bonds are encountered that lie beyond the existing types. In addition, the bond-charge increment methods are not based upon a well-defined physical model and, probably for this reason, are subject to error. This has motivated the development of hybrid approaches in which atomic charges computed with semiempirical quantum methods such as AM1[22] are modified via parametrized bond-charge increments[12] or bond dipoles.[23,24] This important approach markedly improves accuracy but at the cost of computational performance, since semiempirical quantum calculations are time-consuming. It would thus be difficult to use such methods to calculate charges for very large numbers of compounds or in interactive applications that aim to provide a quick response time.

Another approach to assigning partial atomic charges uses the concept of equalization of electronegativities (e.g., refs 25−31). Such methods quantify the idea that highly electronegative atoms, such as oxygens, draw electrons away from less electronegative atoms, such as carbons. These methods are appealing because they are fast, are based upon a reasonable physical picture, and can be transferred across a broad range of molecules without the need to define a large number of different atom types. However, as detailed below, existing electronegativity equalization methods often generate charges that disagree markedly with the touchstone ab initio calculations described above. In addition, the treatment of formal charges has been problematic. Indeed, the pioneering work of Gasteiger and Marsili[26] does not define a method of treating ionized groups; and as previously noted,[11] the charges assigned to ionized groups by the QEq electronegativity equalization method[30] tend to be unrealistically small. Very recently, Bultinck and co-workers presented an electronegativity equalization method parametrized to match charges obtained by Mulliken or natural population analysis[32] of quantum calculations;[33,34] the method was reported to be less accurate when adjusted in relation to ESPs. As in the case of QEq, the charges depend on molecular conformation, so computing charges for a large compound database would require generating a reasonable three-dimensional conformation for each compound. Finally, ionic groups are not specifically considered, and the training and test sets do not appear to include any ions. It is thus of concern that this method may, like QEq, yield less accurate charges for ionic compounds.

Thus, there is still a need for a fast, accurate and broadly applicable method of computing accurate partial atomic charges. The present paper describes a new method that addresses this need. The method uses the electronegativity equalization approach but is parametrized specifically to reproduce ab initio molecular electrostatic potentials. It also uses a unique algorithm to ensure correct treatment of atoms whose equivalence is recognizable only when alternate resonance forms of a molecule are considered, such as the oxygens of a carboxylate group or more widely separated atoms in vinylogous molecules. The accuracy of the method is similar to that of the hybrid charge-assignment methods discussed above, but it is much faster, allowing accurate charges to be generated for ∼8000 drug-like compounds in an hour on a single-processor commodity PC. This combination of accuracy and speed will be valuable for a variety of computational chemistry tasks, including the selection of ligands from large compound databases, de novo ligand design, and calculation of electrostatic descriptors for use in quantitative structure−activity relationships. In addition, the method for identifying alternate resonance forms of molecules has broader applications in molecular modeling and chemical informatics.

2. METHODS

2.1. Charging Algorithm. This section discusses the treatment of molecules with only one important resonance form and then describes the handling of molecules with multiple resonance forms. The last parts of the section discuss parametrization, validation, and application to biomolecular components.

2.1.1. Molecules with a Single Important Resonance Form. Each atom in the molecule is assigned a type based upon its element, bonding pattern, aromaticity, and potentially a small number of additional features. Only certain combinations of these characteristics are encountered in the molecules of interest, so there is not a combinatorial explosion of types. Thus, the 39 atom types listed in Table 1 accommodate all but 8 of the 55 903 compounds in the December 2002 Maybridge HTS database. Existing methods are used to identify aromatic rings,[35,36] and formal charges are assigned based upon bonding patterns; for example, an oxygen with one single bond is assigned a formal charge of −1. Each atom type has its own initial electronegativity and hardness parameter. As described in Section 2.2, these parameters are not derived from first principles but are adjusted to optimize agreement with electrostatic potentials from ab initio quantum calculations.

Each atom i in the molecule is assigned the initial electronegativity e°_i and hardness s°_i associated with its type. The hardnesses are left at their initial values, but the electronegativities of the atoms are modified to new values e_i based upon their locations in the molecule, according to the following formula

$$e_i = e^\circ_i + \alpha_1 \sum_j^{\text{single}} S_{ij}|e^\circ_i - e^\circ_j|^\beta + \alpha_2 \sum_k^{\text{double}} S_{ik}|e^\circ_i - e^\circ_k|^\beta + \alpha_3 \sum_l^{\text{triple}} S_{il}|e^\circ_i - e^\circ_l|^\beta + \alpha_4 \sum_m^{\text{aromatic}} S_{im}|e^\circ_i - e^\circ_m|^\beta - \alpha_5 \sum_n^{1-3} S_{in}|e^\circ_i - e^\circ_n|^\beta$$

$$S_{ab} \equiv \frac{e^\circ_a - e^\circ_b}{|e^\circ_a - e^\circ_b|} \qquad (1)$$

Here the first 4 sums, respectively, run over atoms linked to atom i with single, double, triple, and aromatic bonds, aromaticity taking precedence over bond order; and the fifth sum runs over atoms in a 1−3 bonding relationship to atom i. The α coefficients, along with the exponent β, are

Table 1. Atom Types Used in the Charging Method and Their Fitted Initial Electronegativities (elecneg) and Hardnesses (hard)[a]

			number of bonds					fitted params	
ID	name	element	single	double	triple	charge	special features	elecneg	hard
1	H1	H	1	0	0	0	no	27.4	73.9
2	C3	C	4	0	0	0	no	30.8	78.4
3	C2	C	2	1	0	0	no	33.6	76.4
4	C1a	C	0	2	0	0	no	37.0	65.3
5	C1b	C	1	0	1	0	no	40.0	98.5
6	Car	C	2	1	0	0	aromatic	34.6	84.7
7	O3	O	2	0	0	0	no	45.7	92.6
8	O2	O	0	1	0	0	no	49.5	86.1
9	O3n	O	1	0	0	−1	no	49.3	25.0
10	Oar	O	2	0	0	0	aromatic	45.9	137.
11	N3	N	3	0	0	0	no	44.0	87.6
12	N3s	N	3	0	0	0	planar	43.6	94.4
13	N2	N	1	1	0	0	no	44.0	72.7
14	N1	N	0	0	1	0	no	57.0	111.
15	N3p	N	4	0	0	1	no	42.8	188.
16	N2p	N	2	1	0	1	no	37.6	41.5
17	N1pa	N	0	2	0	1	no	24.0	104.
18	N1pb	N	1	0	1	1	no	39.4	29.7
19	Nar3	N	3	0	0	0	aromatic	43.4	136.
20	Nar2	N	1	1	0	0	aromatic	53.0	102.
21	Narp	N	2	1	0	1	aromatic	38.7	8.64
22	N1m	N	0	1	0	−1	no	31.9	129.
23	N2m	N	2	0	0	−1	no	28.3	20.9
24	N2mR	N	2	0	0	−1	planar	43.6	0.176
25	Cl3	Cl	1	0	0	0	no	37.6	53.5
26	F3	F	1	0	0	0	no	45.2	96.8
27	Br3	Br	1	0	0	0	no	40.1	75.3
28	S3	S	2	0	0	0	no	37.4	69.1
29	S3p	S	3	0	0	1	no	31.8	93.9
30	S4	S	2	1	0	0	no	35.8	93.1
31	S6	S	2	2	0	0	no	31.7	83.2
32	Sar	S	2	0	0	0	aromatic	33.8	88.9
33	S3n	S	1	0	0	−1	no	44.5	24.8
34	S2a	S	0	1	0	0	no	47.5	74.3
35	P3	P	3	0	0	0	no	37.9	72.5
36	P3p	P	4	0	0	1	no	29.6	108.5
37	P5	P	3	1	0	0	no	33.0	86.6
38	I	I	1	0	0	0	no	41.3	109.0
39	Ip	I	2	0	0	1	no	34.1	10.8

[a] Columns 4−6 list the number of single, double and triple bonds to the atom, column 7 lists the atom's formal charge, and column 8 notes any special features of the atom type, where aromatic indicates that the atom is part of an aromatic ring, and planar indicates that the atom is in a planar ring.

parameters that are optimized during the fitting procedure described below. The first 4 modifications of the initial electronegativies here allow for an increase in the polarization of two atoms of unlike electronegativity that are bonded to each other, such as the C and O of a carbonyl group. The effect of this adjustment in electronegativity is similar to that of the screened Coulombic term used in the QEq electronegativity equalization method.[30] However, the present approach eliminates the dependence of the partial atomic charges upon molecular conformation that is characteristic of QEq. Also, the present approach saves computer time during the charge calculation because no charge−charge interactions appear in the energy model (eq 2). The fifth term in eq 1 decreases the transfer of charge between atoms in a 1−3 bonding relationship and is found empirically to improve the ability of the present model to fit the elelectrostatic potentials from ab initio quantum calculations.

A "charge group" is then defined around each formally charged atom. The charge group comprises the formally charged atom itself and every atom to which it is directly bonded. Each such charge group is assigned a nominal charge equal to the formal charge of its formally charged atom. If two such initial charge groups share one or more atoms, then the groups are merged into a single group comprising all the atoms of both groups and with a nominal charge equal to the sum of the nominal charges of the two groups. After all such mergers are completed, any charge group whose nominal charge has become zero is eliminated.

Finally, the atomic charges are calculated by minimizing the following function with respect to the charges

$$E = \sum_i \left(e_i q_i + \frac{1}{2} s_i^{\circ} q_i^2 \right) \tag{2}$$

subject to constraints explained later in this paragraph. Intuitively, E may be viewed as an energy that depends on the atomic charges. The more electronegative an atom, i.e., the greater e_i, the more the energy is lowered by movement of negative charge q_i to the atom. The greater the hardness s_i of an atom, the more it resists accumulation of either positive or negative charge. The minimization of E is subjected to $1 + 2N_{\text{grp}}$ constraints, where N_{grp} is the number of charge groups in the molecule. The first constraint is that the total charge of the molecule must equal the sum of its

formal charges. The $2N_{grp}$ additional constraints are that the total charge of each charge group $j \in [1, N_{grp}]$ must equal the nominal charge Q_j associated with the group, plus or minus a "charge bleed" parameter δ which is constant across all groups and is adjusted as part of the overall parameter setting process (Section 2.2). Thus, up to $\pm\,\delta$ electrons of charge is allowed to bleed out of each charge group and into the rest of the molecule. These constraints can be written as follows:

$$\sum_{i=1}^{N_{atoms}} q_i = \sum_{j=1}^{N_{grp}} Q_j \tag{3}$$

$$Q_j - \delta < \sum_{i}^{N_j} q_i < Q_j + \delta \tag{4}$$

Here, q_i is the charge on atom i, the sum over i ranges over all atoms in the molecule N_{atoms}, and Q_j is the nominal charge of charge group j. Hence, eq 3 expresses the constraint on the total charge of the molecule. In eq 4, $\sum_i^{N_j}$ indicates a sum over all N_j atoms in charge group j, and δ is the bleed parameter.

