The Electronegativity Equalization Method II: Applicability of Different Atomic Charge Schemes

P. Bultinck,*,† W. Langenaeker,‡ P. Lahorte,§ F. De Proft,|| P. Geerlings,|| C. Van Alsenoy,⊥ and J. P. Tollenaere#

Department of Inorganic and Physical Chemistry, Ghent University, Krijgslaan 281 (S-3), B-9000 Gent, Belgium, Molecular Design and Chemoinformatics, Johnson & Johnson Pharmaceutical Research and Development, Turnhoutseweg 30, B-2340 Beerse, Belgium, European Patent Office, Bayerstrasse 34, D-80335 Munich, Germany, Eenheid Algemene Chemie (ALGC), Free University of Brussels (VUB), Pleinlaan 2, B-1050 Brussels, Belgium, Department of Chemistry, Universiteit Antwerpen, Universiteitsplein 1, B-2610 Antwerpen, Belgium, and Department of Medicinal Chemistry, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, P.O. Box 80082, 3508 TB Utrecht, The Netherlands

Received: February 27, 2002; In Final Form: May 14, 2002

The amenability of different schemes for the calculation of atomic charges in the electronegativity equalization method (EEM) is investigated. To that end, a large training set of molecules was composed, comprising H, C, N, O, and F, covering a wide range of medicinal chemistry. Geometries are calculated at the B3LYP/6-31G* level. Atomic charges are calculated using five different methods, belonging to different types of population analysis. Effective electronegativities and hardness values are calibrated from the different quantum chemically calculated atomic charges. The resulting quality of EEM charges is investigated for the different types of atomic charge calculation methods. EEM-derived Mulliken and NPA charges are in good agreement with the ab initio values, electrostatic potential derived, and Hirshfeld charges show no good agreement.

Introduction

Atomic charges are one of the most used properties in chemistry, both in a (semi-) quantitative as well as in a qualitative sense. Differences in, e.g., reactivity between different molecules are often explained in terms of (partial) atomic charges. Atomic charges are also used in many modeling packages, where they may be treated on equal footing as other force field parameters, such as bond lengths, etc. One of the problems is that atomic charges cannot be obtained from experiment. On the other hand, quantum chemistry opens a way to obtain the atomic charges. Unfortunately, there exists no unique method for the quantum chemical calculation of atomic charges. Many different approaches exist, each with their own advantages and disadvantages, making it hard to decide on what method should be used.

The electronegativity equalization method (EEM) is a method that, based on density functional theory, allows the fast calculation of atomic charges in a large set of molecules. This method does however require knowledge of a number of parameters, which have to be calibrated on the basis of a set of molecules with known structures and for which atomic charges have been calculated quantum chemically (see part I of this series[1]). It is clear that a different method for the calculation of atomic charges will generally result in different EEM parameters.

The aim of this article is to evaluate the extent to which different schemes for the calculation of atomic charges can be used in the calibration of EEM parameters, how these parameters differ among each other, and how good the agreement is between the EEM charges and the quantum chemically calculated charges for the different types of population analysis. As such, it is also an extension of the previous work, focusing on the calibration of parameters derived from Mulliken population analysis.[1]

Theoretical Background

1. The Electronegativity Equalization Method. The theoretical background of the electronegativity equalization method was described previously[1] and will not be repeated. The EEM method is based on the idea of the equalization of the electronegativity. When atoms unite to form a molecule or crystal, their electronegativities adjust to the same, common molecular value χ_{eq}.[2] Mortier et al.[3] derived formulas, which express this electronegativity in terms of the atomic charges and values for the effective electronegativity and hardness of the different elements involved. It suffices to note that, for the calculation of atomic charges, one needs to solve a matrix equation (where N denotes the number of atoms in the molecule or crystal):

$$\begin{bmatrix} 2\eta_1^* & 1/R_{12} & \cdots & 1/R_{1N} & -1 \\ 1/R_{21} & 2\eta_2^* & \cdots & 1/R_{2N} & -1 \\ \cdots & \cdots & \vdots & \vdots & \cdots \\ 1/R_{N1} & 1/R_{N2} & \cdots & 2\eta_N^* & -1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_N \\ \chi_{eq} \end{bmatrix} = \begin{bmatrix} -\chi_1^* \\ -\chi_2^* \\ \cdots \\ \chi_N^* \\ Q \end{bmatrix} \quad (1)$$

It is clear that when the structure (and as such the interatomic

---

* To whom correspondence should be addressed. E-mail: Patrick.Bultinck@rug.ac.be. Fax: +32/9/264.49.83.

