High-Speed Calculation of AIM Charges through the Electronegativity Equalization Method

P. Bultinck,*,† R. Vanholme,† P. L. A. Popelier,‡ F. De Proft,§ and P. Geerlings§

Department of Inorganic and Physical Chemistry, Ghent University, Krijgslaan 281 (S-3), B-9000 Gent, Belgium, School of Chemistry, University of Manchester, Manchester M60 1QD, Great Britain, and Eenheid Algemene Chemie (ALGC), Free University of Brussels (VUB), Pleinlaan 2, B-1050 Brussels, Belgium

Received: July 12, 2004; In Final Form: September 6, 2004

The applicability of the electronegativity equalization method (EEM) is investigated for the calculation of atoms-in-molecules charges in prototypical organic molecules. A large training set of molecules was composed, comprising H, C, N, O, and F. 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 least-squares fashion. The quality of the EEM charges is assessed by comparison with B3LYP/6-31G* charges calculated for a test set of amino acids and the neuroleptic fluanisone, not contained in the training set. The EEM approach is found to be a very powerful way to obtain AIM charges without the computational cost of the ab initio approach.

Introduction

One of the most often used concepts in applied quantum chemistry is atomic charges, e.g., for the interpretation of molecular reactivity and charge transfer. Despite the great value of atomic charges in chemistry, there is no unique way to obtain an atomic charge. Thus, many different definitions or algorithms to obtain atomic charges have been proposed. Examples are the basis function based methods such as Mulliken[1] and Löwdin[2] population analysis, electrostatic potential derived charges such as MKS[3,4] and CHELPG,[5] NBO[6] related methods such as natural population analysis (NPA), and methods based on the comparison of promolecular density with the ab initio molecular electron density as in Hirshfeld[7] population analysis. Central to this study is the atoms-in−molecules (AIM) method, introduced by Bader and co-workers.[8−9] The AIM theory, which is rooted in quantum mechanics, generates a variety of atomic properties, among which are atomic charges. This puts AIM in a position to propose an electronegativity scale, as demonstrated in this paper.

The AIM method identifies atoms as regions in real 3D space, bordered by surfaces of zero-flux in the field of the electron density gradient. These regions, called (atomic) basins, obey their own virial theorem, just like the total system. Atomic basins are spanned by gradient paths terminating at a single point, which is called an attractor. When an attractor coincides with a nucleus, the basin and the nucleus together form the atom in the molecule. Atomic basins are denoted as Ω_α^A, meaning the basin of atom α in molecule A. The atomic charge q_α^A for atom α in molecule A is then simply given by[10]

$$q_\alpha^A = Z_\alpha - \int_{\Omega_\alpha^A} \rho^A(\mathbf{r})\, d\mathbf{r} \qquad (1)$$

where the integration of the molecular electron density $\rho^A(\mathbf{r})$ is performed over the atomic basin Ω_α^A and Z_α is the atomic number of the atom α.

Bader and co-workers took Schwinger's quantum action principle as a natural starting point to generalize molecular quantum mechanics to subspace quantum mechanics, the subspace being the AIM atom.[11] This starting point is powerful because from a single dynamical principle (i.e., Schwinger's principle) one can not only define the observables but also obtain their equations of motion and the commutation relations. As a result, one is in a excellent position to define expectation values of atomic properties and derive[12] the so-called atomic theorems, such as the atomic force, virial, continuity, torque, and power theorem. Bader and Zou[13] argued that this approach could be applied to define an atomic population as the expectation value of a quantum observable, being the number operator $\hat{N}$.

A peculiar feature of AIM atoms is that they are not infinite in size. In other words, atoms are confined by sharp boundaries. They do not extend to infinity, except in some radial directions for atoms near the surface of a molecule or crystal. The delineation of atomic basins requires considerable computing time, despite algorithmic advances[14] and hardware improvements. This fact and the lack of analytical integrals make the calculation of the volume integral in eq 1 demanding in CPU time. As a result there is a need to develop a strategy for the fast calculation of AIM charges. Atomic charges are well-known to be very valuable quantities for the interpretation of many chemical observations, such as chemical reactivity[15] and as molecular descriptors in QSAR.[16,17] More specifically, AIM charges have been used in QSPR studies[18,19] and for the calculation of relative pK_a values in solution[20] and for the characterization of atoms and fragments in biomolecules.[21−24]

The electronegativity equalization principle formulated by Sanderson[25−26] states that when molecules are formed, the electronegativities of the constituent atoms become equal, yielding the molecular equalized (Sanderson) electronegativity. The present study is based on the electronegativity equalization method (EEM) of Mortier et al.[27] In this method, the electronegativity of an atom α in an N-atom molecule is shown to be

* To whom correspondence should be addressed. E-mail: Patrick.Bultinck@UGent.be. Fax: +32/9/264.49.83.
† Ghent University.
‡ University of Manchester.
§ Free University of Brussels (VUB).