In summary, the charges generated by the present model are those which minimize the energy in eq 2, subject to the equality constraint in eq 3 and the $2N_{grp}$ inequality constraints in eq 4. The simplicity of the expression to be minimized, eq 2, suggests that the highly efficient method of Lagrangian multipliers might be useful here, but this method of cannot by itself solve a problem with inequality constraints, such as those in eq 4. However, the method of Lagrangian multipliers can be applied once it is recognized that the inequality constraints represent boundaries of the solution space of charges and that the charges that solve the constrained minimization problem must lie either within, or on, the boundaries. As a consequence, the solution can be obtained by applying Lagrangian multipliers with charge group's inequality constraints either left inactive, activated as an equality constraint at the upper end of its range ($q_i = Q_j + \delta$), or active as an equality constraint at the lower end of its range ($q_i = Q_j - \delta$). (The constraint on the total charge is applied in every case.) Thus, $3^{N_{grp}}$ minimizations are carried out, yielding $3^{N_{grp}}$ different charge distributions, each corresponding to its own value of E; the correct solution to the constrained optimization problem, as defined in eqs 2, 3, and 4, is just the charge distribution corresponding to the lowest value of E. Although this method could become time-consuming for a molecule with many charge groups, it was found to be quite efficient for the test cases studied here.

As previously noted,[37] minimization of E in eq 2 subject to the constraint on total charge (eq 3) is mathematically equivalent to equalization of electronegativity as originally proposed by Sanderson,[25] in which the electronegativity of each atom is considered to diminish as its electronic charge increases.[25] Indeed, the matrix equation obtained from applying the method of Lagrangian multipliers to this problem is identical to that obtained from the equalization of electronegativities subject to the same constraint. Thus, the two approaches are, in reality, the same, although formulated differently. The additional constraints on the charge of ionizable groups developed here (eq 4) could also be incorporated into the equalization of electronegativities formalism to yield, again, the same mathematical problem and the same numerical results.

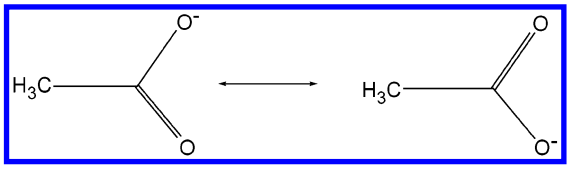

Figure 1. Resonance forms of acetate ion.

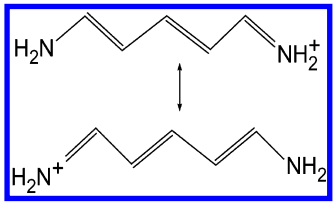

Figure 2. Resonance forms of a vinylogous ion.

2.1.2. Molecules with Multiple Resonance Forms. Computer representations of molecules typically store only a single resonance form, but many molecules can be represented in multiple resonance forms, each of which contributes to the electronic structure. For example, the two resonance forms of the acetate molecule in Figure 1 contribute equally to its charge distribution, so the two oxygens are chemically equivalent and must be assigned equal partial charges. Computing charges based upon the bonding pattern of only one resonance form would lead, incorrectly, to the assignment of different atomic charges to the two oxygens. Because the two oxygens are close to each other, and a carboxylate is a standard chemical group, many charging algorithms handle carboxylate groups correctly by recognizing them as a special case and treating them accordingly. However, a more general treatment of resonance forms is essential in more complex cases. For example, the two nitrogens in the vinylogous compound in Figure 2 do not form a recognized chemical group. Nonetheless, from the two resonance forms shown in the figure, it is clear that the two nitrogens are chemically equivalent and must be assigned equal partial atomic charges. To our knowledge, no algorithm has been described previously for identifying such alternate resonance forms, and thus no prior method of assigning partial atomic charges accounts correctly, generally, and automatically for the existence of alternate resonance forms. A way of identifying alternate resonance forms and using them in the calculation of partial atomic charges is now described.

Classification of Changes in Resonance Form. Changes in resonance form can be divided into two types. The first type involves a formal electron transfer one atom to another, and the second type involves changes in bond order around an aromatic ring without any net electron transfer between atoms. Only the first type of resonance change must be dealt with explicitly in the charging method, as now discussed.

The first type of change in resonance form, which involves formal electron transfer from an electron donor atom to an acceptor atom, requires that the donor and acceptor be connected by a pathway through the molecule made up of an odd number of atoms and an even number of bonds whose bond orders meet criteria described below. When the transfer occurs, the donor becomes an acceptor and the acceptor becomes a donor, while the orders of the bonds in the path

either rise or fall by 1 in an alternating pattern along the path. The carboxylate group in Figure 1 is a simple example of this type of change: in the initial resonance form, the negatively charged oxygen is the donor, the neutral oxygen is the acceptor, and the path contains 1 atom—the carbon of the carboxylate group—and 2 bonds. When the resonance form changes, the bond orders change as shown, the donor becomes its conjugate acceptor, and the acceptor becomes its conjugate donor. For the compound in Figure 2 the initial donor (neutral nitrogen) and acceptor (cationic nitrogen) are joined by a path consisting of 5 atoms and 6 bonds. This type of change in resonance form must be dealt with in the present charging algorithm because the changes from donor to acceptor and acceptor to donor alter atom types as defined in Table 1.

The second type of resonance change, in which the bond orders around a ring are alternately incremented and decremented by one without any net electron transfer between atoms, can occur in an aromatic ring, such as a benzene group. (However, some aromatic rings, such as furan, do not permit such a change.) This type of resonance change does not need to be dealt with in the present charging algorithm because, although it alters bond orders, it does not alter atom types as defined in Table 1, since the number of bonds of each order connected to the atoms involved remains fixed. In addition, all aromatic bonds—whether formally represented as single or double—are treated identically in making the electronegativity changes defined by eq 1, so shifting single and double bonds within an aromatic ring does not influence the resultant electronegativities e_i. Finally, an aromatic ring with alternating bond orders provides an equally good path for resonance changes of the first type (electron transfer—see previous paragraph) no matter which resonance form of the ring itself is considered. For these reasons, the charging algorithm described above does not require that resonance changes of the second type be identified, so this type of resonance change is not considered further here.

Identification of Alternate Donor/Acceptor Resonance Forms. This section describes what is, to our knowledge, the first general algorithm for identifying alternate donor/acceptor resonance forms of a molecule. The algorithm takes as input a single resonance form of the molecule, with an explicit integer bond order for each bond. Thus, "aromatic" or "resonant" are not acceptable bond types, so carboxylates and benzenes, for example, must be represented in a single one of their resonance forms. The algorithm also requires a table that enables potential electron donor and acceptor atoms within the molecule to be identified and that assigns an energy to each donor and acceptor atom type. For example, an oxygen of zero formal charge and having one bond of order 2 is an electron acceptor and has a resonance energy of zero, in arbitrary energy units. Upon accepting an electron, it is converted into its conjugate donor, an oxygen of formal charge −1 with one bond of order 1 and resonance energy 5. This conjugate donor type is also listed in Table 2, along with other donors and acceptors, and the correspondences between conjugate pairs. This table can optionally be expanded to include additional donors and acceptors.

A new resonance form is generated when a donor formally transfers an electron to an acceptor, causing the donor to become its conjugate acceptor and the acceptor to become its conjugate donor (see conjugate IDs in Table 2), while the orders of the bonds connecting the donor and acceptor along a single path change in a prescribed manner: the orders rise by 1 for the odd-numbered bonds in the sequence, starting with the bond to the donor, and the orders of the even numbered bonds in the sequence fall by 1. In order for these changes in bond order to be permitted, the path connecting the donor and acceptor must contain an odd number of atoms and thus an even number of bonds. In addition, the changes in bond order must not raise the order of a bond above 3 (2 in the case of oxygen), and no bond whose order is lowered by 1 can start off as a single bond, because then lowering its order would break it. If a donor and acceptor are connected by a path meeting these criteria, then an alternate resonance form exists and is generated as just described.

Table 2. Definitions of Donor and Acceptor Atoms for Identification of Alternate Donor/Acceptor Resonance Forms[a]

ID	element	formal charge	no. of bonds	bond orders	donor/ acceptor	energy	ID of conjugate
1	O	0	1	2	A	0	2
2	O	−1	1	1	D	5	1
3	S	0	1	2	A	0	4
4	S	−1	1	1	D	5	3
5	N	+1	3	2,1,1	A	5	6
6	N	0	3	1,1,1	D	0	5
7	N	0	2	2,1	A	0	8
8	N	−1	2	1,1	D	5	7
9	N	0	1	3	A	0	10
10	N	−1	1	2	D	5	9

[a] Bond orders: list of the orders of all bonds to the atom; ID of conjugate: ID number of donor/acceptor in this table that is conjugate to the listed acceptor/donor atom.

A notation for describing alternate donor/acceptor resonance forms is now introduced. If a molecule has $i = (1, 2,..., N_f)$ difference resonance forms, and $j = (1, 2,..., N_{d/a})$ atoms that are potential donors or acceptors, then its resonance forms can be described by a vector $\mathbf{f_i}$ having $N_{d/a}$ components: $\mathbf{f_i} = (f_{i,1}, f_{i,2},...f_{i,N_{d/a}})$, where component $f_{ij} = 1$ if donor/acceptor j is in its donor state in resonance form i, and $f_{ij} = 0$ if donor/acceptor j is in its acceptor state in resonance form i. For example, the two forms of the carboxylate in Figure 1 are described by the resonance vectors $f_1 = (0, 1)$ and $f_2 = (1, 0)$. Figure 3 shows a more complex case with 6 potential donors and acceptors and the following 6 resonance forms, numbered as in the figure:

$$f_1 = (1, 0, 0, 1, 0, 1)$$

$$f_2 = (1, 1, 0, 0, 0, 1)$$

$$f_3 = (0, 1, 0, 1, 0, 1)$$

$$f_4 = (1, 0, 0, 0, 1, 1)$$

$$f_5 = (1, 1, 1, 0, 0, 0)$$

$$f_6 = (0, 1, 0, 0, 1, 1)$$

(It is only coincidental that the number of donor/acceptors equals the number of resonance forms in these examples.)

Given an initial resonance form, all possible alternative resonance forms are discovered by the following method,

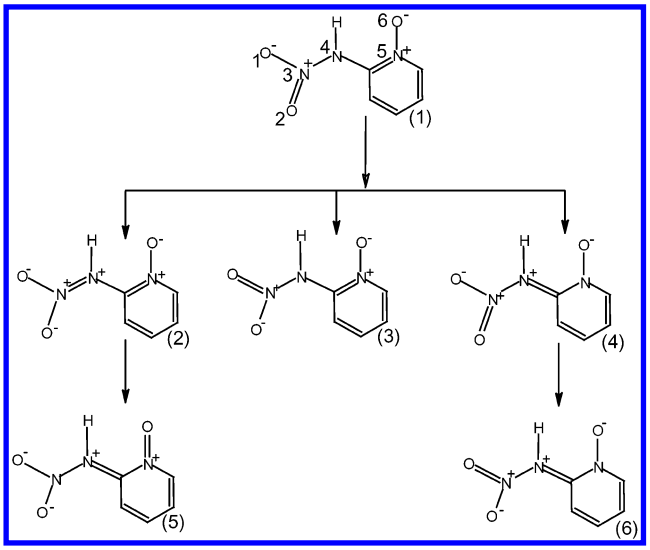

Figure 3. Complex resonance system, with donor and acceptor atoms numbered in initial structure **(1)** at top. Two generations (second and third rows), in addition to the initial resonance form (first row), are required to identify the full set of 6 resonance forms.

which is also diagrammed in Figure 4. The alternative resonance forms are generated in successive generations, starting with the solitary input form which is the first generation $i = 1$. Generations $i > 1$ may include multiple resonance forms, and the total number of generations that will be generated is not known at the outset. Starting with generation $i = 1$, and then repeating for successive generations $i > 1$, the procedure described in the next paragraph is used to identify and carry out all possible formal electron transfers for every form in generation i, thus creating a new generation $i = i + 1$ of resonance forms. All forms generated in this way that do not already exist in any previous generation are saved, but repeats are discarded. Two resonance forms are considered identical when they have the same resonance vector, $\boldsymbol{f}$, as defined above. Successive generations are created until a generation is reached that contains no new forms, at which point the algorithm stops.

