J. Phys. Chem. A **2002,** *106,* 7887−7894

The Electronegativity Equalization Method I: Parametrization and Validation for Atomic Charge Calculations

P. Bultinck,*,† W. Langenaeker,‡ P. Lahorte,§ F. De Proft,‖ P. Geerlings,‖ M. Waroquier,⊥ 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, Laboratory of Theoretical Physics, Ghent University, Proeftuinstraat 86, B-9000 Gent, 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 applicability of the electronegativity equalization method (EEM) is investigated for the fast calculation of atomic charges in organic chemistry, with an emphasis on medicinal chemistry. A large training set of molecules was composed, comprising H, C, N, O, and F, covering a wide range of medicinal chemistry. Geometries and atomic charges are calculated at the B3LYP/6-31G* level, and from the calculated charges, effective electronegativity and hardness values are calibrated in a weighted least-squares fashion. The optimized parameter set is compared to other theoretical as well as experimental values and origins of the differences discussed. An approach toward extension of EEM to include new atoms is introduced. The quality of the EEM charges is assessed by comparison with B3LYP/6-31G* charges calculated for a set of medicinal molecules, not contained in the training set. The EEM approach is found to be a very powerful way to obtain ab initio quality charges without the computational cost of the ab initio approach.

Introduction

In recent decades, the speed of computers has increased dramatically together with the continuing development of algorithms and methods in quantum chemistry. As a consequence, the size of the molecules that can be studied quantum chemically has increased, as well as the number of calculations that can be performed in a limited time frame.

Despite this evolution, there are still many areas where quantum chemical calculations are hardly feasible. The example addressed in this paper is the screening of many thousands of molecules for potentially interesting properties, as is routinely performed in pharmaceutical research. Usually, only a limited amount of time is available in such screenings, and quantum chemical calculations are not an option. On the other hand, quantum chemistry gives access to many highly interesting molecular properties. Consequently, there lies great promise in ways to obtain these descriptors with quantum chemical quality at great speed.

Atomic charges are quite often used as tools for the interpretation of e.g., molecular reactivity. As such, they are valuable parameters in a QSAR environment, and the fast calculation of atomic charges for large molecules, or for a large number of molecules could be very useful.

The scope of this series of articles is to test the application of the electronegativity equalization principle as a tool for generating atomic charges, adhering to the following requirements. The procedure should be sufficiently fast, allowing the calculation of atomic charges in large sets of (large) molecules in a limited time frame (in the order of 1.5 million molecules of medium size an hour on a current PC). The charges obtained from the electronegativity equalization scheme should reproduce ab initio quality charges and should exhibit the ab initio conformation dependence of the charges. The scheme should be extendable to include new elements and should be applicable for every element over a wide range of chemical configurations.

In the first part of the series, the theoretical background is addressed and the calibration process described, the quality of the resulting charges is assessed and the chemical meaning of the parameters obtained from the calibration is discussed. In the second part of this series, the suitability for implementation in the electronegativity equalization method (EEM) of different ab initio charge evaluation schemes is investigated.

Theoretical Background

The electronegativity equalization principle formulated by Sanderson[1−2] states that when molecules are formed the electronegativities of the constituent atoms become equal, yielding the molecular, equalized (Sanderson) electronegativity. Several formalisms have evolved from this principle. Well-known examples are the partial equalization of orbital electronegativity by Gasteiger et al.,[3] the charge equilibration method (Qeq),[4] the atom-bond electronegativity equalization method (ABEEM),[5] and chemical potential equalization methods.[6−7]

---

* 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).

⊥ Laboratory of Theoretical Physics, Ghent University.

\# Utrecht University.

10.1021/jp0205463 CCC: $22.00

The present study is based on the EEM of Mortier et al.[8] In EEM, the electronegativity of an atom α in an *N*-atom molecule is shown to be

$$\chi_{\mathrm{eq}} = \chi_\alpha = \chi_\alpha^0 + \Delta\chi_\alpha + 2(\eta_\alpha^0 + \Delta\eta_\alpha)q_\alpha + \sum_{\beta\neq\alpha}^{N}\frac{q_\beta}{R_{\alpha\beta}} \qquad (1)$$

In this expression, χ_α^0 and η_α^0 represent the isolated atom electronegativity[9] and hardness,[10] respectively, q_α represents the atomic charge on atom α, and $R_{\alpha\beta}$ represents the interatomic distance between atoms α and β. $\Delta\chi_\alpha$ and $\Delta\eta_\alpha$ are corrections to the respective isolated atom values due to the incorporation of the atom in a molecule (or crystal). The external potential represented by the last term accounts for the influence of the surrounding atoms (or molecules). For the actual derivations of the EEM formulas, the reader is referred to Mortier et al.[8] Using χ_α^* $(=\chi_\alpha^0 + \Delta\chi_\alpha)$ and $\eta_\alpha^*(=\eta_\alpha^0 + \Delta\eta_\alpha)$ for the effective atomic electronegativity and hardness values, respectively, the operational, simplified formula used throughout the study is given by

$$\chi_{\mathrm{eq}} = \chi_\alpha = \chi_\alpha^* + 2\eta_\alpha^* q_\alpha + \sum_{\beta\neq\alpha}^{N}\frac{q_\beta}{R_{\alpha\beta}} \qquad (2)$$

The EEM approach exhibits a number of interesting features. First of all, a solid proof has been provided within the context of Hohenberg−Kohn density functional theory for Sandersons electronegativity equalization principle by Parr et al.[9,10] Politzer and Weinstein[11] later confirmed this.