10.1021/jp046928l CCC: $27.50
Published on Web 10/22/2004

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

In this expression, χ_α^* and η_α^* represent the effective atomic electronegativity and hardness respectively, q_α represents the atomic charge on atom α, and $R_{\alpha\beta}$ represents the interatomic distance between atoms α and β. χ_α^* and η_α^* are based on the isolated atom values with added corrections to account for the incorporation of the atom in a molecule (or crystal). For the actual derivations of the EEM formulas, the reader is referred to Mortier et al.[27]

As has been demonstrated on several occasions, EEM holds the potential of generating, at a very modest computational cost, atomic charges that are both connectivity- and geometry-dependent. Bultinck et al. previously described the application of the present EEM scheme to compute atomic charges of different kinds in organic molecules.[28−29] 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 ($\chi_{eq} = \chi_1 = \chi_2 = \cdots = \chi_N$), 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

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

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.[30−35] A recent overview of the status and theory of EEM within the broad context of conceptual DFT can be found in Geerlings et al.[31] Chattaraj et al. recently published a review of reactivity descriptors from DFT, including those related to or obtainable from EEM.[15]

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. These parameters are unknown, so they must be calibrated from a set of known charges. In the present study, we will report the calibration based on a large set of small organic molecules with known AIM charges. The aim of the study is to establish whether EEM can be used to reproduce and predict AIM charges at high speed.

Computational Methods

To test the EEM approach to AIM charges, the same training set was used as in previous studies where the applicability of Mulliken, natural population analysis (NPA), CHELPG, and MKS electrostatic potential derived charges and Hirshfeld charges was tested.[28−29] Not all molecules could be retained, due to the fact that the calculation of AIM charges proved troublesome in some cases. From eq 1, it is clear that the integration has to be performed numerically over the appropriate domain. This may cause errors, and in some cases, the summed atomic populations deviated too much from the total number of electrons in the molecule. This caused us to not include these molecules. The eventual training set is limited to neutral molecules containing H, C, N, O, and F and included 942 atoms, holding 495 H, 316 C, 59 N, 44 O, and 28 F atoms. The calibration set holds a wide variety of different chemical environments for the different elements, inspired on medicinal chemistry. The set of molecules used in the present study is available as Supporting Information.

The geometries of all molecules were optimized at the B3LYP/6-31G* level, using the Gaussian 98 program.[36] AIM charges were calculated at the same level using the Morphy program.[37−38] The ab initio AIM charges calculated at that level will be denoted DFT-AIM. Molecules were considered in equilibrium geometries, although in principle EEM can be applied also to molecules in nonequilibrium geometries such as geometries in crystallographic coordinates.

The calibration of the parameters was carried out largely using the same approach and techniques as reported previously.[28−29] This involves the optimization of the 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 10 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 was evaluated in a least squares way, minimizing

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

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 ($N_{el} = 5$ in this case), M is the number of molecules, and $N_{i,z}$ is the number of atoms of element z in molecule i. The total number of atoms of element z over all molecules is denoted N_z.

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. Equation 4 is used as a 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. The techniques used for this optimization were also described earlier[28−29] and involve a Lamarckian genetic algorithm. This technique was found to be quite efficient, since in every creation of a new generation in the genetic algorithm, a local optimization step is introduced. Menegon et al. used a more limited method in a study on the calibration of CM1 charges.[39] There the genetic algorithm and simplex method are separate steps. First the genetic algorithm is run, after which the simplex method is used to optimize the best fitting parameter set. In the present study, it was found that this approach is more likely to end up in local minima. This is due to the high sensitivity of the fitness function with respect to the calibrated parameters.

The parameter set giving the best fit between the EEM-AIM and DFT-AIM charges is recorded and considered to represent the optimal set of calibrated parameters.

TABLE 1: Optimized Values of χ^* and η^*, as Compared to Previously Calibrated Parameters[a]

atom		present study	Mortier et al.[40−41]	Van Duin et al.[42]	Njo et al.[43]	Bultinck et al. Mulliken[28]	Bultinck et al. NPA[29]	Menegon et al.[39]
H	χ^*	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	η^*	20.57	13.8	15.63	11.99	17.95	19.44	15.52
	N	495	65	n/a	n/a	930	930	n/a
C	χ^*	1.77	2.30	3.62 (sp^3)−3.28 (sp^2)	2.18	5.25	8.49	1.78
	η^*	9.24	9.1	13.77 (sp^3)−12.48 (sp^2)	9.69	9.00	9.15	11.38
	N	316	19	n/a	n/a	602	602	n/a
N	χ^*	14.80	7.20	n/a	5.44	8.80	13.45	3.90
	η^*	10.74	13.2	n/a	11.69	9.39	10.64	10.55
	N	59	1	n/a	n/a	105	105	n/a
O	χ^*	29.46	5.10	n/a	6.38	14.72	27.06	7.18
	η^*	15.85	11.1	n/a	13.66	14.34	19.63	17.81
	N	44	26	n/a	n/a	101	101	n/a
F	χ^*	62.84	n/a	n/a	n/a	15.00	39.18	n/a
	η^*	46.05	n/a	n/a	n/a	19.77	44.10	n/a
	N	28	n/a	n/a	n/a	65	65	n/a