The procedure for identifying the formal electron transfers that are possible for a given resonance form begins by using the criteria in the donor/acceptor table—here Table 2—to identify all atoms that are of donor or acceptor type. For each donor atom, a search is carried out for any path of atoms that connects to an acceptor atom and that meets the criteria for an electron-transfer path given above. If such a path exists, a formal electron transfer is carried out from the donor to the acceptor along the path to generate an alternate resonance form in the next generation of forms.

Averaging of Electronegativities and Hardnesses over Resonance Forms. Not all resonance forms contribute equally to the electronic structure of a molecule, and the present charging algorithm focuses on the more important resonance forms. For example, although an amide group has two resonance forms (Figure 5), the form in which both oxygen and nitrogen carry formal charges contributes less to the electronic structure than that in which both atoms are formally neutral. Differences in importance are identified as follows. Each type of donor and acceptor atom is assigned a resonance energy, in arbitrary units, as listed in the donor/acceptor definition table (Table 2), and the energy of a given resonance form is just the sum of the energies of its donors and acceptors. For example, the two resonance forms of a carboxylate group are of energy 5 because the formally charged oxygen atom (ID 2 in Table 2) that exists in both forms has an energy of 5, while the other oxygen has an energy of zero. As a consequence, both resonance forms are considered equally important. In contrast, when the initially neutral nitrogen of an amide group (ID 4) donates an electron to the initially neutral oxygen (ID 1), the resulting conjugate acceptor and donor (IDs 3 and 2, respectively) are each of energy 5 so the new charge-separated resonance form is less stable than the initial form by $5 + 5 = 10$ energy units. Thus, the resonance energy each resonance form of a molecule is determined by summing the energies of its donor and acceptor atoms as listed in Table 2, and the lower energy resonance forms are considered to contribute most to the overall electronic structure of the molecule.

At least for simple resonance systems, any reasonable set of energy values will correctly identify the key resonance forms. For example, the equivalence of the two resonance forms of a carboxylate will be recognized no matter what energy is assigned to the formally charged oxygen atom; and the secondary status of the charge-separated form of an amide will be recognized so long as the formally charged form of an atom is assigned a higher energy than the neutral form. The specific energies listed in Table 2 were selected based upon a combination of chemical intuition and adjustments to optimize the calculated partial atomic charges; other energy values and schemes could also be used. Note that the resonance energies in Table 2 are distinct from the quantity E in eq 2.

In calculating charges by the present method, the energies of all the resonance forms of the molecule are compared, and only those resonance forms $k = (1, 2, \ldots, N_{low})$ at the lowest energy level—for example both forms of a carboxylate but only the form of an amide that is low in energy because it has no formal charges—are used to compute the atomic charges. Atom types (Table 1) are assigned to each such form k, and eq 1 is used to compute the electronegativity, e_{ik}, of each atom i in resonance form k. Hardnesses, s_{ik}, also are assigned without modification from Table 1. These electronegativities and hardnesses are then averaged over the N_{low} low-energy resonance forms to yield an electronegativity and a hardness for each atom i that reflects the contributions of all the low-energy resonance forms:

$$e_i = \frac{1}{N_{low}} \sum_{k=1}^{N_{low}} e_{ik}$$

$$s_i = \frac{1}{N_{low}} \sum_{k=1}^{N_{low}} s_{ik} \qquad (5)$$

This procedure yields a single averaged value for the electronegativity and hardness of each atom for use in eq 2. Note that, after this average is taken, the two oxygens of a carboxylate have equal electronegativities and equal hardnesses, as is physically appropriate. It is also worth mentioning that other averaging methods could be used; for example, it would be possible to compute partial atomic charges separately for each low-energy resonance form and then average the charges.

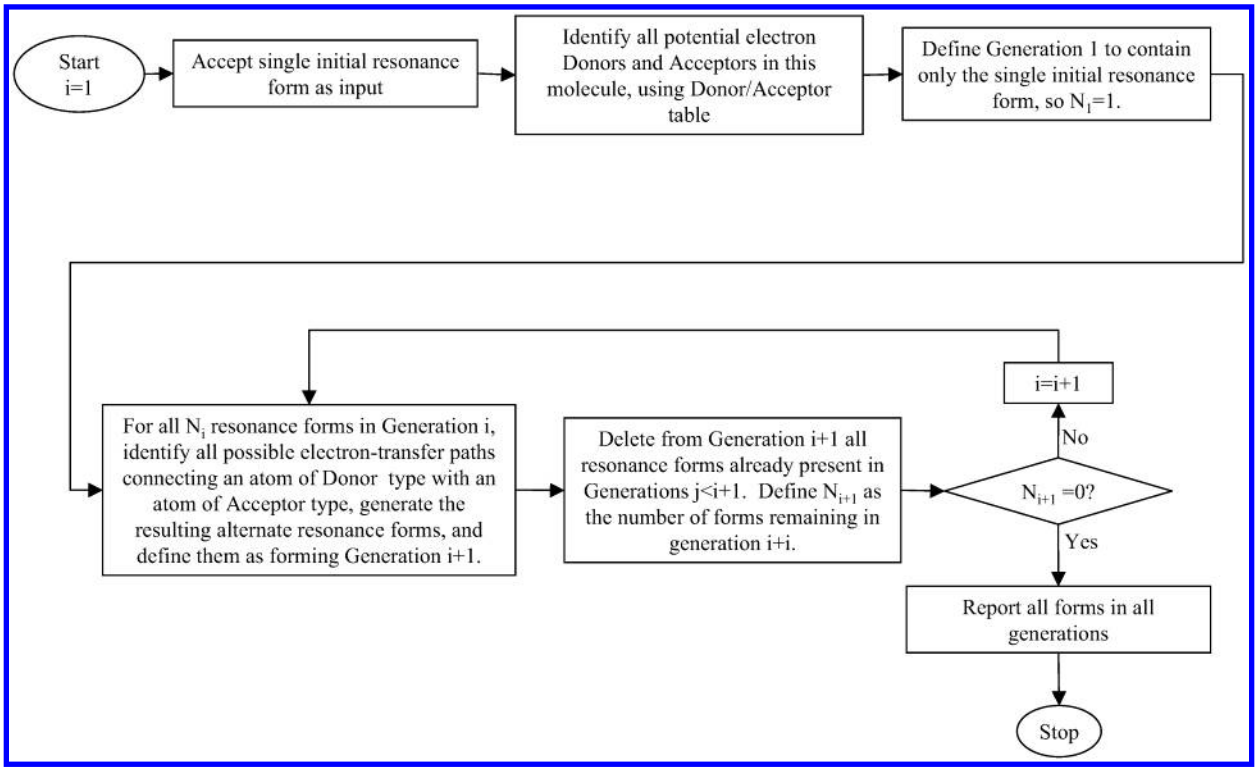

Figure 4. Flowchart of algorithm for identifying alternate resonance forms of compounds.

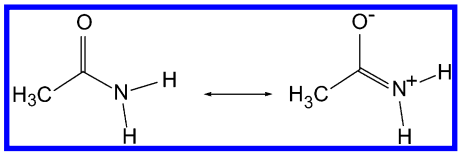

Figure 5. Resonance forms of a amide group.

Note that the procedure for identifying all resonance forms, described above and diagrammed in Figure 4, does not eliminate the high-energy resonance forms until the end of the procedure. This is because it is sometimes necessary to go through a high-energy form in order to identify a low-energy form. Figure 3 shows an example. All the resonance forms in the figure have four formally charged atoms and hence an energy of $4 \times 5 = 20$ arbitrary energy units (see Table 2), except for form **2**, which has 2 additional formal charges for an energy of 30 energy units. Although **2** is a high-energy form that will be discarded when determining the average electronegativities and hardnesses of the atoms (see previous paragraph), form **2** is required as a step in the identification of form **5**, which is of energy 20 and hence will contribute to the averages in eqs 5.

Merging of Charge Groups for Multiple Resonance Forms. As described above (Section 2.1.1), for molecules with only a single resonance form, a charge group is defined around each formally charged atom, and a constraint is applied to keep most of the formal charge within this group of atoms (eq 4). For molecules with multiple low-energy resonance forms, however, a formal charge can move from one atom to another, depending upon the resonance form. This is addressed by creating a separate charge group for each atom that bears formal charge in any resonance form, and assigning fractional nominal charges depending upon the fraction of resonance forms in which the atom is formally charged. For example, the molecule in Figure 2 is assigned two charge groups of nominal charge 0.5, each comprising a nitrogen atom, 2 hydrogen atoms, and a carbon atom. Charge groups that overlap are then merged as described in Section 2.1.1. Thus, the carboxylate in Figure 1 is preliminarily assigned two charge groups of charge −0.5, each comprising an oxygen atom and a carbon atom. However, because the groups share the carboxylate carbon atom, they are merged to a single charge group of nominal charge −1. Furthermore, if the methyl group in the acetate ion Figure 1 were replaced by an ammonium group, which has a formal charge of +1 on the nitrogen atom, then the preliminary charge group associated with the ammonium would include the carboxylate carbon. As a consequence, the ammonium and carboxylate charge groups would overlap, leading to a merger of 3 charge groups with zero net charge, so all the charge groups would be eliminated as described in Section 2.1.1.

2.2. Optimization of Parameters. Parameters of the present charging model were adjusted to maximize the accuracy of electrostatic potentials around a set of representative molecules as computed with the charges generated by the present charging model, relative to reference potentials computed via ab initio quantum mechanics. The molecular weights of the 284 molecules used in this study range between 17 and 374 Daltons, with a mean of 199 Daltons. These compounds were divided into a parametrization set of 253 molecules and a smaller test set of 31 molecules, where both sets contained at least one instance of each atom type in Table 1. The breakdown of atom types in these sets is detailed in Table 3.