Second, EEM holds the potential of generating, at a very modest computational cost, atomic charges that are both connectivity- and geometry-dependent. In an *N*-atom molecule, the atomic charges and the (molecular) electronegativity χ_{eq} are the unknowns, the actual values of which can be determined by solving a set of $N + 1$ linear equations. *N* of these equations are obtained by equilibrating the individual atomic electronegativities to the molecular electronegativity ($\chi_{\mathrm{eq}} = \chi_\alpha = \chi_\beta = ...$), whereas one supplementary equation is obtained by constraining the sum of the atomic charges to equal the total molecular charge ($Q = \Sigma_\alpha^N q_\alpha$). In matrix form, this may be written as

$$\begin{bmatrix} 2\eta_1^* & 1/R_{12} & \cdots & 1/R_{1N} & -1 \\ 1/R_{21} & 2\eta_2^* & \cdots & 1/R_{2N} & -1 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 1/R_{N1} & 1/R_{N2} & \cdots & 2\eta_N^* & -1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_N \\ \chi_{eq} \end{bmatrix} = \begin{bmatrix} -\chi_1^* \\ -\chi_2^* \\ \vdots \\ -\chi_N^* \\ Q \end{bmatrix} \qquad (3)$$

Finally, in addition to the molecular electronegativity and atomic charges, the EEM framework also allows a straightforward and transparent calculation of other fundamental properties such as the total electronic energy, hardness, and reactivity indices, such as Fukui functions and local softness.[12] A recent overview of the status and theory of EEM within the broad context of conceptual DFT can be found in Geerlings et al.[13]

It is clear that when the parameters χ_α^* and η_α^* are known one can use standard matrix algebra to calculate the atomic charges from eq 3. χ_α^* and η_α^*, however, are unknown and cannot be calculated directly. Hence, the need for calibration. In the present study, χ_α^* and η_α^* are calibrated using a large training set, consisting of small organic molecules. The performance of EEM will be judged by testing it on a validation set consisting of molecules with known medicinal use.

Computational Methods

Choice of the Training Set. In view of the investigation of the applicability of EEM to organic molecules, a large training set of small molecules was constructed containing hydrogen, carbon, nitrogen, oxygen, and fluorine atoms. Our main concerns in the design of this set were that it contained all of the most important functional groups used in medicinal chemistry and that it was well “equilibrated”. This means that the calibration set should hold a significant number of atoms of the five included elements. This aspect, which is especially important with respect to the mathematical stability of the ensuing calibration process and the chemical relevance, has been insufficiently stressed in the applications presently available. The list of molecules used is available as Supporting Information. The calibration set consists of 138 molecules, holding 930 hydrogen atoms, 602 carbon, 105 nitrogen, 101 oxygen, and 65 fluorine atoms.

Quantum Chemical Calculations. The geometries of all molecules were optimized at the B3LYP/6-31G* level, using the Gaussian 98 program.[14] Mulliken charges were calculated at the same level. It may be argued that Mulliken population analysis is not an optimal way to calculate atomic charges. In this context, the question, however, may be raised what an optimal way of calculating atomic charges would be like. In previous studies, the Mulliken population analysis was also most often used, and because we want to compare results with these studies, only Mulliken atomic charges are considered in the present study. Admittedly, there are other ways to obtain atomic charges, some of which have a more sound theoretical background. In this study, attention is drawn mainly to the calibration process, and the quality of calibrations, whereas a following report[15] will focus on the capability of EEM to calculate other types of atomic charges.

Calibration of Parameters. The approach adopted by Mortier et al.[16,17] when calibrating the effective atomic electronegativity and hardness consists of first choosing a specific electronegativity and hardness scale for the χ^0 and η^0 values. Consequently, the effective electronegativity and hardness scale is “fixed” by arbitrarily assigning the effective atomic electronegativity of O a value of 8.5 eV. The remaining seven parameters ($\Delta\chi$ and $\Delta\eta$ for H, C, and N, and $\Delta\eta$ for O) are then determined using multiple regression. Fluorine parameters were not calibrated previously. In the present study, no restrictions were imposed on any of the parameters. Software was developed in-house to calibrate the optimal χ^*- and η^* values from the B3LYP/6-31G* molecular equilibrium geometries and atomic charges for the molecules included in the training set. The goal of the calibration is to determine the specific set of ten parameters (χ^* and η^* for H, C, N, O, and F, respectively) that, when inserted in the EEM matrix eq 3, 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 must be evaluated in some fitness function. The simplest such function is a least-squares fit, minimizing

$$\Delta q = \sum_{z=1}^{N_{\mathrm{el}}}\sum_{i=1}^{M}\sum_{\alpha=1}^{N_{i,z}}(q_{\alpha iz}^{\mathrm{EEM}} - q_{\alpha iz}^{\mathrm{DFT}})^2 \qquad (4)$$

where *z* refers to a specific element (here H, C, N, O, or F), *i* to a molecule from the training set, and α 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. Equation 4 is, however, not optimal, because in any realistic training set the number of atoms of the different elements will always differ substantially. As an example, the population of hydrogen atoms in any realistic training set will always be substantially larger than that of any of the other elements. Furthermore, the average charge of the atoms belonging to the different elements is also expected to differ significantly. Consequently, the contributions of the different elements to the minimization criterion (4) may vary substantially from element to element. As a result, the "weight" of a particular element can be artificially altered in the calibration process, which may jeopardize the determination of chemically meaningful values. To alleviate this problem, a weighted minimization criterion is used. The most efficient way proved to be a least-squares fit with the population of each element, where the notation N_Z denotes the number of atoms of element Z:

$$\Delta q = \sum_{z=1}^{N_{\mathrm{el}}} \frac{\sum_{i=1}^{M}\sum_{\alpha=1}^{N_{i,z}} (q_{\alpha iz}^{\mathrm{EEM}} - q_{\alpha iz}^{\mathrm{DFT}})^2}{N_z} \qquad (5)$$