[a] All values are in eV. N is the number of atoms of a particular element in the training set used. n/a = not available.

Results and Discussion

Calibrated Parameters. The calibration of the parameters in EEM is a rather difficult 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,[40−41] van Duin et al.,[42] Njo et al.,[43] and Menegon et al.[39] and the values obtained by Bultinck et al. for the Mulliken and NPA charge schemes.[28−29] First of all, it is seen that the Mortier parameters show large discrepancies with respect to our values. At first glance, this appears to be due to the fact that Mortier et al. constrained the effective electronegativity of oxygen. 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 2, adding a constant to all χ^* only has an effect on the equalized molecular electronegativity χ_{eq}). Even if, as in Table 1, a common value for the effective electronegativity of hydrogen is chosen, one is still confronted with differences. The following factors are expected to play a role in these observations: (i) this study uses different kinds of atomic charges compared to those used by Mortier et al. and the previous calibrations by Bultinck et al. (ii) 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, (iii) a different fitness function is used (iv) as well as a different training set to calibrate the parameters. Similar conclusions can be drawn for the comparison with the other EEM based scales given in Table 1. On the qualitative level, the same trend is observed in almost all of the electronegativity scales, including our scheme, except for the inverted sequence for the electronegativity and hardness for N and O in the study by Mortier et al. Njo et al.[43] used HF/STO-3G calculations similar to those by Mortier et al. and found no such inversion. Equation 2 shows that if only a single atom of a specific element is present in the calibration set there is a clear over-parametrization so no real significance can be attributed to the parameters of this element. This is the case for N in the calibration by Mortier et al. An equally good set of parameters for nitrogen, fulfilling the expected trends, can be obtained by simple algebraic manipulation of the parameters.

The calibrated parameters for the hardness are also given in Table 1, together with values based on other EEM calibrations.[28−29,39−43] 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.[30−31] Table 1 illustrates again the differences, even in trends, that are found when considering either ground states or valence states. From Table 1, 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.[40,41] for nitrogen compared to ours can be shown to induce a larger effective hardness in the scale by Mortier et al. compared to ours. The higher effective hardness is needed to compensate for the larger effective electronegativity. Recently, Menegon et al.[39] found that when calibrating atomic hardness carbon exhibits a higher hardness than nitrogen. This is contradictory to our results and to the other scales in Table 1, as well as to experimental data.[44−45] Again, several reasons may be identified, among which the Hamiltonian used (PM3 versus B3LYP/6-31G* in the present study), the charge model (CM1 versus AIM), and the use of the Klopman−Ohno−Mataga−Nishimoto (KOMN) approximation.[46−48] As will be shown below, it was found that using this latter interaction term did not yield such an inversion in the present study. Clearly, there are many possible effects that can influence the results, including also the optimization method used. Another difference may lie in the fact that Menegon et al.[39] did not consider aromatic or conjugated systems, containing, on average, softer carbon atoms. In the current calibration set, a wide selection of different valencies is included for carbon, including carbon atoms in aromatic and conjugated systems.

The most comparable sets of parameters are those obtained previously by Bultinck et al. for different atomic charges.[28−29] It was shown by these authors that Mulliken and NPA charges were by far the atomic charge definitions that could most accurately be obtained from EEM. Moreover, they were obtained using exactly the same methodology as used in the present study and nearly exactly the same training set of molecules. For C, H, N, and O, one notices that the parameters obtained from AIM calibration follow the same trends as the Mulliken and NPA based ones.[28−29] A remarkable feature is that the electronegativity of the carbon atom is not much higher than that of hydrogen. In the NPA and Mulliken based schemes, this difference was much more outspoken.[28−29] Another effect is