HyperChem Lite (Release 1.0, Hypercube, Inc., 1996) was used to set up each molecule, and the program GAMESS[38] was then used to further minimize the energy with respect to conformation, i.e., to generate a stable, low-energy conformation, and to compute electrostatic potentials at sampling points around each molecule. All quantum calculations were done at the 6-31G* level, except that the SBKJC

Table 3. Composition of Training and Test Sets of Molecules[a]

	number of molecules			number of atoms		
type	training set	test set	total	training set	test set	total
H1	249	30	279	1693.0	194.0	1887.0
C3	178	17	195	463.0	42.0	505.0
C2	121	13	134	227.2	24.3	251.5
C1a	10	2	12	11.5	1.3	12.8
C1b	19	2	21	26.5	1.7	28.2
Car	68	14	82	364.8	82.7	447.5
O3	89	4	93	135.3	7.0	142.3
O2	114	12	126	157.5	15.9	173.4
O3n	35	5	40	34.5	5.1	39.6
Oar	3	1	4	2.7	1.0	3.7
N3	55	4	59	69.3	4.0	73.3
N3s	19	4	23	23.7	4.3	28.0
N2	24	5	29	28.3	4.3	32.6
N1	19	4	23	18.0	2.7	20.7
N3p	17	1	18	18.0	1.0	19.0
N2p	7	3	10	6.7	3.0	9.7
N1pa	6	3	9	2.5	2.0	4.5
N1pb	6	2	8	3.5	1.0	4.5
Nar3	16	2	18	16.0	1.7	17.7
Nar2	24	4	28	42.2	3.2	45.3
Narp	6	1	7	5.3	1.0	6.3
N1m	7	4	11	3.0	2.3	5.3
N2m	3	2	5	1.7	0.8	2.4
N2mR	2	2	4	0.8	0.8	1.7
Cl3	18	1	19	33.0	1.0	34.0
F3	5	1	6	13.0	3.0	16.0
Br3	9	2	11	10.0	5.0	15.0
S3	28	3	31	34.5	3.0	37.5
S3p	5	1	6	5.0	1.0	6.0
S4	9	1	10	9.0	1.0	10.0
S6	16	1	17	16.0	1.0	17.0
Sar	8	2	10	7.5	2.0	9.5
S3n	4	2	6	3.0	2.0	5.0
S2a	12	2	14	11.0	2.0	13.0
P3	6	1	7	6.0	1.0	7.0
P3p	2	1	3	2.0	1.0	3.0
P5	15	1	16	15.0	1.0	16.0
I	5	2	7	5.0	3.0	8.0
Ip	3	1	4	3.0	1.0	4.0

[a] Number of molecules: number of molecules containing atom of specified type (see Table 1); number of atoms: number of atoms of specified type across all molecules in the training, test, or combined sets. Fractional numbers result from averaging of resonance forms. For example, an atom type present in only 1 of 2 resonance forms contributes 0.5 to the number of atoms.

effective core potential was used for iodine. The electrostatic potentials were computed with the CHELPG method,[17] as implemented in GAMESS, with RMAX = 3 Å and DELR = 0.8 Å, where RMAX is the maximum distance of any point to the closest atom and DELR is the distance between neighboring sampling points.

The following 85 parameters were adjusted: the values of e° and s° for each atom type in Table 1; and the global parameters α_1, α_2, α_3, α_4, α_5, β in eq 1, and δ in eq 4. The error in the electrostatic potential for molecule i was defined as

$$D_i \equiv \left(\frac{\sum_{j=1}^{n^{\phi}_i} (\phi_{ij}^{\text{quantum}} - \phi_{ij}^{\text{model}})^2}{n^{\phi}_{\text{i}}} \right)^{1/2} \quad (6)$$

and the error for all $N_{\text{par}} = 253$ molecules in the parametrization set is computed as

$$D_{\text{par}} \equiv \left(\frac{\sum_{i=1}^{N_{\text{par}}} D_i^2}{N_{\text{par}}} \right)^{1/2} \quad (7)$$

where n^{ϕ}_i is the number of sampling points around molecule i, $\phi_{ij}^{\text{quantum}}$ is the electrostatic potential at point j around molecule i from the quantum calculation, and ϕ_{ij}^{model} is the electrostatic potential at sampling point j of molecule i computed based upon the present charging model using Coulomb's law in vacuo. The optimization algorithm is based upon the optimizer used in the Mining Minima algorithm,[39] but other global optimization methods could also be used. The optimized parameters were tested by computing the measure of error defined in eq 7, but now using the N_{test} test molecules, rather than the N_{par} molecules in the parametrization set; the resulting measure of error is termed D_{test}.

2.3. Application to Amino Acids and Nucleotides. To permit comparison with existing force fields, charges were computed by the present method for the 20 common amino acids, with several variant ionization states, and for the four common deoxyribonucleotides. Each individual amino acid was capped at the N-terminus with a CH_3CO group and at the C-terminus with an $NHCH_3$ group in order to form amide groups as if the residue were part of a protein. The appropriate charge constraint was applied to the molecule as a whole, rather than to the amino acid moiety, so the net charges of the amino acid moieties reported here need not be integer. The deoxyribonucleotides were similarly capped at the 3′ and 5′ hydroxyls with PO_3CH_3 groups to mimic residues in a chain, and charges are reported only for one repeating unit. As in the case of the amino acids, only the total charges were constrained, so the charges of the atoms shown here need not sum to an integer total charge.

3. RESULTS

Section 3.1 evaluates the atomic charges generated by the present method with its current parameters against the results of ab initio quantum calculations and also compares with other charging models and force fields. Section 3.2 discusses the values of the optimized parameters, Section 3.3 reviews results from the method of generating alternate resonance structures, and Section 3.4 examines the speed and applicability of the method. Here "VC" indicates the present method of computing partial atomic charges, while "VC/2003" denotes the present method along with the present parameters.

3.1. Charges. 3.1.1 Accuracy of VC/2003 Charges. The accuracy of VC/2003 charges is assessed by comparison between ESPs computed from these charges with those obtained from ab initio quantum calculations, as discussed in Section 2.2. It is of interest to compare the VC/2003 charges with those obtained by CHELPG fitting,[17] which represent the mathematically optimal charges for reproducing the ab initio ESPs; that is, the atomic charges that minimize D_i (eq 7) for each molecule i. (Note that CHELPG charges, unlike VC charges, are not suitable for use in molecular simulations because they depend on molecular conformation and do not constrain chemically equivalent atoms to have equal charges.)

Extensive optimization led to a parameter set for which the errors on both the parametrization and test sets are very low, $D_{par} = 3.44$ and $D_{test} = 4.72$ kcal/(mol-electron), while for the parametrization and test sets combined, the overall error is 3.60. For comparison, CHELPG charges yield overall errors D_i of $D_{par} = 1.93$ and $D_{test} = 2.20$ kcal/(mol-electron) for the parametrization and test sets, respectively. The largest value of D_i for any molecule in the parametrization set is $D_i = 9.50$, for thiocyanate. CHELPG yields $D_i = 3.6$ for this compound and the root-mean-square deviation between the atomic charges themselves for VC/2003 and CHELPG is 0.29 electrons. The largest value of D_i from VC/2003 for any molecule in the test set is 12.09, for *p*-diiodobenzene. CHELPG yields $D_i = 6.11$ for this compound, and the root-mean-square deviation between the atomic charges themselves for VC/2003 and CHELPG is 0.05 electrons. The largest value of D_i from CHELPG for a compound in the parametrization and test set, respectively, are 5.77 for iodopyridine and 6.11 for *p*-diiodobenzene. The VC/2003 charges can also be compared directly to the CHELPG charges: the root-mean-squared deviation of the VC/2003 charges from the CHELPG charges across all 284 molecules is 0.17 electrons, and the maximum charge deviation for any molecule is 0.464 electrons, for 3-iminoallen-1-one. Interestingly, the root-mean-squared deviation of the VC charges from Mulliken and Lowdin charges are significantly lower, at 0.14 and 0.21 electrons, respectively. For this molecule, the VC/2003 charges give an RMSD relative to the quantum electrostatic potentials of 5.54 kcal/(mol-electron), and the CHELPG charges give a deviation of 2.10 kcal/(mol-electron).

Overall, then, VC/2003 charges do a good job of reproducing potentials from ab initio quantum calculations, and the differences between VC/2003 and CHELPG charges result in part from the additional constraints on the VC/2003 charges, which are required to be independent of conformation and the same for chemically equivalent atoms. The following section compares VC/2003 charges with other charging methods that are subject to the same constraints.

3.1.2. Comparison with Other Charging Models. The accuracy of the VC method in reproducing ab initio quantum potentials is compared with that of several other charging models in Table 4. Published atomic charges[11] from various models were used to compute values of D_i, based upon the ab initio electrostatic potentials (Section 2.2). These are compared with the values of D_i from CHELPG and VC/2003 charges. The VC method, which is applicable to a broad range of compounds and uses 85 adjustable parameters, yields more accurate potentials than any other method tested here except for CHELPG which, as noted above, is not suitable for a force field. The MMFF[11] and OPLS[40] charges are also quite accurate, but the MMFF charging method requires a large number ($N \sim 500$) of parameters[11] and does not, to our knowledge, handle complex resonance forms automatically. OPLS parameters, although well optimized, are not available for a broad variety of molecules. The VC/2003 charges for water (0.353,−0.707) are somewhat less accurate than those of MMFF (0.430,−0.860) or OPLS (TIP3P 0.417,−0.834). It is thus worth commenting that, when careful tuned charges are available for a molecule, such as water, it is probably preferable to use the more specific charges than those obtained by a more general model.

Table 4. Root-Mean-Square Deviations of Electrostatic Potentials (kcal/(mol-electron)) Computed with Partial Atomic Charges Relative to ab Initio Electrostatic Potentials (D_i) for Various Charge Sets[a]

compound	CHELPG	VC/2003	MMFF	OPLS	QEq	GM
imidazole	2.08	3.80	4.30	4.03	9.89	9.84
imidazolium	1.43	5.13	4.22	6.59	8.68	4.70
methylamine	1.99	2.79	2.65	2.50	3.95	5.55
methylammonium	1.09	1.14	1.67	2.87	n/a	10.10
acetic acid	0.78	1.59	2.37	2.45	7.03	5.52
acetate	0.91	1.87	3.50	2.54	5.92	6.33
pyridine	1.83	2.38	4.80	4.79	3.88	3.93
aminobenzene	1.91	2.96	3.88	3.17	6.34	6.37
water	1.60	2.36	1.82	1.69	2.53	6.92
methanol	1.31	1.98	1.87	2.01	5.40	3.93
acetone	0.78	2.03	2.34	1.87	3.82	3.36
dimethyl ether	1.09	1.37	2.25	1.62	n/a	1.31
mean	**1.40**	**2.45**	**2.97**	**3.01**	**5.74**	**5.66**

[a] MMFF: Merck molecular force field;[11] OPLS: OPLS force field;[40] QEq: QEq electronegativity equalization method;[30] GM: Gasteiger−Marsili partial equalization of electronegativities method.[26] All charges except for CHELPG[17] and VC/2003 are drawn from a previous publication.[11]

Table 5. Root-Mean-Square Deviations of Electrostatic Potentials (kcal/(mol-electron)) Computed with Partial Atomic Charges Relative to ab Initio Electrostatic Potentials (D_i) for RESP[9,10] and AM1-BCC Charge Sets,[41] Compared with CHELPG[17] and VC/2003 Charges[a]

compound	CHELPG	VC/2003	RESP	AM1-BCC
imidazole	2.08	3.80	4.37	3.05
methanol	1.31	1.98	1.94	1.88
glucose	1.23	2.42	4.57	3.73
indole	2.16	2.88	2.82	3.68
aspirin	1.70	2.46	2.44	2.93
mean	**1.70**	**2.71**	**3.23**	**3.05**

[a] The RESP[9,10] and AM1-BCC charges are drawn from a previous publication.[12]

The QEq[30] and Gasteiger−Marsili[26] electronegativity equalization methods are appealing because they achieve broad applicability with only on the order of 50 parameters, some of which are not, in fact, considered to be adjustable terms. However, they yield relatively inaccurate electrostatic potentials, as assessed by the ab initio ESPs. Also, as previously pointed out,[11] QEq allows charge to spread from ionized groups more than seems physically reasonable. This was illustrated[11] with the zwitterion $CO_2^-(CH_2)_6NH_3^+$, for which QEq places a net charge of −0.092 on the NH_3^+ moiety. In contrast, the VC method constrains charges to remain on ionized groups (to within $\pm\,\delta$) so it assigns a net charge of 0.459 to this NH_3^+ group, a value close to that of 0.497 from the MMFF model.[11]

The AM1-BCC method of calculating partial atomic charges[12,41] uses semiempirical quantum mechanics to generate an initial set of charges and then corrects them by adding bond charge corrections designed to improve the agreement with reference charges from the RESP method,[9,10] which is based upon ab initio quantum mechanics. AM1-BCC achieves high accuracy at intermediate computational cost. Table 5 compares the accuracy of AM1-BCC charges as well as the RESP charges to which they are parametrized with that of VC/2003 and CHELPG charges. These results suggest that VC/2003 charges are as accurate as AM1-BCC, despite the fact that VC/2003 charges rely on fewer parameters than

AM1-BCC and can be computed much more rapidly. Surprisingly, the RESP charges do not appear to be especially accurate in the cases of imidazole and glucose. This issue is considered in the Discussion section.