The actual calibration is a stepwise process. First, χ^* and η^* values for all elements are assigned randomly. These values are then used to calculate EEM charges on all atoms through standard matrix algebra (see eq 3). Equation 5 is used as fitness function to evaluate the quality of the fit between the DFT charges and the EEM charges. This fitness function is then minimized by updating the χ^* and η^* values by means of a combination of a local and global optimizer. First many randomly chosen sets of parameters were submitted to the simplex method in multidimensions.[18] Each time 10000 sets were submitted to the simplex method, the 100 best unique sets (i.e., those sets that give the lowest value for the fitness function) were used as input in a genetic algorithm. The new sets of parameters found by mating of the parent sets or originating from mutations are locally optimized in the simplex method prior to using them as parents in the next generation. This combination of a global minimization technique and a local optimizer proved the best approach for the very cumbersome calibration. The parameter set giving the best fit between the EEM and DFT charges is recorded and considered to represent the optimal set of calibrated parameters.

Results and Discussion

Calibrated Parameters. The calibration of the parameters proved to be a highly cumbersome task. This is mainly due to the sensitivity of the fitness function for the parameters, as well as to the existence of a very large number of local minima. In Table 1, the calibrated effective atomic electronegativities and hardness parameters calibrated in the procedure described above are presented, together with the original parameter-values obtained by Mortier.[16−17] First of all, it is seen that the Mortier parameters show large discrepancies with respect to our values. At first glance, it may look as this may be due to the fact that Mortier et al. constrained the effective electronegativity of oxygen to 8.5 eV (0.3124 au). However, the effective electronegativity parameters for all elements can be calibrated only up to a certain constant. Consequently, a constant may be added to all effective electronegativities, without influencing the resulting charges as it will merely influence the equalized molecular electronegativity value (see eq 3 adding a constant to all χ^* only has an effect on the equalized molecular χ_{eq}). Even if a common value for the effective electronegativity of hydrogen is chosen, one is still confronted with the same differences. The following factors are expected to play a role in these observations: (i) Mortier et al. used HF/STO-3G calculations to obtain charges to calibrate effective electronegativities and hardnesses, whereas in the present study, B3LYP/6-31G* calculations are used, (ii) a different fitness function is used, and (iii) a different training set is used to calibrate the parameters.

TABLE 1: Optimized Values of χ^* and η^*, as Compared to the Parameters from the Work of Mortier et al.[16−17a]

atom		present study	Mortier et al.
H	χ^*	1.00	4.4
	η^*	17.95	13.8
	N	930	65
C	χ^*	5.25	5.7
	η^*	9.00	9.1
	N	602	19
N	χ^*	8.80	10.6
	η^*	9.39	13.2
	N	105	1
O	χ^*	14.72	8.5
	η^*	14.34	11.1
	N	101	26
F	χ^*	15.00	not calibrated
	η^*	19.77	not calibrated
	N	65	not calibrated

[a] All values are in eV. N is the number of atoms of a particular element in the training set of 138 molecules.

TABLE 2: Optimized χ^* Values (This Work and Mortier[16,17]) as Compared to Electronegativity Values from Other Scales (All Values in eV, Unless Indicated Otherwise, n/a = Not Available)

scale	H	C	N	O	F
this work	1.00	5.25	8.80	14.72	15.00
Mortier[16−17] (EEM-based)	4.4	5.7	10.6	8.5	n/a
van Duin et al.[19] (EEM-based)	5.96	sp^3 8.58 sp^2 8.24	n/a	n/a	n/a
Njo et al.[20] (EEM-based)	3.12	4.30	7.56	8.50	n/a
Sanderson (Pauling units)[1−2]	2.31	2.47	2.93	3.46	3.92
Boyd, Edgcombe (Pauling units)[21]	n/a	2.6	3.08	3.62	4.00
Pauling, Allred (Pauling units)[22−23]	2.20	2.55	3.04	3.44	3.98
Gordy (Pauling units)[24]	2.17	2.52	3.01	3.47	3.94
Allred, Rochow[25]	2.20	2.50	3.07	3.50	4.10
Lackner, Zweig[27] (Mulliken ground-state version)	7.17	6.24	6.96	7.59	10.40
Bergman, Hinze[28] (Mulliken valence state version)	7.18	8.07	9.68	11.91	13.66
Allen (Pauling units)[29]	2.30	2.54	3.06	3.61	4.19