† Department of Inorganic and Physical Chemistry, Ghent University.

‡ Johnson & Johnson Pharmaceutical Research and Development.

§ European Patent Office.

|| Free University of Brussels (VUB).

⊥ Universiteit Antwerpen.

Utrecht University.

10.1021/jp020547v CCC: $22.00

distances R_{xy}) and the total charge (Q) of the molecule is known, and the values for the effective electronegativity and hardness (χ_α^* and η_α^*) for the different atoms are known, eq 1 allows the calculation of the N atomic charges (q_x) through standard matrix algebra. χ_α^* and η_α^*, however, are unknown and cannot be calculated directly. Values for these parameters can be obtained through calibration. To that end, both the structures and atomic charges in a set of molecules should be known to calibrate values for χ_α^* and η_α^*. The procedure used for this calibration was described in detail previously.[1] Not for every atom a different effective electronegativity and hardness is calibrated, but rather a common value is calibrated for all atoms of a specific element. This is justified by the fact that distinguishing different sets of values for different valencies of atoms of the same element[1] does not substantially improve the agreement between the DFT and EEM charges.

2. Atomic Charges. The calibration of the effective electronegativity and hardness values requires knowledge of quantum chemically calculated atomic charges. Unfortunately, there is no unique way to obtain atomic charges from quantum chemical calculations. As a result, depending on the type of charges used in the calibration, different values for the effective electronegativity and hardness of the different elements will result. There are several commonly used methods to obtain atomic charges from quantum chemical calculations. These may be divided as follows:[4] methods based on partitioning the wave function in terms of the contributing basis functions, methods based on fitting the electrostatic potential, and methods based on partitioning the charge density in terms of atoms. In the present study, EEM parameters will be derived from DFT calculated structures and atomic charges of the different types.

EEM calibrations have most often been carried out using Mulliken charges. Mulliken population analysis uses the **DS** matrix to distribute electrons into atomic contributions (**D** is the density matrix, **S** is the overlap matrix). Mulliken population analysis is clearly a population analysis of the first kind, where the partitioning is carried out in terms of the basis functions. One of the most important criticisms against the Mulliken population analysis is the fact that contributions involving basis functions on different atoms are simply partitioned equally between the two atoms, whereas it may be argued that the most electronegative atom should receive most of the shared electrons. The Mulliken population analysis is only one representative of this class of population analyses. In fact, both the Mulliken and Löwdin population analysis are special cases of what may be written in general as $\mathbf{S}^n\mathbf{D}\mathbf{S}^{1-n}$ population analyses (where $n = 0$ for Mulliken and $n = 0.5$ for Löwdin). Several criticisms can be formulated against population analyses based on these methods. Some diagonal elements of the **DS** matrix (in Mulliken population analysis) can exceed two, thereby harboring more than two electrons which is a violation of the Pauli principle. Some elements of the matrix may become negative, implying a negative number of electrons. Furthermore, Mulliken charges are rather basis-set-dependent.[5] When diffuse basis functions are added, their populations still count for the atom they are centered on, despite the fact that they contribute to the wave function at large distances from the nucleus of the atom.

Quite often in, e.g., medicinal chemistry, molecules interact at distances which are of the same order as the van der Waals radii. Given the importance of the electrostatic potential in determining these interactions, several schemes have evolved from attempts to fit atomic charges which generate an electrostatic potential that best matches that generated from the wave function, under the constraint that the total molecular charge is conserved. Extra constraints may be added, such as the requirement that the atomic charges should reproduce the molecular dipole moment. Different methods have been published, differing in the number and the sampling of the points on the van der Waals surface or some other surface where the electrostatic potential is calculated. Methods may also differ in the way the fitting is done or the kind and number of constraints in the fitting procedure. A major drawback of this method is that the high number of points on the surface used compared to the number of atomic charges that need to be fitted makes the methods statistically underdetermined.[6] The electrostatic potential on the van der Waals surface, or some other usually user-determined surface, is mainly determined by the atoms nearest to the molecular surface. Atoms embedded in the inside of the molecule contribute less to the electrostatic potential on the surface. As a result, there is quite some ambiguity in their charges. The full set of atomic charges is a highly redundant set, and many different sets of charges may yield an almost fully equivalent fit to the wave-function-based electrostatic potential. Despite these shortcomings, these types of atomic charge calculations are widely used. In the present study, charges obtained through the CHELPG[7] algorithm and the Merz–Singh–Kollman[8] schemes are used. The latter scheme has been tested in the context of EEM previously by Njo et al.[10]