3.1.3. Application to Amino Acids and Nucleotides. It is of interest to examine VC/2003 charges for standard biochemical groups, such as amino acids, for which existing force fields are relatively well parametrized. Table 6 shows the results for the 20 common amino acids, including several variant ionization states, and Table 7 shows the results for the four common deoxyribonucleotides. (In viewing these results, it should be kept in mind that the VC/2003 charges presented here differ from the force field charges in that they are extracted from capped model compounds and the net charge of the extracted atoms was not required to sum to an integer; the AMBER and CHARMM parameters, in contrast, require the charge of each separate residue to sum to −1, 0, or 1 electrons.) Overall, the VC/2003 charges are physically reasonable and appear very similar to those of the CHARMM22[6,7] and AMBER94[2] force fields for these biochemical components. Perhaps the most noticeable difference between VC/2003 and the other two charge sets is that VC/2003 makes the main-chain amide nitrogen more polar. However, the VC/2003 result is physically reasonable, especially given that the nitrogen charges assigned by VC/2003, about −0.65, are consistent with the charges assigned by all three charge sets to the chemically similar side-chain amide nitrogens of GLN and ASN: −0.9, −0.6, and −0.8 for AMBER94, CHARMM22, and VC/2003, respectively. Furthermore, the value of −0.65 is rather close to the charge of −0.55 for the AMBER ff02EP force field (http://www.amber.ucsf.edu/amber/dbase.html). Summary statistics shown in Figure 6 confirm that VC/2003 charges are highly similar to those of CHARMM and AMBER: the RMSD values between the charges from VC/2003, CHARMM, and AMBER94 are all within 0.10 ± 0.01 electrons for the amino acids and 0.138 ± 0.016 electrons for the nucleotides. It should be noted that CHARMM and AMBER charges are available for only a limited set of compounds, many of which are shown in Tables 6 and 7, whereas VC/2003 charges can be generated rapidly for a wide range of compounds.

3.2. Parameters. Table 1 shows the atomic electronegativities and hardnesses ($e°$ and $s°$, respectively) that result from the parametrization procedure described in Optimization of Parameters, and Table 8 lists the optimized values of the global parameters. The average electronegativity of each element—for example, the average for the four types of oxygen in Table 1—correlates well with its Pauling electronegativity, as shown in Figure 7.

3.3. Resonance Forms. The present method of generating resonance forms efficiently treats even complex resonance systems that might be difficult to analyze by hand. Thus, it correctly identifies both resonance forms of the compound in Figure 2 and of other vinylogous compounds as well as the more complicated system shown in Figure 3. Note that resonance form **5** in Figure 3 cannot be derived by a single formal electron transfer starting from the initial conformation **1** but only via an electron transfer starting from one of the second generation of resonance forms. Moreover, resonance form **5** can be reached only via an intermediate resonance form, **2**, which is of high energy relative to the other forms. These examples show that identification of alternative donor/acceptor resonance forms is a nontrivial task that is best accomplished by the use of a systematic algorithm.

Correctly averaging over resonance forms is essential in order to equalize the charges of chemically equivalent atoms, such as the oxygens in a carboxylate group or the widely separated nitrogens in the vinylogous system in Figure 2. Table 9 compares the atomic charges of this compound from VC/2003 with those from Gasteiger−Marsili[26] as implemented in the program Babel[42] and the program Quanta.[18,19] Neither of these two methods detects alternate resonance forms, so both yield markedly incorrect charges for this compound. Atom numbering is provided in Figure 8.

3.4. Efficiency and Applicability. The VC charging method is suitable for processing large numbers of compounds because it is fast and broadly applicable. Thus the current implemention averages about 0.45 s per compound in the December 2002 Maybridge HTS compound catalog (Maybridge PLC) on a 2.26 GHz Pentium 4 processor, and it is able to process all but 8 of the approximately 56 000 compounds. The code that detects aromaticity, resonance forms, and charge groups is written in the Python scripting language, and considerable overall speed-up is available, if needed, by conversion of time-consuming segments of Python code to a compiled language such as C.

4. DISCUSSION

The present charging method provides accurate charges, relative to electrostatic potentials from ab initio quantum calculations, along with computational speed, and broad applicability. It is anticipated that this approach will be useful in a range of applications, including ligand−receptor docking and scoring, processing of virtual compound libraries, chemical design, conformational analysis, free energy calculations, and the calculation of molecular descriptors for use in quantitative structure−activity and structure−property relationships (QSAR and QSPR). In particular, because the charges are optimized to fit ab initio potentials at the 6-31G* level, are independent of conformation, and assign equal charges to chemically equivalent atoms, they should be compatible with existing force fields for molecular simulations, such as AMBER and CHARMM. Indeed, VC/2003 charges for common biochemical components are quite similar to those from the CHARMM and AMBER force fields. Thus, it would be appropriate to model a protein−ligand system using, for example, AMBER parameters for a protein, TIP4P parameters for any explicit water molecules, and VC/2003 charges for a variety of drug candidates. It would also be reasonable to compute charges for a protein with a covalently modified amino acid by assigning the protein backbone the usual force field charges and using VC/2003 to compute the side-chain charges, constraining the net charge of the side-chain atoms to sum to the value required by the force field.

Other electronegativity equalization methods, such as Gasteiger−Marsili and QEq, also provide speed and broad applicability but are not particularly accurate relative to reference ab initio calculations. In addition, QEq charges, unlike Gasteiger−Marsili and the present model, depend on the conformation of the molecule. Although it is reasonable that atomic charges should vary somewhat with conformation, it has not, to our knowledge, been shown that the

Table 6. Comparison of VC/2003 Charges for Amino Acid Residues with Those from the CHARMM22 and AMBER94 Force Fields

atom	AMBER94	CHARMM22	VC/2003	atom	AMBER94	CHARMM22	VC/2003	atom	AMBER94	CHARMM22	VC/2003
					Alanine						
N	−0.4157	−0.47	−0.655	CB	−0.1825	−0.27	−0.125	HB3	0.0603	0.09	0.049
H	0.2719	0.31	0.333	HB1	0.0603	0.09	0.049	C	0.5973	0.51	0.628
CA	0.0337	0.07	0.179	HB2	0.0603	0.09	0.049	O	−0.5679	−0.51	−0.549
HA	0.0823	0.09	0.026								
					Arginine						
N	−0.3479	−0.47	−0.639	HG2	0.0285	0.09	0.067	NH1	−0.8627	−0.8	−0.705
H	0.2747	0.31	0.352	HG3	0.0285	0.09	0.067	HH11	0.4478	0.46	0.358
CA	−0.2637	0.07	0.195	CD	0.0486	0.2	0.177	HH12	0.4478	0.46	0.358
HA	0.156	0.09	0.045	HD2	0.0687	0.09	0.053	NH2	−0.8627	−0.8	−0.705
CB	−0.0007	−0.18	−0.068	HD3	0.0687	0.09	0.053	HH21	0.4478	0.46	0.358
HB2	0.0327	0.09	0.067	NE	−0.5295	−0.7	−0.578	HH22	0.4478	0.46	0.358
HB3	0.0327	0.09	0.067	HE	0.3456	0.44	0.356	C	0.7341	0.51	0.646
CG	0.039	−0.18	−0.062	CZ	0.8076	0.64	0.478	O	−0.5894	−0.51	−0.533
					Asparagine						
N	−0.4157	−0.47	−0.654	HB2	0.0797	0.09	0.046	HD21	0.4196	0.32	0.336
H	0.2719	0.31	0.334	HB3	0.0797	0.09	0.046	HD22	0.4196	0.3	0.336
CA	0.0143	0.07	0.176	CG	0.713	0.55	0.646	C	0.5973	0.51	0.629
HA	0.1048	0.09	0.028	OD1	−0.5931	−0.55	−0.548	O	−0.5679	−0.51	−0.548
CB	−0.2041	−0.18	−0.096	ND2	−0.9191	−0.62	−0.763				
					Aspartic Acid (Anionic)						
N	−0.5163	−0.47	−0.666	CB	−0.0303	−0.28	−0.117	OD1	−0.8014	−0.76	−0.721
H	0.2936	0.31	0.319	HB2	−0.0122	0.09	0.032	OD2	−0.8014	−0.76	−0.721
CA	0.0381	0.07	0.163	HB3	−0.0122	0.09	0.032	C	0.5366	0.51	0.615
HA	0.088	0.09	0.013	CG	0.7994	0.62	0.751	O	−0.5819	−0.51	−0.560
					Cysteine						
N	−0.4157	−0.47	−0.652	CB	−0.1231	−0.11	0.000	HG	0.1933	0.16	0.177
H	0.2719	0.31	0.336	HB2	0.1112	0.09	0.044	C	0.5973	0.51	0.631
CA	0.0213	0.07	0.175	HB3	0.1112	0.09	0.044	O	−0.5679	−0.51	−0.546
HA	0.1124	0.09	0.030	SG	−0.3119	−0.23	−0.295				
					Glutamine						
N	−0.4157	−0.47	−0.654	HB3	0.0171	0.09	0.048	NE2	−0.9407	−0.62	−0.764
H	0.2719	0.31	0.333	CG	−0.0645	−0.18	−0.081	HE21	0.4251	0.32	0.335
CA	−0.0031	0.07	0.177	HG2	0.0352	0.09	0.045	HE22	0.4251	0.3	0.335
HA	0.085	0.09	0.027	HG3	0.0352	0.09	0.045	C	0.5973	0.51	0.628
CB	−0.0036	−0.18	−0.087	CD	0.6951	0.55	0.645	O	−0.5679	−0.51	−0.549
HB2	0.0171	0.09	0.048	OE1	−0.6086	−0.55	−0.549				
					Glutamic Acid (Anionic)						
N	−0.5163	−0.47	−0.665	HB2	−0.0173	0.09	0.035	CD	0.8054	0.62	0.751
H	0.2936	0.31	0.320	HB3	−0.0173	0.09	0.035	OE1	−0.8188	−0.76	−0.721
CA	0.0397	0.07	0.165	CG	0.0136	−0.28	−0.100	OE2	−0.8188	−0.76	−0.721
HA	0.1105	0.09	0.014	HG2	−0.0425	0.09	0.032	C	0.5366	0.51	0.615
CB	0.056	−0.18	−0.099	HG3	−0.0425	0.09	0.032	O	−0.5819	−0.51	−0.560
					Glycine						
N	−0.4157	−0.47	−0.649	HA2	0.0698	0.09	0.030	C	0.5973	0.51	0.632
H	0.2719	0.31	0.334	HA3	0.0698	0.09	0.030	O	−0.5679	−0.51	−0.548
CA	−0.0252	−0.02	0.139								
					Histidine (Neutral HD1)						
N	−0.4157	−0.47	−0.657	HB3	0.0402	0.09	0.041	NE2	−0.5727	−0.7	−0.516
H	0.2719	0.31	0.330	CG	−0.0266	−0.05	0.037	CD2	0.1292	0.22	0.157
CA	0.0188	0.07	0.172	ND1	−0.3811	−0.36	−0.328	HD2	0.1147	0.1	0.075
HA	0.0881	0.09	0.023	HD1	0.3649	0.32	0.308	C	0.5973	0.51	0.625
CB	−0.0462	−0.09	−0.067	CE1	0.2057	0.25	0.268	O	−0.5679	−0.51	−0.552
HB2	0.0402	0.09	0.041	HE1	0.1392	0.13	0.063				
					Histidine (Neutral HE1)						
N	−0.4157	−0.47	−0.657	HB3	0.0367	0.09	0.041	HE2	0.3339	0.32	0.308
H	0.2719	0.31	0.330	CG	0.1868	0.22	0.225	CD2	−0.2207	−0.05	−0.032
CA	−0.0581	0.07	0.172	ND1	−0.5432	−0.7	−0.520	HD2	0.1862	0.09	0.091
HA	0.136	0.09	0.023	CE1	0.1635	0.25	0.268	C	0.5973	0.51	0.625
CB	−0.0074	−0.08	−0.081	HE1	0.1435	0.13	0.063	O	−0.5679	−0.51	−0.552
HB2	0.0367	0.09	0.041	NE2	−0.2795	−0.36	−0.326				
					Histidine (Cationic)						
N	−0.3479	−0.47	−0.641	HB3	0.081	0.09	0.060	NE2	−0.1718	−0.51	−0.326
H	0.2747	0.31	0.349	CG	−0.0012	0.19	0.134	HE2	0.3911	0.44	0.384
CA	−0.1354	0.07	0.190	ND1	−0.1513	−0.51	−0.331	CE1	−0.017	0.32	0.066
HA	0.1212	0.09	0.043	HD1	0.3866	0.44	0.384	HE1	0.2681	0.18	0.113
CB	−0.0414	−0.05	−0.046	CD2	−0.1141	0.19	0.143	C	0.7341	0.51	0.643
HB2	0.081	0.09	0.060	HD2	0.2317	0.13	0.105	O	−0.5894	−0.51	−0.535