Table 2 presents a number of values from different well-established electronegativity scales. Included are scales based on previous EEM calibrations[16−17,19−20] and scales based on atomic electron densities,[1−2] topological properties of electron density distributions of molecules,[21] bond dissociation (BD) energies[22] (introduced by Pauling[23]), the electrostatic potential[24] and force as experienced by the valence electrons,[25] the ionization potential (IP) and electron affinity (EA)[26] as ground-state[27] and valence-state properties,[28] and the average energy of valence shell s and p electrons (free-atom spectroscopic electronegativity).[29]

We focus on the trends observed in the different classes, rather than on the quantitative numbers. In general, the same trend is observed in almost all of the electronegativity scales, including our scheme. When comparing values from the EEM calibrations by Mortier et al.[16,17] and the developers of the DMM force field,[19−20] a number of clear differences are found. These are in part due to the fact that the ab initio charges were obtained using different quantum chemical methods and basis sets and the fact that different calibration sets were used. This makes a quantitative comparison impossible, but a striking discrepancy between the parameters sets given by Mortier et al. and ours lies in the inverted sequence for the electronegativity and hardness for N and O. Njo et al. used HF/STO-3G calculations similar to those by Mortier, and found no such inversion. Equation 2 shows that if only a single atom of a specific element is present in the calibration set any change in the effective electronegativity (denoted $\Delta\chi^*$) can perfectly be compensated by an opposite change in the effective hardness (the magnitude of the change in η^* is then $-\Delta\chi^*/q_\alpha$). This is the case for N in the calibration by Mortier et al. With an increasing number of valencies for the nitrogen atoms, such a compensation gives rise to a poorer fitness function, making us reject that set of parameters.

Concerning the other, non-EEM-based scales in Table 2, a relatively high value is found for hydrogen in the Mulliken scale (based on ionization potentials and electron affinities), especially in the ground-state version.[27] This might be due to the fact that hydrogen is a special case having only one electron in its neutral form. To identify the electronic chemical potential with Mulliken's electronegativity defined as (IP + EA)/2, one has to assume a quadratic relation between the energy, *E*, and the number of electrons, *n*, in combination with a finite difference approximation to calculate the second derivative of *E* with respect to *n*.[12] This might be too crude an approximation in the (special) case of hydrogen.

When comparing electronegativity values determined from different schemes, it is important to note that there is a difference between ground state and valence state values. Ground-state models are determined for the isolated atom, whereas in a valence state, one considers the atom in a molecule. The valence state of the atom in the molecule corresponds to the state of the atom when taken out of the molecule, without reorganization of its electronic charge distribution. The latter case represents a perturbed situation, because the electron cloud of the atom has adapted itself to the rest of the molecule. In the present calibration, χ^* and η^* values are determined, which are values for valence states. Although it might seem better justified to compare our values to previously reported valence state values, such a comparison is still not straightforward. Values may differ between different valence states for the same atom. Depending on the range of valence states included in the set of molecules used to calibrate χ^* and η^* values, different numbers could be found. We have chosen in the first instance not to distinguish between different valence states or chemical surroundings (meaning taking into account the nature of the bonds the atom is in). In doing so, the additional work of having the computations spend time in trying to identify hybridization or bond connectivity can be avoided, allowing a faster algorithm. Whether it is justifiable not to distinguish between different valencies will be examined below. It is important, however, to note that the trend in electronegativity values from the current calibration agrees with that in most other scales.

The calibrated parameters for the hardness are given in Table 3, together with values based on other EEM calibrations[16−17,19−20] and some experimental values calculated from the ionization potentials and electron affinities for ground states and valence states. In terms of DFT, the hardness is defined as the second derivative of the energy, *E*, with respect to the number of electrons, *n*, the curvature of the *E*(*n*) curve at the point n_0, where n_0 is the number of electrons for the neutral system. Table 3 illustrates again the differences, even in trends, that are found when considering either ground states or valence states. From Table 3, it is found that the present calibration gives nitrogen a lower hardness compared to oxygen. This illustrates again the importance of having the calibration set spanning a wide range of valencies. The larger electronegativity given by Mortier et al. for nitrogen compared to ours should give rise to a larger effective hardness in the scale by Mortier et al. compared to ours. As noted above, the increase of the effective hardness needed to compensate a larger effective electronegativity is $-\Delta\chi^*/q_\alpha$, where the nitrogen atom in the set by Mortier et al. has a negative charge.

TABLE 3: Optimized η^* Values (This Work and Mortier[16,17]) Compared to Other EEM Based Values and Available Experimental Values (All Values in eV, n/a = Value Not Available)

scale	H	C	N	O	F
this work	17.95	9.00	9.39	14.34	19.77
Mortier[16−17] (EEM-based)	13.8	9.1	13.2	11.1	n/a
van Duin et al.[19] (EEM-based)	15.63	sp^3 13.77 sp^2 12.48	n/a	n/a	n/a
Njo et al.[20] (EEM-based)	11.99	9.69	11.69	13.66	n/a
Lackner, Zweig[27] (Mulliken ground-state version)	6.45	4.99	7.59	6.14	7.07
Bergman, Hinze[28] (Mulliken valence state version)	6.42	6.35	7.29	8.26	9.19

Quality of the EEM Charges. Figure 1a−e gives the EEM charges (using the optimal effective electronegativity and hardness values) versus the DFT charges. Also included are the parameters for the best fitting linear function between both types of charges. There is an obvious linear correlation between both sets of charges. This illustrates the ability of the principle of electronegativity equalization to yield quantitative atomic charges. It may be argued that for fluorine, and to some extent for oxygen as well, the agreement is much less pronounced. This is due to the more limited range of charge values for the F and O atoms. This is clear from the average absolute differences between the DFT and the EEM charges. For H, C, N, O, and F, these are respectively 0.02, 0.06, 0.06, 0.04, and 0.02, showing that the accuracy of the EEM method is similar for all elements.