A rigorous way to obtain atomic charges is by analyzing the wave function, but not in terms of basis functions. The amount of electronic charge at some point in space can readily be calculated from the square of the wave function or is obtained directly from density functional theory. Nuclei are considered as point charges. The problem, however, is how to derive atomic charges. This is the consequence of the fact that quantum chemical methods such as density functional theory in fact do not consider a molecule as a set of atoms held together by chemical bonds but as a set of nuclei surrounded by electrons. The reason for the existence of different schemes to obtain atomic charges from the wave function lies in the existence of different ways to define an atom in a molecule. Once an atomic basin is defined, one can simply calculate the atomic charge by integrating the total electronic charge from the wave function in the atomic basin and subtract it from the nuclear charge. An example of such an approach is the AIM (atoms in molecules) method developed by Bader et al.[9] In this method, atomic basins are identified on the basis of the properties of the electron density and its derivatives. Unfortunately, this is computationally a quite demanding technique. Another technique which is inspired on the idea of dividing electron density in terms of atoms is the Hirshfeld approach.[11] This is a conceptually simple and efficient model to obtain atomic charges. In this method, one distinguishes the actual molecule and a so-called promolecule. The actual molecule is the molecule with its electron density as produced by the quantum chemical method. The promolecule is constructed by placing the same atoms as in the actual molecule on exactly the same positions in space. The electron density in this promolecule, however, is not the result of a quantum chemical calculation on the molecule, but at every point r in space, the electron density is taken as the sum of the isolated atom densities at this same point. These isolated atom densities are derived from calculations using the same theoretical method, basis set, etc., as for the calculations on the actual molecule. A weight factor for a specific atom is introduced as the ratio of the amount of isolated atom electron density to the total *promolecule* density at different points in space. For each of these points, the amount of electron density due to an atom A in the *actual* molecule is then obtained as the weight factor of atom A times the amount of electron density of the *actual* molecule at that point. By integrating over the entire space for atom A, and subtracting this amount of electronic charge from its nuclear charge, one obtains the atomic charge. The main assumption of this method is naturally that the weight of atom A in the actual molecule is adequately/correctly represented by the promolecule. Their applicability and influence of different parameters on the resulting charges was checked, and Hirshfeld charges were found to be instructive values in interpreting chemical observations.[12−14] This particular scheme for atomic charge calculations also merits special attention in the definition of what an atom is in a molecule.[15]

Still another approach is that used in natural population analysis. This method by Weinhold et al.[16] is based on blocking out the density matrix (obtained from a quantum chemical calculation) in terms of the different atoms. Each atom has its own block in the density matrix, and after a few steps (see ref 16 for details on the NBO methodology, these steps involve diagonalization and orthogonalization), one obtains the so-called natural atomic orbitals. The diagonal elements in the different blocks are the atomic orbital populations, and by summing these, the total electronic charge on the atom is found, and as a result the atomic charge. Natural population analysis is a very fast method, because it requires only matrix diagonalization steps and orthogonalization. Natural population analysis has been shown to be quite useful in discerning intra- and intermolecular interactions, because even small shifts in their values are often already relevant.[17,18]

Computational Methods

Choice of the Training Set and Quantum Chemical Calculations. As the values for χ_α^* and η_α^* cannot readily be calculated, they need to be determined by calibration. To that end, one needs a set of molecules, for which both structures and atomic charges (calculated using the different methods described above) are known. The same calibration set as in part I of the present study has been used.[1] This set holds 138 molecules containing only H, C, N, O, and F and covers a wide range of functional groups in medicinal and organic chemistry. The geometries of all molecules were optimized at the B3LYP/6-31G* level, using the Gaussian 98 program.[19] Mulliken, CHELPG, Merz−Singh−Kollman (MK), NPA, and Hirshfeld charges were calculated for all molecules. All atomic charges were calculated at the B3LYP/6-31G* level of theory. Hirshfeld charges were calculated from the DFT generated densities using the Hirshfeld program by Rousseau et al.[13]

Calibration of Parameters and Fitness Evaluation. The calibration procedure used, is the same as used previously for the calibration of Mulliken derived effective hardness and electronegativity values.[1] The calibration procedure will therefore not be discussed in detail. The goal of the calibration is to determine the specific set of 10 parameters (χ^* and η^* for H, C, N, O, and F, respectively) that, when inserted in the EEM matrix eq 1, will yield EEM charges that differ minimally from the corresponding quantum chemically calculated charges in the training set. The quality of the fit between the B3LYP/6-31G* and the EEM charges is evaluated using the following fitness function:[1]

$$\Delta q = \sum_{z=1}^{N_{\mathrm{el}}} \frac{\sum_{i=1}^{M}\sum_{j=1}^{N_{i,z}} (q_{jiz}^{\mathrm{EEM}} - q_{jiz}^{\mathrm{DFT}})^2}{N_z} \qquad (2)$$