Table 6 (Continued)

atom	AMBER94	CHARMM22	VC/2003	atom	AMBER94	CHARMM22	VC/2003	atom	AMBER94	CHARMM22	VC/2003
					Isoleucine						
N	−0.4157	−0.47	−0.657	HG21	0.0882	0.09	0.047	CD1	−0.066	−0.27	−0.111
H	0.2719	0.31	0.330	HG22	0.0882	0.09	0.047	HD11	0.0186	0.09	0.047
CA	−0.0597	0.07	0.172	HG23	0.0882	0.09	0.047	HD12	0.0186	0.09	0.047
HA	0.0869	0.09	0.024	CG1	−0.043	−0.18	−0.072	HD13	0.0186	0.09	0.047
CB	0.1303	−0.09	−0.047	HG12	0.0236	0.09	0.045	C	0.5973	0.51	0.625
HB	0.0187	0.09	0.043	HG13	0.0236	0.09	0.045	O	−0.5679	−0.51	−0.551
CG2	−0.3204	−0.27	−0.113								
					Leucine						
N	−0.4157	−0.47	−0.657	CG	0.3531	−0.09	−0.029	CD2	−0.4121	−0.27	−0.113
H	0.2719	0.31	0.330	HG	−0.0361	0.09	0.043	HD21	0.1	0.09	0.047
CA	−0.0518	0.07	0.174	CD1	−0.4121	−0.27	−0.113	HD22	0.1	0.09	0.047
HA	0.0922	0.09	0.024	HD11	0.1	0.09	0.047	HD23	0.1	0.09	0.047
CB	−0.1102	−0.18	−0.090	HD12	0.1	0.09	0.047	C	0.5973	0.51	0.625
HB2	0.0457	0.09	0.045	HD13	0.1	0.09	0.047	O	−0.5679	−0.51	−0.551
HB3	0.0457	0.09	0.045								
					Lysine (Cation)						
N	−0.3479	−0.47	−0.647	HG2	0.0103	0.09	0.057	HE3	0.1135	0.05	0.042
H	0.2747	0.31	0.342	HG3	0.0103	0.09	0.057	NZ	−0.3854	−0.3	−0.439
CA	−0.24	0.07	0.186	CD	−0.0479	−0.18	−0.072	HZ1	0.34	0.33	0.315
HA	0.1426	0.09	0.036	HD2	0.0621	0.09	0.057	HZ2	0.34	0.33	0.315
CB	−0.0094	−0.18	−0.077	HD3	0.0621	0.09	0.057	HZ3	0.34	0.33	0.315
HB2	0.0362	0.09	0.057	CE	−0.0143	0.21	0.136	C	0.7341	0.51	0.637
HB3	0.0362	0.09	0.057	HE2	0.1135	0.05	0.042	O	−0.5894	−0.51	−0.541
CG	0.0187	−0.18	−0.061								
					Methionine						
N	−0.4157	−0.47	−0.654	HB3	0.0241	0.09	0.048	HE1	0.0684	0.09	0.043
H	0.2719	0.31	0.333	CG	0.0018	−0.14	0.012	HE2	0.0684	0.09	0.043
CA	−0.0237	0.07	0.177	HG2	0.044	0.09	0.041	HE3	0.0684	0.09	0.043
HA	0.088	0.09	0.027	HG3	0.044	0.09	0.041	C	0.5973	0.51	0.628
CB	0.0342	−0.18	−0.090	SD	−0.2737	−0.09	−0.193	O	−0.5679	−0.51	−0.548
HB2	0.0241	0.09	0.048	CE	−0.0536	−0.22	−0.027				
					Phenylalanine						
N	−0.4157	−0.47	−0.653	CG	0.0118	0	−0.042	CD2	−0.1256	−0.11	−0.110
H	0.2719	0.31	0.335	CD1	−0.1256	−0.11	−0.110	HD2	0.133	0.11	0.107
CA	−0.0024	0.07	0.176	HD1	0.133	0.11	0.107	CE2	−0.1704	−0.11	−0.107
HA	0.0978	0.09	0.028	CE1	−0.1704	−0.11	−0.107	HE2	0.143	0.11	0.107
CB	−0.0343	−0.18	−0.053	HE1	0.143	0.11	0.107	C	0.5973	0.51	0.632
HB2	0.0295	0.09	0.046	CZ	−0.1072	−0.11	−0.107	O	−0.5679	−0.51	−0.547
HB3	0.0295	0.09	0.046	HZ	0.1297	0.11	0.107				
					Proline						
N	−0.2548	−0.29	−0.537	HA	0.0641	0.09	0.027	HG2	0.0213	0.09	0.048
CD	0.0192	0	0.148	CB	−0.007	−0.18	−0.085	HG3	0.0213	0.09	0.048
HD2	0.0391	0.09	0.032	HB2	0.0253	0.09	0.048	C	0.5896	0.51	0.628
HD3	0.0391	0.09	0.032	HB3	0.0253	0.09	0.048	O	−0.5748	−0.51	−0.549
CA	−0.0266	0.02	0.175	CG	0.0189	−0.18	−0.082				
					Serine						
N	−0.4157	−0.47	−0.655	CB	0.2117	0.05	0.175	HG	0.4275	0.43	0.384
H	0.2719	0.31	0.332	HB2	0.0352	0.09	0.028	C	0.5973	0.51	0.627
CA	−0.0249	0.07	0.161	HB3	0.0352	0.09	0.028	O	−0.5679	−0.51	−0.550
HA	0.0843	0.09	0.025	OG	−0.6546	−0.66	−0.561				
					Threonine						
N	−0.4157	−0.47	−0.656	HB	0.0043	0.09	0.024	HG22	0.0642	0.09	0.048
H	0.2719	0.31	0.331	OG1	−0.6761	−0.66	−0.566	HG23	0.0642	0.09	0.048
CA	−0.0389	0.07	0.158	HG1	0.4102	0.43	0.383	C	0.5973	0.51	0.626
HA	0.1007	0.09	0.024	CG2	−0.2438	−0.27	−0.127	O	−0.5679	−0.51	−0.551
CB	0.3654	0.14	0.215	HG21	0.0642	0.09	0.048				
					Tryptophane						
N	−0.4157	−0.47	−0.656	CD1	−0.1638	0.03	−0.011	CZ3	−0.1972	−0.11	−0.110
H	0.2719	0.31	0.331	HD1	0.2062	0.11	0.092	HZ3	0.1447	0.11	0.104
CA	−0.0275	0.07	0.173	NE1	−0.3418	−0.61	−0.321	CZ2	−0.2601	−0.11	−0.122
HA	0.1123	0.09	0.025	HE1	0.3412	0.38	0.310	HZ2	0.1572	0.11	0.104
CB	−0.005	−0.18	−0.056	CE2	0.138	0.13	0.081	CH2	−0.1134	−0.11	−0.110
HB2	0.0339	0.09	0.042	CD2	0.1243	−0.02	−0.030	HH2	0.1417	0.11	0.104
HB3	0.0339	0.09	0.042	CE3	−0.2387	−0.11	−0.115	C	0.5973	0.51	0.626
CG	−0.1415	−0.03	−0.057	HE3	0.17	0.11	0.104	O	−0.5679	−0.51	−0.550
					Tyrosine						
N	−0.4157	−0.47	−0.653	CG	−0.0011	0	−0.042	HH	0.3992	0.43	0.383
H	0.2719	0.31	0.335	CD1	−0.1906	−0.11	−0.110	CD2	−0.1906	−0.11	−0.110
CA	−0.0014	0.07	0.177	HD1	0.1699	0.11	0.108	HD2	0.1699	0.11	0.108
HA	0.0876	0.09	0.029	CE1	−0.2341	−0.11	−0.121	CE2	−0.2341	−0.11	−0.121
CB	−0.0152	−0.18	−0.053	HE1	0.1656	0.11	0.108	HE2	0.1656	0.11	0.108
HB2	0.0295	0.09	0.046	CZ	0.3226	0.11	0.151	C	0.5973	0.51	0.630
HB3	0.0295	0.09	0.046	OH	−0.5579	−0.54	−0.514	O	−0.5679	−0.51	−0.547

Table 6 (Continued)

atom	AMBER94	CHARMM22	VC/2003	atom	AMBER94	CHARMM22	VC/2003	atom	AMBER94	CHARMM22	VC/2003
					Valine						
N	−0.4157	−0.47	−0.656	CG1	−0.3192	−0.27	−0.112	HG21	0.0791	0.09	0.048
H	0.2719	0.31	0.331	HG11	0.0791	0.09	0.048	HG22	0.0791	0.09	0.048
CA	−0.0875	0.07	0.173	HG12	0.0791	0.09	0.048	HG23	0.0791	0.09	0.048
HA	0.0969	0.09	0.024	HG13	0.0791	0.09	0.048	C	0.5973	0.51	0.626
CB	0.2985	−0.09	−0.045	CG2	−0.3192	−0.27	−0.112	O	−0.5679	−0.51	−0.551
HB	−0.0297	0.09	0.043								
					Lysine (Neutral)						
N	−0.4157		−0.658	CG	0.0661		−0.073	HE2	−0.0335		0.027
H	0.2719		0.329	HG2	0.0104		0.043	HE3	−0.0335		0.027
CA	−0.072		0.173	HG3	0.0104		0.043	NZ	−1.0358		−0.789
HA	0.0994		0.022	CD	−0.0376		−0.086	HZ2	0.386		0.333
CB	−0.0484		−0.090	HD2	0.0115		0.043	HZ3	0.386		0.333
HB2	0.034		0.043	HD3	0.0115		0.043	C	0.5973		0.624
HB3	0.034		0.043	CE	0.326		0.149	O	−0.5679		−0.552
					Aspartic Acid (Neutral)						
N	−0.4157	−0.47	−0.652	HB2	0.0488	0.09	0.048	OD2	−0.6376	−0.61	−0.538
H	0.2719	0.31	0.335	HB3	0.0488	0.09	0.048	HD2	0.4747	0.44	0.385
CA	0.0341	0.07	0.178	CG	0.6462	0.75	0.681	C	0.5973	0.51	0.630
HA	0.0864	0.09	0.029	OD1	−0.5554	−0.55	−0.548	O	−0.5679	−0.51	−0.547
CB	−0.0316	−0.21	−0.097								
					Glutamic Acid (Neutral)						
N	−0.4157	−0.47	−0.653	HB3	0.0256	0.09	0.049	OE1	−0.5838	−0.55	−0.549
H	0.2719	0.31	0.334	CG	−0.0174	−0.21	−0.082	OE2	−0.6511	−0.61	−0.539
CA	0.0145	0.07	0.178	HG2	0.043	0.09	0.046	HE2	0.4641	0.44	0.384
HA	0.0779	0.09	0.028	HG3	0.043	0.09	0.046	C	0.5973	0.51	0.629
CB	−0.0071	−0.18	−0.086	CD	0.6801	0.75	0.680	O	−0.5679	−0.51	−0.548
HB2	0.0256	0.09	0.049								

variations predicted by QEq are accurate. RESP charges, although considered to be highly reliable, are time-consuming to compute, especially for complex compounds, because they involve quantum calculations at the 6-31G* level.