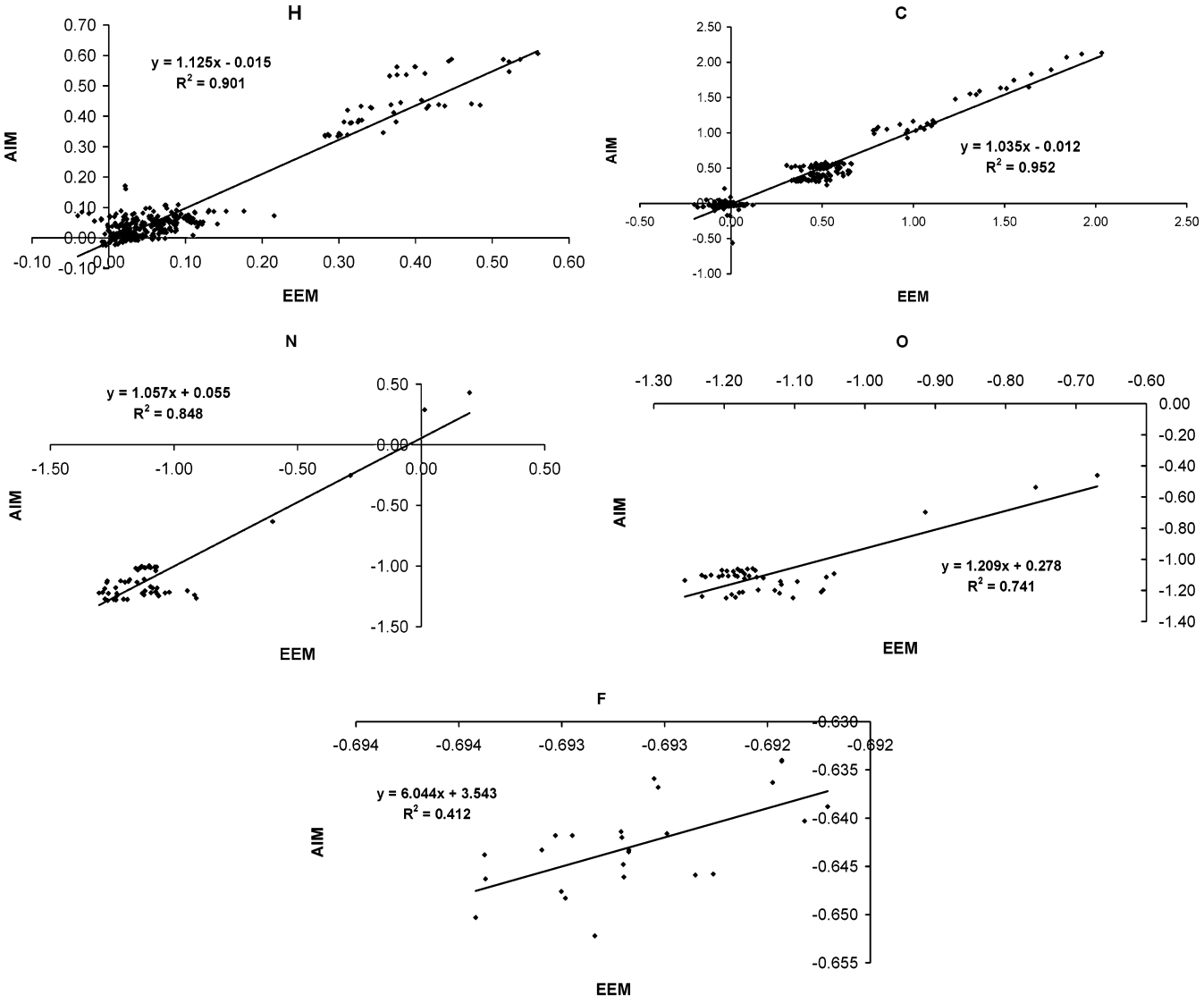

Figure 1. Comparison of EEM-AIM and DFT-AIM point charges for all atoms in the training set (separate correlations are shown for H, C, N, O, and F); EEM-AIM charges were obtained using the parameters calibrated in this work and listed in Table 1.

apparent from the calibrated parameters for fluorine. There are only minor differences in charge for the different fluorine atoms in the different molecules. As is clear from Figure 1, the range of charges of the H, C, N, O, and F atoms is 0.63, 2.70, 1.72, 0.79, and 0.02, respectively. This means that calibrating two parameters for fluorine may be doubtful from a statistical point of view. Since most fluorine atoms bear very similar DFT-AIM charges, there may be many different solutions that all yield nearly the same quality of fit between the DFT-AIM and EEM-AIM charges. Therefore, the same numerical significance should not be attached to the fluorine parameter values compared to those of the other atoms.