TABLE 1: Correlation Coefficients (R^2 in %) and Linear Regression Equations for Each Combination of Two Atomic Charge Schemes Based on B3LYP/6-31G* Calculations[a]

Scheme 1	Scheme 2	a	b	R^2
Mulliken	CHELPG	0.84	0.00	73
Mulliken	MK	0.90	0.00	79
Mulliken	NPA	1.29	0.00	96
Mulliken	Hirshfeld	0.29	0.00	82
CHELPG	MK	1.01	0.00	96
CHELPG	NPA	1.07	0.00	65
CHELPG	Hirshfeld	0.28	0.00	78
MK	NPA	1.10	0.00	73
MK	Hirshfeld	0.28	0.00	78
NPA	Hirshfeld	0.22	0.00	79

[a] a and b are obtained from linear regression of the type: $Q_{\mathrm{Scheme1}} = a.Q_{\mathrm{Scheme2}} + b$.

where z refers to a specific element (here H, C, N, O, or F), i to a molecule from the training set and j to an atom of element z in molecule i. The upper summation indices are the total number of elements ($N_{\mathrm{el}} = 5$ in this case), M is the number of molecules, and $N_{i,z}$ is the number of atoms of element z in molecule i. N_z is the total number of atoms of element z in the calibration set. Because the different schemes to obtain atomic charges yield (highly) different values for the charges, a second fitness function was also used; based on relative deviations. This entails the minimization of eq 3 in terms of the effective electronegativity and hardness parameters:

$$\Delta q = \sum_{z=1}^{N_{\mathrm{el}}} \frac{\sum_{i=1}^{M}\sum_{j=1}^{N_{i,z}} |(q_{jiz}^{\mathrm{EEM}} - q_{jiz}^{\mathrm{DFT}})/q_{jiz}^{\mathrm{DFT}}|}{N_z} \qquad (3)$$

Calibrations using both (2) and (3) were performed. For both calibrations, the optimal effective electronegativity and hardness values were taken, and the agreement between the EEM and DFT derived charges within each type of population analysis examined through linear regression. In the following, only those results for the parameters that show the best correlation between the EEM and DFT charges are presented. The actual calibration is a stepwise process. First, χ^* and η^* values for all elements are assigned randomly. These values are used to calculate EEM charges on all atoms through standard matrix algebra (see eq 1). Equations 2 or 3 are used as fitness function to evaluate the quality of the fit between the DFT charges and the EEM charges. The fitness function is then minimized by updating the χ^* and η^* values to obtain a smaller value for the fitness function. As minimization algorithm a combination of random sampling and the simplex method in multidimensions,[20] followed by a genetic algorithm is used. The exact numerical techniques used were described previously.[1]

Results and Discussion

Comparison of Charges. For all molecules in the test set, the atomic charges were calculated using the Mulliken, CHELPG, MK, NPA and Hirshfeld methods. From the results one clearly sees that the different techniques yield quite different charges. Not only absolute charges are different, trends may also differ among the different atomic charge calculation schemes.

Table 1 presents the correlation coefficients for the linear regression $Q_{\mathrm{Scheme1}} = a.Q_{\mathrm{Scheme2}} + b$ for each possible combination of charge schemes. In these correlations, no distinction was made between the different elements. From the results, it is seen

TABLE 2: Optimized Values of χ^* and η^* for the Different Types of Population Analysis[a]

atom		Mulliken	CHELPG	MK	NPA	Hirshfeld
H	χ^*	1.00	1.00	1.00	1.00	1.00
	η^*	17.95	15.29	14.30	19.44	19.31
	N	930	930	930	930	930
C	χ^*	5.26	2.77	3.15	8.49	2.22
	η^*	9.00	8.82	8.36	9.15	9.98
	N	602	602	602	602	602
N	χ^*	8.80	6.31	5.48	13.45	3.51
	η^*	9.39	9.03	8.75	10.64	10.99
	N	105	105	105	105	105
O	χ^*	14.72	33.94	26.35	27.06	71.76
	η^*	14.34	35.57	29.06	19.63	156.79
	N	101	101	101	101	101
F	χ^*	15.00	9.88	8.46	39.18	11.78
	η^*	19.77	19.53	17.13	44.10	43.03
	N	65	65	65	65	65