A few remarks should be made concerning the spread of charges in Figure 1. For the hydrogen atoms in the training set, two regions exhibiting a higher density of points in the plot are found. This is simply a consequence of the chemistry of the training set. The high number of points in with lower charges are mainly associated with hydrogen atoms bound to carbon. Two regions with a relatively high point density occur at higher positive charges. These are mainly associated with hydrogen atoms bound to nitrogen and oxygen. One could argue that distinguishing between the different connectivities of hydrogen atoms, each with their own hydrogen effective electronegativity and hardness, would further improve the fit. Mortier et al. followed this path for hydrogen, where they calibrated different parameters for positively charged hydrogen atoms and hydride atoms. In the case of other atoms, Mortier also notes the "surprisingly" good ability of a single value for the effective electronegativity and hardness to yield high quality charges in many different chemical surroundings.[30] The charge pattern of nitrogen also shows a number of outliers. These are mainly associated with nitro compounds. In these cases, nitrogen is

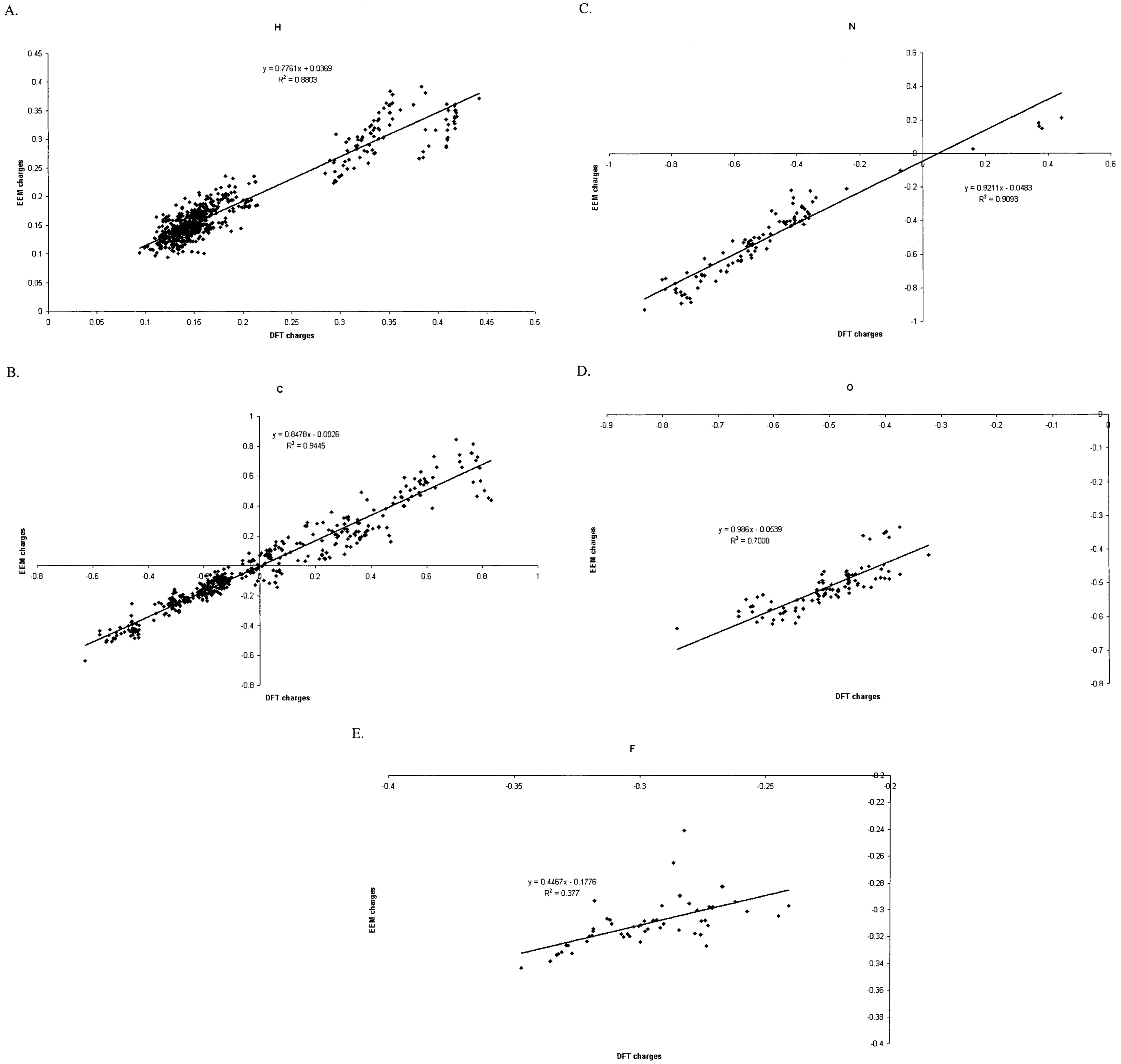

Figure 1. Comparison of EEM and DFT Mulliken point charges for all atoms in the training set (separate correlations are shown for H, C, N, O, and F). EEM charges were obtained using the parameters calibrated in this work and listed in Table 1.

positively charged according to the DFT calculations, which agrees with what is expected from simple Lewis structures and formal charges. It is worth mentioning that the present EEM calibration succeeds at giving positive charges for the nitrogen atoms in the nitro compounds. Previous calibrations did not succeed in this.[31] This may also indicate that the necessity of Mortier et al. to distinguish different parameters for positively and negatively charged hydrogen atoms might equally be due to incomplete calibrations. The present medicinal chemistry biased calibration set does, however, not contain hydride atoms.