The AM1-BCC method yields results very close to RESP at about 100 times less computational cost. However, the semiempirical quantum calculations used in AM1-BCC, which include a conformational optimization step, can require a few minutes for a drug-like compound on a fast PC, making the method on the order of 100 times slower than the method presented here. It is also worth noting that AM1-BCC charges depend to some degree upon the molecular conformation, while conformational independence is preferred in most molecular modeling applications. In addition, AM1-BCC does not, to our knowledge, automatically assign equal charges to chemically equivalent atoms. Rather, information about chemical equivalence must be supplied manually by the user. The VC approach generates charges that are independent of conformation and are equal for chemically equivalent atoms. The AM1-BCC method also poses a greater parametrization challenge than the VC method because it has over 3 times as many adjustable parameters.

It was initially surprising that the potentials generated by RESP charges deviate so strongly from the ab initio potentials for imidazole (RMSD 4.37 (kcal/mol-electron)) and glucose (RMSD 4.57) (Table 5), given that the optimally fitted CHELPG charges give RMSDs of only 1.2 and 2.1, respectively, while VC/2003 charges give RMSDs between the CHELPG and RESP values. In the case of glucose, this difference could result in part from a difference between the conformations used to generate the RESP charges[12] and the conformation for which electrostatic potentials were compared here (Section 2.2). However, conformational differences could not explain the difference for imidazole since this compound is rigid, and the same protonation state is used here as before.[12] It seems likely that the deviation of potentials computed with RESP charges results from the fact that RESP sacrifices accuracy in nonpolar parts of molecules in order to maximize accuracy in polar regions and thereby to correctly capture strong polar interactions.[9,10] Thus, for trans-*N*-methylacetamide, the RESP fitting method approximately doubles the RMS error relative to mathematically optimal charges, much as seen here. (See Table 10 of ref 10.) It is thus worth asking whether VC/2003 charges should be modified by making a similar tradeoff between the accuracy of potentials around polar and nonpolar parts of molecules; various methods of making this change can readily be envisioned. However, focusing on the accuracy of polar regions may, in some cases, impose nonphysical charges on nonpolar atoms and published RESP charges for amino acids hint that this could be a significant concern. For example, RESP charges for nominally nonpolar carbons in the leucine side-chain are as large as 0.353 and −0.412 (Table 6). In addition, the carbon alpha charges of arginine and lysine, at about −0.25, are considerably larger than those of the other amino acids, about ± 0.025 (Table 6); this difference presumably has to do with the effect of the side-chain cations on the fitting process.

The method of identifying alternate resonance forms is a central innovation of the present method. To our knowledge, this is the first general algorithm for systematically addressing this problem. Its development was motivated initially by the need for an automated way of ensuring that all chemically equivalent atoms are assigned equal charges, but the algorithm has other uses as well. One application is in the detection of chemical symmetries; for example, the symmetry of the vinylogous compound in Figure 2 can be detected only if both resonance forms are considered. Resonance

Table 7. Comparison of VC/2003 Charges for Deoxyribonucleic Acid Residues with Those from the CHARMM27 and AMBER94 Force Fields

atom	AMBER94	CHARMM27	VC/2003	atom	AMBER94	CHARMM27	VC/2003	atom	AMBER94	CHARMM27	VC/2003
						A					
P	1.1659	1.5	1.105	H1′	0.1838	0.09	0.005	C2	0.5716	0.5	0.441
O1P	−0.7761	−0.78	−0.780	N9	−0.0268	−0.05	−0.279	H2	0.0598	0.13	0.043
O2P	−0.7761	−0.78	−0.780	C8	0.1607	0.34	0.262	N3	−0.7417	−0.75	−0.543
O5′	−0.4954	−0.57	−0.452	H8	0.1877	0.12	0.060	C4	0.38	0.43	0.333
C5′	−0.0069	−0.08	0.166	N7	−0.6175	−0.71	−0.527	C3′	0.0713	0.01	0.205
H5′	0.0754	0.09	0.022	C5	0.0725	0.28	0.196	H3′	0.0985	0.09	0.020
H5″	0.0754	0.09	0.022	C6	0.6897	0.46	0.369	C2′	−0.0854	−0.18	−0.121
C4′	0.1629	0.16	0.191	N6	−0.9123	−0.77	−0.759	H2′	0.0718	0.09	0.041
H4′	0.1176	0.09	0.020	H61	0.4167	0.38	0.328	H2″	0.0718	0.09	0.041
O4′	−0.3691	−0.5	−0.465	H62	0.4167	0.38	0.328	O3′	−0.5232	−0.57	−0.456
C1′	0.0431	0.16	0.422	N1	−0.7624	−0.74	−0.543				
						T					
P	1.1659	1.5	1.105	H1′	0.1804	0.09	0.007	N3	−0.434	−0.46	−0.643
O1P	−0.7761	−0.78	−0.780	N1	−0.0239	−0.34	−0.523	H3	0.342	0.36	0.327
O2P	−0.7761	−0.78	−0.780	C6	−0.2209	0.17	0.058	C2	0.5677	0.51	0.819
O5′	−0.4954	−0.57	−0.449	H6	0.2607	0.17	0.072	O2	−0.5881	−0.41	−0.566
C5′	−0.0069	−0.08	−0.169	C5	0.0025	−0.15	−0.073	C3′	0.0713	0.01	0.208
H5′	0.0754	0.09	0.026	C7	−0.2269	−0.11	−0.092	H3′	0.0985	0.09	0.024
H5″	0.0754	0.09	0.026	H71	0.077	0.07	0.044	C2′	−0.0854	−0.18	−0.119
C4′	0.1629	0.16	0.194	H72	0.077	0.07	0.044	H2′	0.0718	0.09	0.045
H4′	0.1176	0.09	0.024	H73	0.077	0.07	0.044	H2″	0.0718	0.09	0.045
O4′	−0.3691	−0.5	−0.462	C4	0.5194	0.5	0.655	O3′	−0.5232	−0.57	−0.454
C1′	0.068	0.16	0.441	O4	−0.5563	−0.45	−0.555				
						G					
P	1.1659	1.5	1.105	H1′	0.1746	0.09	0.008	N2	−0.923	−0.68	−0.775
O1P	−0.7761	−0.78	−0.780	N9	0.0577	−0.02	−0.272	H21	0.4235	0.32	0.322
O2P	−0.7761	−0.78	−0.780	C8	0.0736	0.25	0.265	H22	0.4235	0.35	0.322
O5′	−0.4954	−0.57	−0.450	H8	0.1997	0.16	0.063	N3	−0.6636	−0.74	−0.543
C5′	−0.0069	−0.08	0.169	N7	−0.5725	−0.6	−0.524	C4	0.1814	0.26	0.183
H5′	0.0754	0.09	0.025	C5	0.1991	0	0.210	C3′	0.0713	0.01	0.208
H5″	0.0754	0.09	0.025	C6	0.4918	0.54	0.637	H3′	0.0985	0.09	0.023
C4′	0.1629	0.16	0.194	O6	−0.5699	−0.51	−0.556	C2′	−0.0854	−0.18	−0.118
H4′	0.1176	0.09	0.023	N1	−0.5053	−0.34	−0.641	H2′	0.0718	0.09	0.045
O4′	−0.3691	−0.5	−0.462	H1	0.352	0.26	0.327	H2″	0.0718	0.09	0.045
C1′	0.0358	0.16	0.425	C2	0.7432	0.75	0.599	O3′	−0.5232	−0.57	−0.454
						C					
P	1.1659	1.5	1.105	C1′	−0.0116	0.16	0.442	H42	0.4314	0.33	0.333
O1P	−0.7761	−0.78	−0.780	H1′	0.1963	0.09	0.008	N3	−0.7748	−0.66	−0.567
O2P	−0.7761	−0.78	−0.780	N1	−0.0339	−0.13	−0.522	C2	0.7959	0.52	0.816
O5′	−0.4954	−0.57	−0.448	C6	−0.0183	0.05	0.062	O2	−0.6548	−0.49	−0.565
C5′	−0.0069	−0.08	0.170	H6	0.2293	0.17	0.074	C3′	0.0713	0.01	0.209
H5′	0.0754	0.09	0.027	C5	−0.5222	−0.13	−0.132	H3′	0.0985	0.09	0.025
H5″	0.0754	0.09	0.027	H5	0.1863	0.07	0.087	C2′	−0.0854	−0.18	−0.118
C4′	0.1629	0.16	0.195	C4	0.8439	0.65	0.436	H2′	0.0718	0.09	0.046
H4′	0.1176	0.09	0.025	N4	−0.9773	−0.75	−0.766	H2″	0.0718	0.09	0.046
O4′	−0.3691	−0.5	−0.461	H41	0.4314	0.37	0.333	O3′	−0.5232	−0.57	−0.453

detection should also be useful in assigning atom types so that appropriate force field parameters may be assigned to atoms in molecules. Thus, although existing methods of assigning atom types can recognize a carboxylate group as such and hence assign equivalent types to the two carboxylate oxygens, many existing atom-typing algorithms will not correctly type atoms in vinylogous molecules, because the resonance donor and acceptor atoms are far apart. In the present application, only the most highly occupied resonance forms are considered (Section 2.1.2); i.e., those with lowest resonance energy. However, for atom-typing, it will be necessary to consider higher-energy forms as well. For example, it should be possible to recognize not only a standard amide group as such but also a vinylogous amide group, by considering the contribution of the charge-separated resonance form to the overall electronic structure of the compound. Finally, there may well be further applications of this algorithm such as in the prediction of ionization constants (pK_as) and in the context of chemical education.