Quality of the EEM-AIM Charges. Figure 1 gives the EEM-AIM charges (using the optimal effective electronegativity and hardness values from Table 1) versus the DFT-AIM charges using the globally best set of parameters. 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 quantitatively reliable atomic charges. It may be argued that for fluorine, and to a somewhat smaller extent for oxygen, the agreement is much less pronounced. This is due to the more limited range of charge values for the F and O atoms. Especially for fluorine, a better correlation would only be possible if EEM would allow reproducing very small differences in atomic charge. This is clear from the average absolute differences Δ_X

$$\Delta_X = \frac{\sum_{z=1}^{N_X} |q_z^{\mathrm{EEM-AIM}} - q_z^{\mathrm{DFT-AIM}}|}{N_X} \qquad (5)$$

between the DFT-AIM and the EEM-AIM charges. For X = H, C, N, O, and F, these are respectively 0.03, 0.07, 0.10, 0.09, and 0.05, showing that the accuracy of the EEM-AIM method is quite similar for all elements. If one compares these numbers to the range of the values of the DFT-AIM charges for the different elements as can be seen from Figure 1, one immediately sees that EEM is unable to reproduce the very small changes in atomic charge between different fluorine atoms.

A few remarks should be made concerning the spread of charges in Figure 1. For the hydrogen atoms in the training set, three 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 points with lower charge are associated with hydrogen atoms bound to carbon. Two regions with a

relatively high point density occur at higher positive charges. These are associated with hydrogen atoms bound to nitrogen and oxygen. One could argue that distinguishing between the different charge states of hydrogen atoms, for example positively and negatively charged atoms, each with their own hydrogen effective electronegativity and hardness, would further improve the accuracy of EEM. Mortier et al.[40,41] followed this path for hydrogen, where they calibrated different parameters for positively charged hydrogen atoms and hydride-type atoms. Although no hydride-type atoms are present in our calibration set, a similar approach could be followed to allow separate calibrations for hydrogen in different chemical surroundings, for example distinguishing between hydrogen atoms bound to carbon, nitrogen, and oxygen. This does, however, introduce extra computational work since one then would need to identify first to which class each hydrogen atom belongs and to calibrate separate parameters for each type of hydrogen. Given the already very satisfactory results for hydrogen, we did not pursue this path. In case of other atoms, Mortier et al. also note the "surprisingly" good ability of a single value for the effective electronegativity and hardness to yield high quality charges in many different chemical surroundings.[41] The charge pattern of nitrogen also shows a number of outliers. These are mainly associated with nitro compounds. In these cases, nitrogen is positively charged according to the DFT calculations, which agrees with what is expected from simple Lewis structures and formal charges. It is gratifying that the present EEM calibration succeeds at giving positive charges for the nitrogen atoms in the nitro compounds. One set of parameters for nitrogen is able to give good EEM-AIM charges over a very wide range of chemical configurations. 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 and that maybe a set of parameters exists that could be used for both positively and negatively charged hydrogen atoms. The fact that EEM quantitatively predicts atomic charges for very different valencies of the same element, using the same parameters, shows that the calibrated parameters are applicable within a very large range of valencies. To further investigate whether this is effectively the case and how much accuracy would be gained from distinguishing different "types" of atoms of the same element, the effect of including hybridization will be discussed below.

The time savings with EEM versus ab initio AIM are substantial. EEM calculations for all molecules in the training set take less than one tenth of a second on a current personal computer and allow calculating charges for at least one million molecules of the size as those contained in the current training set per hour. The AIM calculations require first the ab initio calculations to be performed and then the time needed for finding the atomic boundaries and to carry out the numerical integration in eq 1. This could easily cost several minutes for one molecule on a current personal computer.

Applicability of EEM. In the section above, the EEM-AIM and DFT-AIM 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 amino acids were chosen as a test set. The alanine, aspartic acid, leucine, serine, and valine amino acids were chosen. Conformations were obtained through stochastic conformational analysis[49] with MM3,[50] and the lowest energy conformation was optimized at the B3LYP/6-31G* level. DFT-AIM charges were calculated at this same level. Using the globally best calibrated effective electronegativity and hardness parameters, the EEM-AIM charges were calculated. In Table 2, the average absolute differences between the DFT-AIM charges and EEM-AIM charges are given. These data were calculated over all atoms in every molecule using eq 5 and summing over all atoms together. The results clearly show that although the calibration set used does not hold any of these amino acids, the charges predicted by EEM agree very well with the DFT-AIM charges. Taking into account the speed of the EEM-AIM 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. The gain in speed will become even more apparent for larger molecules or molecules with many different conformations. Another instance where EEM can be very advantageous lies in drug design where atomic charges are often used in the screening stage of often hundreds of thousands of molecules or to calculate molecular descriptors for reactivity studies or QSAR.[17,35] To examine whether EEM-AIM charges perform well in this area, fluanisone was chosen as a test molecule. This neuroleptic is depicted in Figure 2 and was used previously to examine the applicability of EEM for different charge schemes.[28−29] The performance of EEM was found to be quite good, giving an average absolute deviation between the DFT-AIM and EEM-AIM charges over all 51 atoms of 0.04 au and a correlation coefficient of 0.9665 considered over all atoms in the molecule.