The fact that EEM quantitatively predicts atomic charges for very different valencies, using the same parameters, shows that the calibrated parameters are applicable within a very large range of valencies. This justifies the decision not to distinguish between different valencies of atoms of the same element, thereby saving a lot of additional computational work to identify which "type" of atoms of each element are present in each molecule.

Applicability of EEM. In the section above, the EEM and DFT charges for the molecules contained in the calibration set were compared showing fine agreement. However, to test the applicability of EEM, one needs to consider molecules that are not part of the calibration set. To that end, a number of medicinal molecules, more precisely neuroleptics, were used for comparison. Structures were taken from Tollenaere et al.[32] and are shown in Figure 2. Molecules used in this study were: fluanisone, benperidol, pipamperone, spirilene, oxypertine, and azaperone.

Because the aim of this study is not a full conformational study of these six molecules, only a relatively limited conformational analysis of the molecules was performed, using the MMFF[33−37] force field and AM1 calculations. Both the global minimum from the MMFF conformational analysis and that from the AM1 conformational analysis were further optimized on the B3LYP/6-31G* level, and Mulliken charges were calculated at that level. Using the calibrated effective electrone-

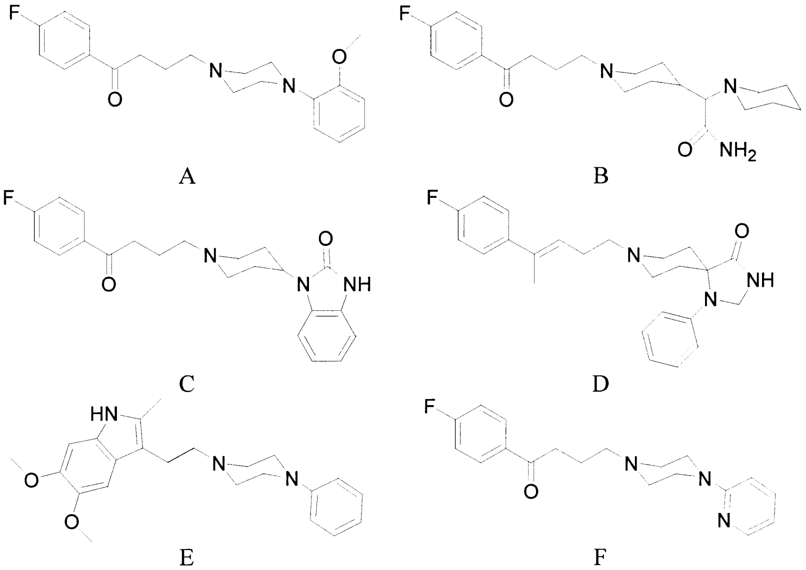

Figure 2. Chemical structures of the neuroleptics used as an EEM validation set. A = fluanisone, B = benperidol, C = pipamperone, D = spirilene, E = oxypertine, and F = azaperone.

TABLE 4: Calibrated Effective Electronegativity and Hardness Parameters Using the Original Calibration Set, and a Reduced Calibration Set Designed for the Study of Neuroleptics (All Values in eV)[a]

atom		original set	reduced set
H	χ^*	1.00	1.00
	η^*	17.95	15.73
	N	930	53
	dev.	0.01	0.01
C	χ^*	5.25	5.06
	η^*	9.00	9.03
	N	602	45
	dev.	0.06	0.06
N	χ^*	8.80	13.72
	η^*	9.39	13.97
	N	105	5
	dev.	0.07	0.04
O	χ^*	14.72	14.25
	η^*	14.34	14.27
	N	101	3
	dev.	0.03	0.03
F	χ^*	15.00	16.39
	η^*	19.77	24.79
	N	65	2
	dev.	0.04	0.00

[a] Dev. describes the average absolute error between the DFT and EEM charges for both calibration sets. The column *N* gives the number of atoms of each element in both calibration sets.

gativity and hardness parameters from Table 1, the EEM charges were calculated. The speed of the EEM calculation of atomic charges is illustrated by the fact that the calculation of atomic charges in all 12 structures takes less than 50 ms on a current personal computer. In Table 4, the average absolute differences between the DFT charges and EEM charges is given. The results clearly show that although the calibration set used does not hold any of these neuroleptics, or fragments thereof, the charges predicted by EEM agree very well with the DFT charges. Taking into account the speed of the EEM calculations, it may be concluded that EEM performs in an excellent way, costing several orders of magnitude less computational effort than the B3LYP/6-31G* calculations. Mulliken population analysis requires the SCF procedure to be carried out, which takes, for example, one single neuroleptic molecule 32 min on a PC. For the 12 molecules together, the EEM calculations take less than 50 ms, illustrating the impressive speed of the EEM approach.