[a] All values are in eV. For ease of comparison, all sets of parameters are transferred to a value of 1.00 for the effective electronegativity of hydrogen. N denotes the number of atoms of each element in the test set (equal for all types of population analysis).

that there may indeed be important differences between different schemes to calculate atomic charges. For all possible combinations, the constant b is always very close to zero. The correlation coefficient shows that there are some cases where there is some important similarity, most notably between NPA and Mulliken charges. Given the fact that good performance of the EEM method for the Mulliken charges[1] was found, similar or better performance for the NPA based EEM method is expected. The Mulliken charges do not show similarity with any other technique, so for these techniques, no prediction of their applicability to EEM can be made. Another good correlation from Table 1 is that between the CHELPG and MK atomic charges. This is not unexpected because they belong to the same type of population analysis and differ mainly in the numeric techniques used.

Concerning the values of the charges, there are a number of clear differences between different types of atomic charges. Hirshfeld charges are consistently much smaller than any other set of charges. For the other combinations, the term a is always relatively near to one. A special remark should be made on the charges produced by the CHELPG and MK methods. Quite a lot of negatively charged hydrogen atoms were found, whereas these are not expected from the Lewis structures and chemical intuition. None of the other methods to calculate atomic charges yield any negatively charged hydrogen atoms. This behavior for the ESP fitting charge schemes is due to the problems mentioned earlier, such as the statistical underdetermination, redundancies in the fitting, and embedded atoms. As such, the negative charge on the hydrogen atoms is not due to chemical reasons. Because all atomic charges in the calibration set are used to calibrate the effective electronegativity and hardness parameters against, and because it is not possible to detect which hydrogen atoms would have a chemically reasonable charge and which not, we expect problematic behavior for the ESP derived schemes.

Calibrated Parameters. In Table 2, the calibrated effective atomic electronegativities and hardness parameters calibrated in the procedure described above are presented for the different types of population analysis. For ease of comparison, the value for the effective electronegativity of hydrogen is adjusted to a value of 1.00 eV. For all other elements, this simply involves adding or subtracting this same term to the effective electronegativity. The possibility of changing the reference for the electronegativity scale among the different types of population analysis is due to the fact that the effective electronegativities can only be determined up to a constant. From eq 1, it is clear that changing the effective electronegativity of all atoms by adding the same term will solely affect the equalized electronegativity (it will be raised or lowered with this same term) but not the charges resulting from (1). The effective hardnesses cannot be changed in a similar way. Table 2 shows that the absolute values of the different parameters may differ quite strongly between different types of population analysis. This is not unexpected given the differences in charges resulting from the different types of population analysis. More importantly, there are also differences in relative magnitudes from the different types of population analysis. Most notably, in the electrostatic potential and Hirshfeld-charge-derived parameters, oxygen can have a higher effective electronegativity and hardness than fluorine. This is quite contrary to chemical intuition. Different factors play a role. In the case of the ESP-derived charges, we already discussed the numerical problems arising in these schemes. These are expected to influence the set of charges in the calibration set and, as such, the parameters resulting from the calibrations. These may influence the parameters in some unpredictable way. For the Hirshfeld scheme, a different effect is responsible. An analysis of the spread of oxygen charges for all types of population analysis reveals that the Hirshfeld charges show the smallest spread. This has an important numerical consequence. Considering eq 1, this means that a random change in χ^* can be compensated by a change in η^* (with the chance in η^* proportional to the change in χ^* as described in part I[1]) without much effect on the resulting charges. This means that a different value for the effective electronegativity can be chosen and compensated by an adjusted hardness, thereby leading to almost exactly the same DFT-EEM agreement. If the charges are spread more, such a compensation is no longer possible. This is reminiscent of the earlier described problem[1] when not including sufficient different valencies for an element. For the Mulliken and NPA methods, no such problems arise as in the ESP- and Hirshfeld-based schemes. Both for the Mulliken and NPA based methods, the expected trends are found for both the effective electronegativity and hardness.

Quality of the EEM Charges. To calibrate the effective electronegativity and hardness values, the fitness function (2) or (3) was minimized by comparing the EEM charges and DFT charges. The quality of the EEM charges is expressed as the average absolute difference between the EEM and DFT charges. As pointed out before, there may be important differences between the charges from different types of population analysis. For example, Hirshfeld charges are relevantly smaller than the charges from any other method. One should therefore not only consider the absolute deviations but also the relative deviations. Both values describing the agreement are given in Table 3. In the calculation of the relative deviations, charges between -0.05 and $+0.05$ are omitted, because the slightest difference between the DFT and EEM charges could have a very large influence on the relative error (e.g., a difference of 0.001 is chemically irrelevant, but on an almost zero charge, such as 0.0001 contributes 1000% to the relative error).