TABLE 2: Average Absolute Differences Δ_{All} (in au) between the DFT-AIM Charges and EEM-AIM Charges for All Atoms in the Validation Set of Amino Acids[a]

	Δ_{All}	R^2		Δ_{All}	R^2		Δ_{All}	R^2
Ala	0.084	98.38	Leu	0.071	97.66	Val	0.075	97.87
Asp	0.077	99.35	Ser	0.074	99.03	All	0.076	98.59

[a] R^2 (in %) is the correlation coefficient over all atoms in every molecule. The row "all" gives the average absolute difference and R^2 over all atoms in all molecules.

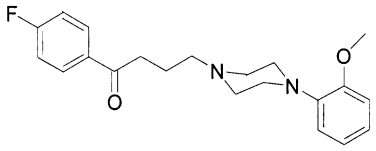

Figure 2. Lewis structure of fluanisone.

Klopman−Ohno−Mataga−Nishimoto (KOMN) Expression. The influence of the presence of other atoms in the molecules was up to now modeled through $\Sigma_{\beta\neq\alpha}^{N} q_\beta/R_{\alpha\beta}$. This electrostatic model is a computationally interesting choice but is not the only possible choice. The electrostatic model gives potentials that are quite hard, and an often used alternative is the KOMN expression.[46−48] Using

$$\eta_{\alpha\beta}(R_{\alpha\beta}) = \frac{1}{\sqrt{{R_{\alpha\beta}}^2 + \dfrac{1}{(\eta_\alpha + \eta_\beta)^2}}} \tag{6}$$

the EEM equations become

$$\chi_{eq} = \chi_\alpha = \chi_\alpha^* + \sum_\beta^N \eta_{\alpha\beta}(R_{\alpha\beta}) q_\beta \tag{7}$$

This EEM expression has the disadvantage that more computational manipulations need to be carried out, making the computer code slower. On the other hand, the interaction term becomes less hard, and a unified expression can be used for