If an even better agreement is desired, one might choose to calibrate parameters specifically for a certain class of molecules, in this case neuroleptics. The parameters were therefore calibrated again, using as a calibration set one structure of both fluanisone and spirilene. Both the original and new parameters are given in Table 4. The table shows how the difference between the EEM and DFT charges is reduced when using the parameter sets calibrated from the fluanisone/spirilene set. However, the improvement is only small. Given the amount of additional work required to recalibrate the parameters and the rather limited improvement of the agreement between the DFT and EEM charges, such recalibration is unnecessary. Concerning the values for both the effective electronegativity and hardness for the calibrations with the two sets, some clear differences are found. This shows that the calibration set is a bias. Also, for a smaller calibration set, certainly a specific one aimed at only a limited range of valencies for all elements, the above-described electronegativity−hardness compensation may play a role. The parameters obtained from the smaller calibration set therefore have less chemical relevance as those obtained from the larger calibration set.

In deriving the EEM equations (as, e.g., eq 2), the parameters χ^* and η^* clearly have a firm chemical meaning representing the effective electronegativity and hardness, which can be considered as valence state type atoms in molecules properties. When the parameters are calibrated, solutions may be found that are mathematically equivalent but chemically are highly counterintuitive (e.g., when only a limited range of valencies for a certain element is considered). The chemical meaning of the calibrated parameters allows the discrimination between chemically sound and less meaningful solutions, even if these are mathematically equivalent.

Extension to New Atom Types. As noted above, calibration of the parameters proved not to be a simple task and required much effort to minimize the fitness function (5). It is clear that when new atoms will be included, the dimensionality in the calibration will increase, resulting in an even more cumbersome calibration. It would therefore be very useful if extension of the EEM method to include new atom types would only require calibration of the new atom types, retaining the values previously calibrated for the other atom types. In an attempt to establish whether such an approach is feasible, independent calibrations

TABLE 5: Changes in Calibrated Parameters (in eV, for Convenience, All Effective Electronegativity Values Were Standardized to $\chi^* = 1.00$ for Hydrogen) upon Extending the Range of Elements and Valencies of Atoms in the Calibration Sets[a]

atom		CH	CHN	CHO	CHNO	CHNOF
	M	14	38	34	90	138
H	χ^*	1.00	1.00	1.00	1.00	1.00
	η^*	15.83	19.01	18.06	17.95	17.95
	N	98	276	210	643	930
C	χ^*	5.22	5.99	5.88	5.58	5.26
	η^*	9.32	9.35	9.32	9.17	9.00
	N	50	155	104	393	602
N	χ^*		10.47		9.56	8.80
	η^*		10.12		9.79	9.39
	N		34		83	105
O	χ^*			14.04	15.91	14.72
	η^*			13.54	15.01	14.34
	N			24	88	101
F	χ^*					15.00
	η^*					19.77
	N					65

[a] N denotes the number of atoms of each element in the each set and M denotes the number of molecules in each set.

for different sets of molecules were performed. The sets differ among each other by the number of atom types included in the calibration set. Sets were constructed comprising only C and H (set 1); C, H, and O (set 2); C, H, and N (set 3); C, H, N, and O (set 4); and ultimately C, H, N, O, and F (set 5). These sets were derived from the full calibration set by keeping only those molecules that only contain atoms of the types retained in the reduced calibration sets.

It was found that keeping the previously calibrated parameters fixed, resulted in a very poor calibration. This is not surprising, given the sensitivity of the fitness function on the parameters. On the other hand, inclusion of a new atom type did not need the assigning of new random values for the previously calibrated parameters, and thereby augmenting the calibration dimensionality. The most efficient approach turned out to be the use of the previously determined parameters, supplemented with a random choice of values for the new atom type(s) and minimization of the entire set of parameters in the simplex procedure or genetic algorithm. Thus, instead of having both the random assignment step and minimization procedure working in $(2X + 1)$ dimensions for X atom types ($2X$ parameters for X atom types plus the fitness function), extension of the method toward a new atom type will result in 2 additional dimensions in the random step, followed by $(2(X + 1) + 1)$ dimensions in the minimization procedure. This is of course an enormous reduction of the computing efforts, because if, e.g., 10 random values for each new parameter are chosen, the number of resulting sets in the calibration can be reduced from $10^{2(X+1)}$ to 10^2 for the addition of a single new atom type, or for addition of Y new atom types from $10^{2(X+Y)}$ to 10^{2Y}. That the previously determined values need to be reoptimized through the minimization procedure is the consequence of, on one hand, the large sensitivity of the fitness function on the parameter values, and on the other hand, the necessity of letting the parameters adapt to new types of chemical surroundings. Table 5 gives the optimal values for both the effective electronegativity and hardness when the calibration sets are extended to include new atom types. When comparing small calibration sets, large differences in the calibrated parameters may be found, but when moving to sets with an increasing number of elements, the differences grow smaller.

Following the previous discussion on the applicability of EEM parameters for a specific element over a wide range of chemical situations (e.g., different hybridization states), it is clear that the previous paragraph and method for extension of the EEM calibration space is also valid for instances where new, separate parameters need to be calibrated for a different valencies of an atom. At present, however, no situations have been found where the predicted charges show such poor agreement with the DFT charges that separate parameters need to be calibrated for specific valencies.