The table shows that Mulliken population analysis performs quite well, as found earlier.[1] It is found that natural population analysis (NPA) shows a relatively good agreement between the EEM and DFT charges. Figure 1a–e shows the correlation for the different elements in the test set for the NPA method. The agreement is relatively good. This agrees with the findings in Table 3. It should be noted that in some cases the correlation

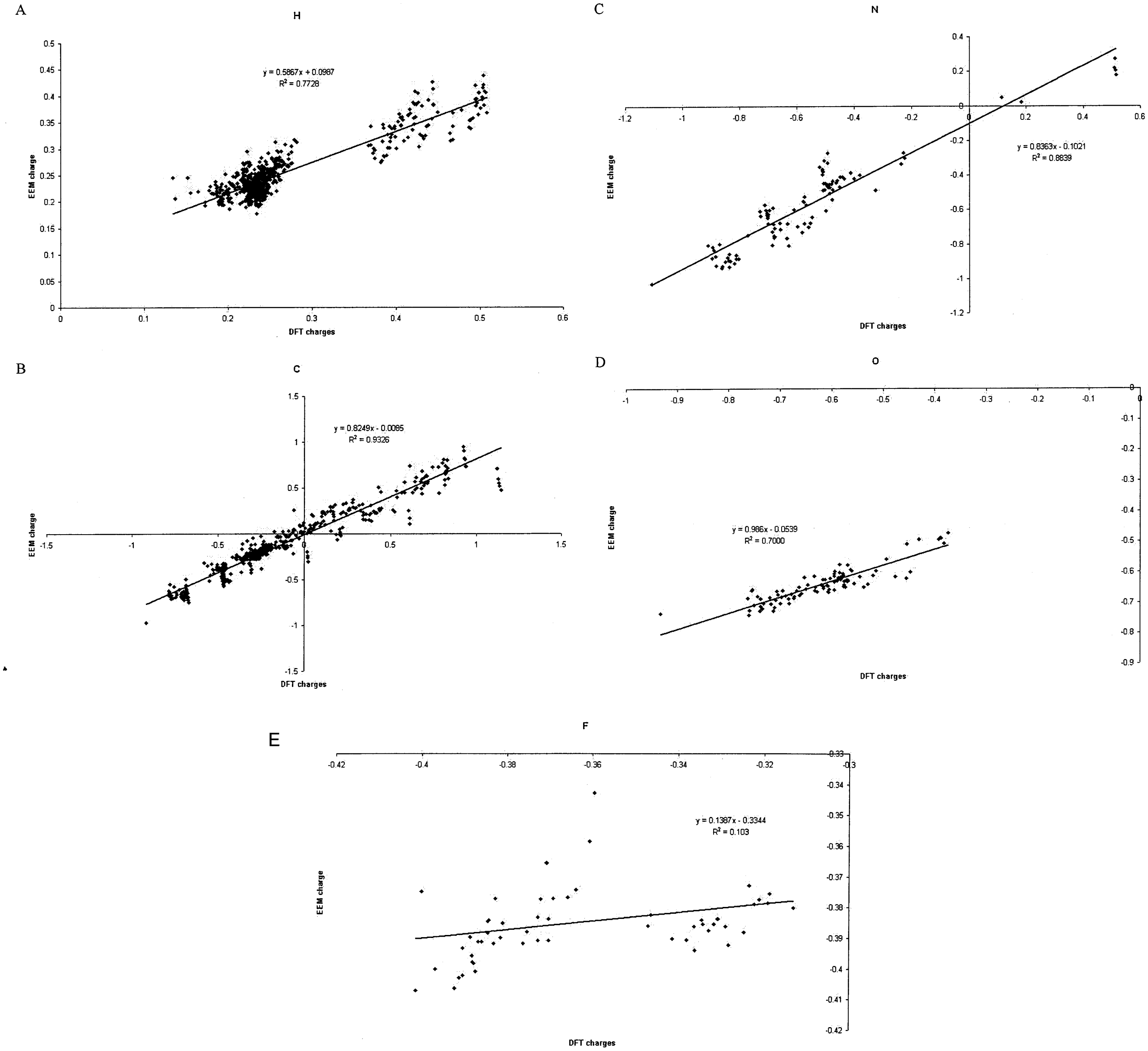

Figure 1. Illustration of the agreement between EEM and DFT charges based on the NPA population analysis. Separate graphs illustrate the agreement for the different elements involved.