It is worth noting that numerous variants of the VC method presented here can readily be envisioned. At the simplest level, the numerical parameters could be further adjusted to tune the model for specific classes of compounds or to adjust the model to mimic existing force field charges (e.g., AMBER94) or charges from other methods (e.g., RESP). Also, the criteria used to categorize atoms by type (Table 1), and thus the number of atom types, could be modified. Alternative methods of adjusting the electronegativities of atoms depending upon their location in a molecule (eq 1) could also be used, and it may be worth considering more sophisticated methods for merging the contributions of different resonance forms. For example, the accuracy of the charges might be increased by including small contributions from low-occupancy resonance forms, based upon their resonance energies (Section 2.1.2). It would also be possible

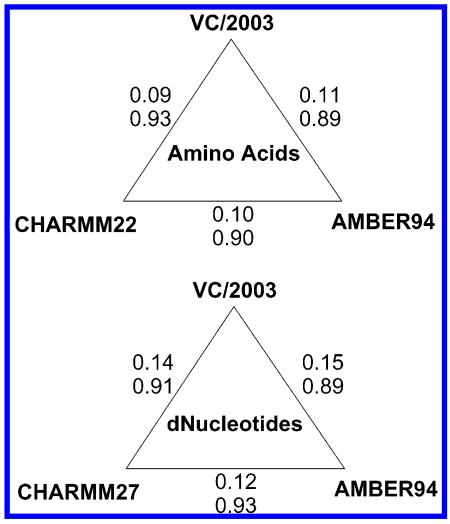

Figure 6. Comparison of VC/2003 charges with AMBER94 and CHARMM charges (CHARMM22 for amino acids, CHARMM27 charges for nucleotides). All charges are listed in Tables 6 and 7. Top number: root-mean-square deviation of atomic charges; bottom: correlation coefficient.

Table 8. Optimized Values of the Global Parameters[a]

param	value	param	value	param	value
α_1	1.00	α_4	0.86	β	1.378
α_2	1.74	α_5	0.0570	δ	0.545
α_3	1.67				

[a] The units of δ are electron charges.

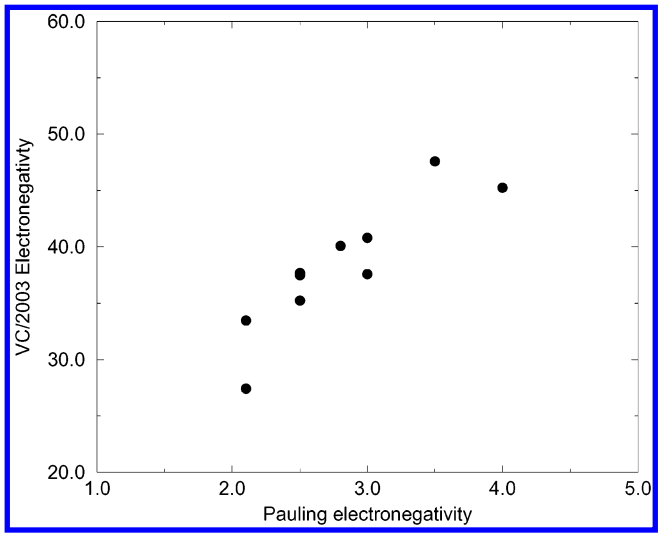

Figure 7. Comparison of VC/2003 electronegativities averaged by element (arbitrary units) with Pauling electronegativities.

to account for alternate resonance forms by computing charges separately for each form and then averaging the charge. However, this procedure would slow the calculations somewhat relative to the present approach, in which only one charge calculation is done even if multiple resonance forms are present.

5. CONCLUSIONS

The present method, VC/2003, efficiently provides partial atomic charges that agree well with ab initio quantum calculations and that are independent of molecular conformation. Its accuracy results in part from its ability to automatically detect and account for alternate resonance forms, a capability that is, to our knowledge, unique.

Table 9. Atomic Partial Charges for the Vinylogous Compound in Figures 2 and 8[a]

atom(s)	VC/2003	GM	Quanta
1	0.367	0.155	0.15
1′	0.367	0.352	0.25
2	0.367	0.155	0.15
2′	0.367	0.352	0.25
3	−0.641	−0.404	−0.30
3′	−0.641	0.010	−0.40
4	0.135	−0.005	−0.20
4′	0.135	0.154	0.55
5	0.133	0.080	0.17
5′	0.133	0.097	0.17
6	−0.052	−0.046	0.00
6′	−0.052	−0.031	−0.20
7	0.143	0.063	0.12
7′	0.143	0.064	0.17
8	−0.046	−0.059	0.00
9	0.143	0.062	0.12

[a] VC/2003: VC charge model with current parameters; GM: Gasteiger−Marsili charge model[26] from Babel;[42] Quanta: charges from the charge template method in the program QUANTA, with excess charge distributed over carbons and nonpolar hydrogens.[18,19]

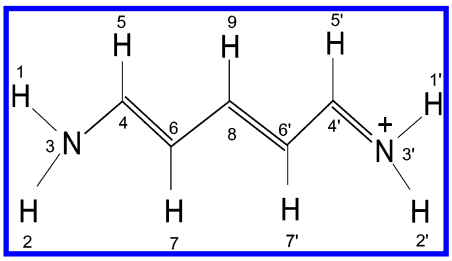

Figure 8. Numbering scheme used in Table 9 for compound in Figure 2.

REFERENCES AND NOTES

(1) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765−784.

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

(3) van Gunsteren, W. F.; Berendsen, H. J. C., *Groningen Molecular Simulation (GROMOS) Library Manual*; Biomos: Nijenborgh 16, 9747 AG Groningen, The Netherlands, 1987.

(4) Scott, W.; Hunenberger, P.; Tironi, I.; Mark, A.; Billeter, S.; Fennen, J.; Torda, J.; Huber, T.; Kruger, P.; vanGunsteren, W. The GROMOS biomolecular simulation package. *J. Phys. Chem. A* **1999**, *103*, 3596−3607.

(5) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.

(6) Molecular Simulations Inc. Waltham, MA. CHARMm Version 22., Polar hydrogen parameter set for CHARMm Version 22. Waltham, MA. 1992.

(7) MacKerell, Jr., A. D.; Wiókiewicz-Kuczera, J.; Karplus, M. An all-atom empirical energy function for the simulation of nucleic acids. *J. Am. Chem. Soc.* **1995**, *117*, 11946−11975.

(8) MacKerell, A.; Bashford, D.; Bellott, M.; Dunbrack, R.; Evanseck, J.; Field, M.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D.; Prodhom, B.; Reiher, W.; Roux, B.; Schlenkrich, M.; Smith, J.; 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. B* **1998**, *102*, 3586−3616.

(9) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. Application of RESP charges to calculation conformational energies, hydrogen bond energies, and free energies of solvation. *J. Am. Chem. Soc.* **1993**, *115*, 9620−9631.
(10) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A well-behaved electrostatic potential based method using charge-restraints for deriving charges: The RESP model. *J. Phys. Chem.* **1993**, *97*, 10269−10280.
(11) Halgren, T. A. Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *J. Comput. Chem.* **1996**, *17*, 520−552.
(12) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I., Fast efficient generation of high-quality atomic charges. AM1-BCC Model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132−146.
(13) Momany, F., Determination of Partial Atomic Charges from Ab Initio M. E. P. Applications to Formamide, Methanol and Formic Acid. *J. Phys. Chem.* **1978**, *82*, 592.
(14) 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.
(15) Singh, U. C.; Kollman, P. A. An Approach to Computing Electrostatic Charges for Molecules. *J. Comput. Chem.* **1984**, *5*, 129−145.
(16) Chirlian, L.; Francl, M. Charges Fit to Electrostatic Potentials. *J. Comput. Chem.* **1987**, *8*, 894.
(17) 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.
(18) Momany, F. A.; Rone, R. Validation of the general-purpose Quanta-(r)3.2/CHARMM(r) force-field. *J. Comput. Chem.* **1992**, *13*, 888−900.
(19) Molecular Simulations Inc. Waltham, MA, Quanta2000.
(20) Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate Calculation of Hydration Free Energies Using Macroscopic Solvation Models. *J. Phys. Chem.* **1994**, *98*, 1978−1988.
(21) Bush, B. L.; Bayly, C. I.; Halgren, T. A. Consensus bond-charge increments fitted to electrostatic potential or field of many compounds: Application to MMFF94 training set. *J. Comput. Chem.* **1999**.
(22) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.
(23) 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.
(24) Li, J.; Zhu, T.; Cramer, C.; Truhlar, D. G., A New Class IV Charge Model for Extracting Accurate Partial Charges from Wave Functions. *J. Phys. Chem. A* **1998**, *102*, 1820.
(25) Sanderson, R. T. *Science* **1951**, *144*, 670.
(26) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity − a Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219−3228.
(27) Mortier, W. J.; Ghosh, S. K.; Shankar, S. Electronegativity equalization method for the calculation of atomic charges in molecules. *J. Am. Chem. Soc.* **1986**, *108*, 4315−4320.
(28) Mullay, J. A simple method for calculating atomic charges in charged molecular systems of biochemical interest. *J. Comput. Chem.* **1990**, *12*, 369−375.
(29) Tai, K.; Grant, J. A.; Jhon, M. S.; Scheraga, H. A. Determination of net atomic charges using a modified partial equalization of orbital electronegativity method. 2. Application to ionic and aromatic molecules as models for polypeptides. *J. Phys. Chem.* **1990**, *94*, 0−4746.
(30) Rappe, A. K.; Goddard, W. A. III Charge Equilibration Method for Molecular Dynamics Simulations. *J. Phys. Chem.* **1991**, *95*, 3358−3363.
(31) Cong, Y.; Yang, Z.-Z. General atom-bond electronegativity equalization method and its application in prediction of charge-distributions in polypeptide. *Chem. Phys. Lett.* **2000**, 324−329.
(32) Reed, A. E.; Curtis, L. A.; Weinhold, F. Intermolecular Interactions from a Natural Bond Orbital, Donor-Acceptor Viewpoint. *Chem. Rev.* **1988**, *88*, 899−926.
(33) Bultinck, P.; Langenaeker, W.; Lahorte, P.; De Proft, F.; Geerlings, P.; Waroquier, M.; Tollenaere, J. P. The electronegativity equalization method I: Parametrization and validation for atomic charge calculations. *J. Phys. Chem. A* **2002**, *106*, 7887−7894.
(34) Bultinck, P.; Langenaeker, W.; Lahorte, P.; De Proft, F.; Geerlings, P.; Van Alsenoy, C.; Tollenaere, J. P. The electronegativity equalization method II: Applicability of different atomic charge schemes. *J. Phys. Chem. A* **2002**, *106*, 7895−7901.
(35) Roos-Kozel, B. L.; Jorgensen, W. L. Computer-assisted mechanistic evaluation of organic reactions. 2. Perception of rings. *J. Chem. Inf. Comput. Sci.* **1981**, *21*, 101−111.
(36) Hanser, T.; Jauffret, P.; Kaufmann, G. A new algorithm for exhaustive ring perception in a molecular graph. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1146−1152.
(37) Morales, J.; Martinez, T. J. Classical fluctuating charge theories: the maximum entropy valence bond formalism and relationships to previous models. *J. Phys. Chem. A* **2001**, *105*, 2842−2850.
(38) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. General atomic and molecular electronic−structure system. *J. Comput. Chem.* **1993**, *14*, 1347−1363.
(39) Head, M. S.; Given, J. A.; Gilson, M. K. "Mining Minima": Direct computation of conformational free energy. *J. Phys. Chem.* **1997**, *101*, 1609−1618.
(40) Jorgensen, W. L.; Tirado-Rives, J. The OPLS Potential Function for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666.
(41) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges: II. Parametrization and validation. *J. Comput. Chem.* **2002**, *23*, 1623−1641.
(42) Walters, P.; Stahl, M. www.ccl.net/cca/software/MS-WIN95-NT/babel/index.shtml.

CI034148O