Conclusions

Calibration of the effective electronegativity and hardness parameters for C, H, N, O, and F was performed using a large set of molecules representing a wide range of functional groups in medicinal chemistry. It was found that, using these parameters, one can calculate atomic charges in molecules at a very impressive speed, without important loss in accuracy compared to the DFT calculated charges. The resulting calibrated parameters were also found to be applicable over a very wide range of valencies. As such, EEM is a promising method for the very fast calculation of atomic charges, showing ab initio quality without the cost of such calculations.

Indications are that the calibration of EEM parameters can be extended to include new elements, without the need to repeat the entire calibration process. Previously determined values for the effective electronegativity and hardness can be retained, and merely need to be slightly reoptimized, thereby opening the way to efficiently extending the number of elements that can be covered in EEM.

The calibrated parameters were found to depend on the calibration set used. Only parameters based on calibration sets holding wide ranges of valencies for every element should therefore be considered as chemically relevant. Such effective electronegativity and hardness parameters were found to exhibit the same trends as other well-established scales. The calibrated parameters also depend on what charges have been used in the calibration, including the quantum chemical method used, basis set, etc.

Acknowledgment. P.B. wishes to thank Janssen Research Foundation and 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.

Supporting Information Available: The list of molecules used in the calibration. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Sanderson, R. T. *Science* **1951**, *114*, 670.
(2) Sanderson, R. T. *Polar Covalence*; Academic Press: New York, 1983.
(3) Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219.
(4) Rappé, A. K.; Goddard, W. A., III. *J. Phys. Chem.* **1991**, *95*, 3358.
(5) Yang, Z.-Z.; Wang, C.-S. *J. Phys. Chem. A* **1997**, *101*, 6315.
(6) York, D. M.; Yang, W. *J. Chem. Phys.* **1996**, *104*, 159.
(7) Itzkowitz, P.; Berkowitz, M. L. *J. Phys. Chem. A* **1997**, *101*, 5687.
(8) Mortier, W. J.; Ghosh, S. K.; Shankar, S. *J. Am. Chem. Soc.* **1986**, *108*, 4315.
(9) Parr, R. G.; Donnelly, R. A.; Levy, M.; Palke, W. E. *J. Chem. Phys.* **1978**, *68*, 3801.
(10) Parr, R. G.; Pearson, R. G. *J. Am. Chem. Soc.* **1983**, *105*, 7512.
(11) Politzer, P.; Weinstein, H. *J. Chem. Phys.* **1979**, *71*, 4218.
(12) Parr, R. G.; Yang, W. *Density Functional Theory of Atoms and Molecules*, Oxford University Press: New York, 1989.

(13) Geerlings, P.; De Proft, F.; Langenaeker, W. *Chem. Rev.* **2002**, submitted for publication.

(14) 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.

(15) Bultinck, P.; Langenaeker, W.; Lahorte, P.; De Proft, F.; Geerlings, P.; Van Alsenoy, C.; Tollenaere, J. P. *J. Phys. Chem. A* **2002**, *106*, 7895.

(16) Van Genechten, K. A.; Mortier, W. J.; Geerlings, P. *J. Chem. Phys.* **1987**, *86*, 5063.

(17) Uytterhoeven, L.; Lievens, J.; Van Genechten, K.; Mortier, W. J.; Geerlings, P. Preprints Conf. Eberswalde (GDR), 1987.

(18) 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.

(19) van Duin, A. C. T.; Baas, J. M. A.; van de Graaf B. *J. Chem. Soc., Faraday Trans.* **1994**, *90*, 2881.

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

(21) Boyd, R. J.; Edgecombe, K. E. *J. Am. Chem. Soc.* **1988**, *110*, 4182.

(22) Allred, A. L. *J. Inorg. Nucl. Chem.* **1961**, *17*, 215.

(23) Pauling, L. *J. Am. Chem. Soc.* **1932**, *54*, 3570.

(24) Gordy, W. *Phys. Rev.* **1946**, *69*, 604.

(25) Allred, A. L.; Rochow, E. G. *J. Inorg. Nucl. Chem.* **1958**, *5*, 264.

(26) Mulliken, *J. Chem. Phys.* **1934**, *2*, 782.

(27) Lackner, K. S.; Zweig, G. *Phys. Rev. D* **1983**, *28*, 1671.

(28) Bergman; Hinze, J. *Angew. Chem., Int. Ed.* **1996**, *35*, 150.

(29) Allen, L. C. *J. Am. Chem. Soc.* **1989**, *111*, 9003.

(30) Mortier, W. J. *Academiae analecta, Mededelingen van de Koninklijke Academie voor Wetenschappen, Letteren en Schone Kunsten*, **1990**, *52*, 29.

(31) Lahorte, P. *Charge Polarization in Molecular Mechanics Calculations*, M.Sc. Thesis (unpublished), Ghent University, 1996

(32) Tollenaere, J. P.; Moereels, H.; Raymaekers, L. A. *Atlas of the three-dimensional structure of drugs*; Elsevier/North-Holland Biomedical Press: Amsterdam, 1979.

(33) Halgren, T. A. *J. Comput. Chem*. **1996**, *17*, 490.

(34) Halgren, T. A. *J. Comput. Chem*. **1996**, *17*, 520.

(35) Halgren, T. A. *J. Comput. Chem*. **1996**, *17*, 553.

(36) Halgren, T. A. *J. Comput. Chem*. **1996**, *17*, 587.

(37) Halgren, T. A. *J. Comput. Chem*. **1996**, *17*, 616.