TABLE 3: Average Absolute Deviations and Average Relative Deviations (in %) for the Different Atom Types and the Combined Set of All Atoms, Describing the Agreement between EEM and DFT Derived Charges[a]

	H	C	N	O	F	all atoms
Mulliken	0.02	0.06	0.06	0.04	0.02	0.03
CHELPG	0.05	0.12	0.17	0.06	0.04	0.08
MK	0.05	0.12	0.20	0.06	0.04	0.08
NPA	0.02	0.09	0.08	0.05	0.03	0.05
Hirshfeld	0.01	0.02	0.05	0.06	0.02	0.02
Mulliken	10	28	12	8	7	19
CHELPG	93	55	53	19	20	81
MK	76	53	63	19	22	74
NPA	9	30	16	8	8	20
Hirshfeld	28	32	59	29	18	34

[a] EEM charges were calculated with the parameter sets given in Table 2 for each type of population analysis.

coefficient is quite low. One should however bear in mind that in these cases the spread in DFT charges is also very low. The other types of population analysis clearly perform less. Especially, the ESP derived charges show poor agreement. This is contrary to the finding of Njo et al.,[10] who found that MK charges can also be predicted quite reliably using the EEM scheme.[10] They do however also mention that it does not perform as well as the combination of EEM and Mulliken charges. The most likely reason for the different conclusion about the performance of EEM for MK charges is the more extended range of valencies in the present, more extended calibration set.

The quality of Mulliken charges based on EEM was investigated previously and was shown to be very good.[1] One of the most important objections against the Mulliken population analysis is its arbitrary division of the electron density between two atoms, more precisely, it is divided in half irrespective of the difference of the electronegativity of the two partners in the bond. A method that does not suffer from this shortcoming is NPA, which also performs quite well in EEM. Both techniques also succeed in giving parameters that are applicable within a very large range of valencies. They are also the only

TABLE 4: Average Absolute Deviations and Average Absolute Relative Deviations (in %) for the Different Atom Types in the Neuroleptics, Describing the Agreement between EEM and DFT Derived Charges[a]

	H	C	N	O	F	all atoms
Mulliken	0.01	0.06	0.07	0.04	0.04	0.04
CHELPG	0.05	0.12	0.12	0.08	0.04	0.08
MK	0.04	0.11	0.14	0.08	0.04	0.08
NPA	0.02	0.07	0.10	0.05	0.07	0.05
Hirshfeld	0.01	0.02	0.04	0.02	0.03	0.02
Mulliken	7	29	14	6	12	16
CHELPG	583	63	31	21	21	330
MK	182	58	109	24	23	126
NPA	9	24	19	11	20	15
Hirshfeld	44	35	57	32	32	42

[a] EEM charges were calculated with the parameter sets given in Table 2 for each type of population analysis.

techniques that yield parameters for nitrogen, applicable for both $N^{\delta+}$ and $N^{\delta-}$ atoms.

ESP charges perform poorly, especially for hydrogen. This is due to the reasons mentioned earlier, which cause many hydrogen atoms to bear a negative charge.

Validation of EEM. The parameters derived from the NPA-based calibrations yield charges which agree fine with the original NPA charges. To test the applicability of the EEM method based on NPA charges, the same validation set of molecules was used as previously for the Mulliken-based set of EEM parameters.[1] This set holds 12 structures of six neuroleptics, medium size molecules holding all elements calibrated in the present study. The average absolute and relative differences between the DFT calculated and EEM predicted charges (using the parameters from Table 2) per element are shown in Table 4. The results clearly reveal that Mulliken and NPA are the best performing types of population analysis in the EEM context. Given the agreement between the DFT and EEM charges, both in absolute and relative errors, and the speed of the EEM calculations, excellent performance of the Mulliken−EEM and NPA−EEM methods may be concluded. The calculations are equally long for all types of charges to be generated through EEM (contrary to the B3LYP calculations) and take less than 30 ms on a current Pentium IV PC per type to calculate all atomic charges in the 12 structures. Mulliken population analysis, the least costly population analysis, requires the SCF procedure to be carried out, which takes, for example, one single neuroleptic molecule 32 min on the same PC. For the 12 molecules together, the EEM approach is 10^6 times faster than the DFT calculations. This speed-up is even more impressive for more CPU time-consuming kinds of population analysis.