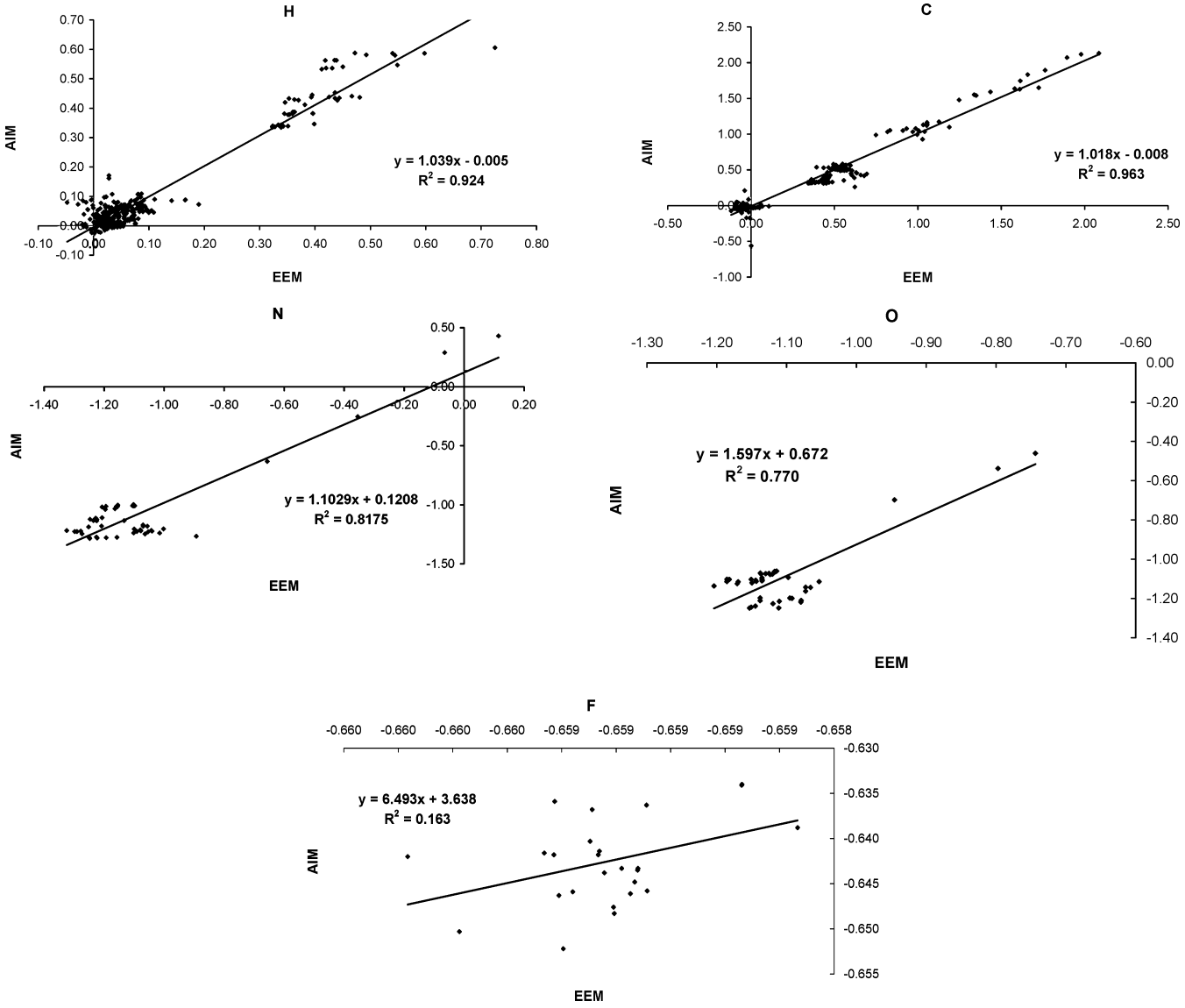

Figure 3. Comparison between EEM-AIM and DFT-AIM point charges for all atoms in the training set using the KOMN interaction term.

the second and third terms of the right-hand side of eq 2. The improvement by using the KOMN equation was investigated.

Again extensive calibrations were performed using the Lamarckian genetic algorithm. Although it is impossible to show conclusively that the global minimum has been found, the thoroughness of the calibrations makes us conclude that only a minor improvement of the correlations was brought about by replacing the original interaction model with the KOMN model. Figure 3 and Table 3 indeed show that the improvement was fairly modest. In some cases, the effect of introducing the KOMN equation even proved counterproductive, as is clear from Table 3. For some elements, the average absolute deviation is larger when using the KOMN expression, whereas for other elements, the use of the KOMN equation improves the agreement between DFT-AIM and EEM-AIM charges. The fact that further calibration is unlikely to improve the results much more, combined with the extra computational cost associated with the use of the KOMN formula, makes us opt for the continued use of the simple $\Sigma_{\beta\neq\alpha}^{N} q_\beta/R_{\alpha\beta}$ electrostatic interaction model.

The parameters obtained using both interaction models naturally undergo changes, but the important feature that the electronegativity of carbon is only modestly above that of hydrogen remains in both cases. Another observation is that the hardness of hydrogen and oxygen are reversed when going from the electrostatic to the KOMN interaction term. The trends in parameters can thus also depend on the interaction model used. Inversion of the hardness sequence between these two atoms appeared already several times in EEM calibrations in other studies as well.[28]

The fact that the atomic charges and the correlation with the DFT-AIM charges do not undergo important changes does not necessarily mean that derived quantities would not undergo

TABLE 3: Calibrated Effective Electronegativity and Hardness Parameters (in eV) Using the Original Electrostatic Potential Model and the KOMN Model[a]

atom		original	KOMN	atom		original	KOMN
H	χ^*	1.00	1.00	H	N	495	495
	η^*	20.57	15.00		Δ_H	0.033	0.027
C	χ^*	1.77	1.64	C	N	316	316
	η^*	9.24	7.33		Δ_C	0.069	0.056
N	χ^*	14.80	13.90	N	N	59	59
	η^*	10.74	9.84		Δ_N	0.103	0.120
O	χ^*	29.46	34.19	O	N	44	44
	η^*	15.85	18.98		Δ_O	0.093	0.086
F	χ^*	62.84	82.63	F	N	28	28
	η^*	46.05	82.88		Δ_F	0.050	0.017

[a] Δ_X denotes the average absolute deviation (in au) for every element X separately.