The ESP derived charges do not yield useful EEM parameters. The reason for this is again probably due to the existence of many solutions for the derivation of atomic charges. As a result, one can even obtain quite different ab initio charges on atoms that, although in different molecules, are chemically quite similar. Also, there are quite a number of atoms whose charges cannot be derived from the electrostatic potential to a sufficient degree of confidence. This is the case for embedded atoms. Calibration of parameters against Hirshfeld charges does not yield a very useful set of parameters either. Apparently, the assumption of conservation of the weight factor between the promolecule and the actual molecule is too crude an approximation.

Conclusions

Using the same calibration set throughout, the amenability of different schemes for the quantum chemical calculation of atomic charges to EEM was tested. Important differences between the different schemes were found. Mulliken population analysis performs quite well, as does NPA, but other types of population analyses were found to yield poor agreement between the DFT and EEM calculated charges. As a conclusion, it is recommended to base EEM work on NPA charges, because Mulliken population analysis suffers from well-known drawbacks, and the other schemes do not perform satisfactorily.

We found that the parameters returned by the calibration are applicable over a very wide range of valencies for each element. Different factors that influence the values of the parameters were identified. These include the influence of the calibration set, the population analysis used to calibrate the parameters against, and the numerical character of the EEM equations especially in case of the electrostatic potential derived charge schemes.

The EEM procedure constitutes a very useful approach in screening compounds or explaining observed properties, because they allow the calculation of ab initio quality charges in large sets of (large) molecules without the cost of the actual ab initio calculations.

Acknowledgment. P.B. thanks Janssen Research Foundation, Ghent University and the Fund for Scientific Research-Flanders (Belgium) for their grants to the Quantum Chemistry group at Ghent University. P.L. is grateful to Ghent University for a doctoral scholarship. P.G. thanks the Free University of Brussels and the Fund for Scientific Research-Flanders (Belgium) for continuous support to his group.

References and Notes

(1) Bultinck, P.; Langenaeker, W.; Lahorte, P.; De Proft, F.; Geerlings, P.; Waroquier, M.; Tollenaere, J. P. *J. Phys. Chem. A* **2002**, *106*, 7887.

(2) Sanderson, R. T. *Polar Covalence*; Academic Press: New York, 1983.

(3) Mortier, W. J.; Ghosh, S. K.; Shankar, S. *J. Am. Chem. Soc.* **1986**, *108*, 4315.

(4) Jensen, J. *Introduction to Computational Chemistry*; John Wiley and Sons: New York, 1999.

(5) Hehre, W. J.; Radom, L.; Schleyer, P. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.

(6) Francl, M. M.; Carey, C.; Chilian, L. E.; Gange, D. M. *J. Comput. Chem.* **1996**, *17*, 367.

(7) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361.

(8) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129.

(9) Bader, A. M. *Atoms in Molecules*; Clarendon Press: Oxford, U.K., 1990.

(10) Njo, S. L.; Fan, J.; van de Graaf, B. *J. Mol. Catal. A* **1998**, *134*, 79.

(11) Hirshfeld, F. L. *Theor. Chim. Acta* **1977**, *44*, 129.

(12) De Proft, F.; Martin, J. M. L.; Geerlings, P. *Chem. Phys. Lett.* **1996**, *250*, 393.

(13) Rousseau, B.; Peeters, A.; Van Alsenoy, C. *Chem. Phys. Lett.* **2000**, *324*, 189.

(14) De Proft, F.; Van Alsenoy, C.; Peeters, A.; Langenaeker, W.; Geerlings, P. *J. Comput. Chem.* **2002**, accepted for publication.

(15) Nalewajski, R. F.; Parr, R. G. *Proc. Nat. Acad. Sci. U.S.A.* **2000**, *97*, 8879.

(16) Reed, A. E.; Curtiss, L. A.; Weinhold, F. *Chem. Rev.* **1988**, *88*, 899.

(17) Bultinck, P.; Van Alsenoy, C.; Goeminne, A. *J. Phys. Chem. A* **2001**, *105*, 9203.

(18) Gu, Y.; Kar, T.; Scheiner, S. *J. Am. Chem. Soc.* **1999**, *121*, 9411.

(19) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi,

I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A.7; Gaussian, Inc.: Pittsburgh, PA, 1998.

(20) Walters, F. H.; Parker, L. R., Jr.; Morgan, S. L.; Deming, S. N. *Sequential Simplex Optimization*; CRC Press: Boca Raton, FL, 1991. An electronic reprint of this book may be obtained from www.multisimplex.com. The authors wish to thank MultiSimplex AB for making this book available.