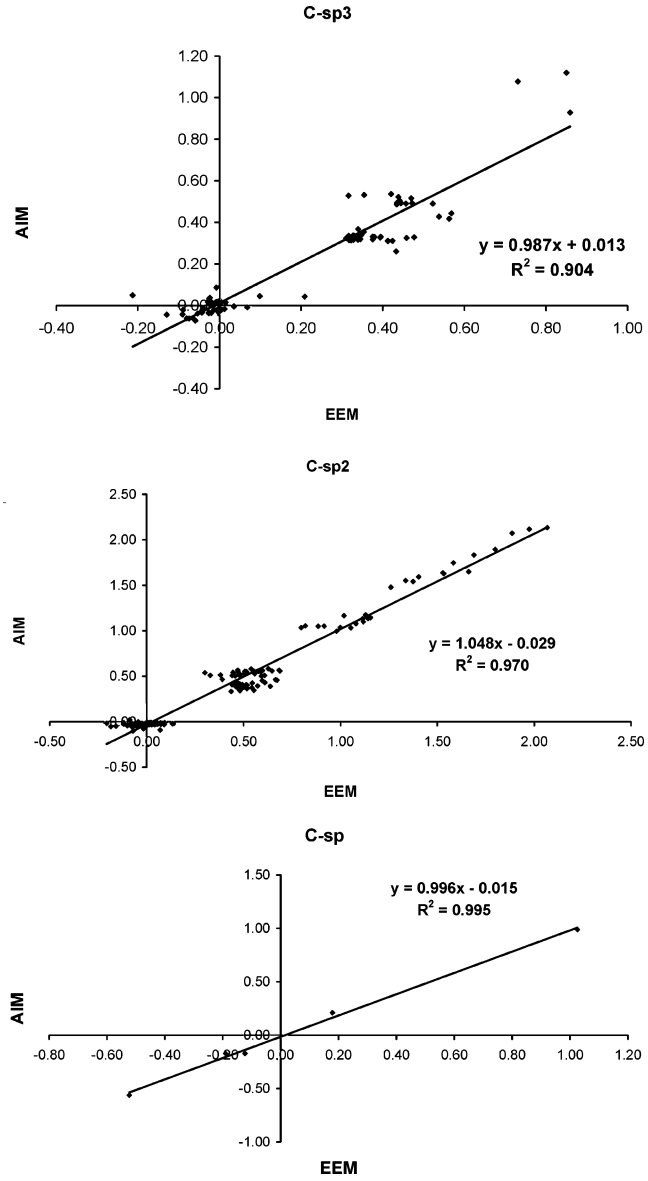

Figure 4. Comparison between EEM-AIM and DFT-AIM point charges for all carbon atoms in the different hybridization states.

changes. Fukui functions derived from EEM for example, depend on the characteristics of the hardness matrix, especially the degree of diagonal dominance.[32−34,51] This will naturally differ between the both types of hardness matrix.

Effect of Hybridization. Although the agreement between the EEM and AIM charges was found to be very good, a possible path to further improvement lies in considering different states of hybridization for the different elements. Such a path was previously also taken in the EEM calibrations by van Duin et al.[42] To test whether calibrating separate parameter sets for different hybridization states effectively results in a better correlation, we distinguished between carbon in sp^3, sp^2, and sp hybridization states. The results of this test are shown in Figure 4. Table 4 gives the average absolute deviation between the different atoms in the different hybridization states and the average absolute deviation over all atoms of the same element. As Table 4 shows, no sufficiently meaningful differences in agreement between the DFT-AIM and EEM-AIM are found which justify the extra cost of having to identify the hybridization state of every (carbon) atom in every molecule. This again accentuates the validity of using a single parameter set per

TABLE 4: Calibrated Parameters (in eV) for Carbon Including Hybridization, Compared to the Original Calibration[a]

atom		without hybridization	including hybridization
Csp^3	χ^*		1.57
	η^*		9.88
	N		116
	Δ_{Csp^3}	0.061	0.048
	R^2	90.73	90.44
Csp^2	χ^*		1.30
	η^*		9.20
	N		193
	Δ_{Csp^2}	0.068	0.071
	R^2	97.14	96.96
Csp	χ^*		4.67
	η^*		6.56
	N		7
	Δ_{Csp}	0.236	0.034
	R^2	76.02	99.94
C	χ^*	1.77	
	η^*	9.24	
	N	316	316
	Δ_C	0.069	0.062
	R^2	96.25	96.25

[a] N denotes the number of atoms of each type in the calibration set. Δ_X denotes the average absolute deviation for carbon for every hybridization state X separately and for all carbon atoms together when not distinguishing between different hybridization states. R^2 (in %) is the correlation coefficient between DFT-AIM and EEM-AIM charges.

element, as was implied by the surprisingly good ability of using a single parameter set for a large range of valencies, as first described by Mortier et al.[41] An exception lies in the case of sp hybridized carbon atoms. There the correlation coefficient increases quite a lot, and the average absolute difference lowers substantially. Obviously, care should be taken in drawing conclusions since only 7 such atoms were present in the test set. The most important reason this improvement is observed lies in an sp carbon atom of propyn, which clearly was an outlier in Figure 1. For that atom, an AIM charge is obtained of −0.56, whereas the EEM-AIM charge was 0.01 when not considering hybridization. When making a separate calibration for sp hybridized carbon, the EEM-AIM value is −0.52. Table 4 shows that the optimized sp^3 and sp^2 carbon atom parameters do not deviate very much from those for carbon when not considering hybridization. Compared to the sp^3 and sp^2 parameters, the values for the parameters of sp carbon differ more from those obtained for carbon when not considering hybridization. This is not unexpected, given the high and dominant population of sp^3 and sp^2 atoms in the original training set of molecules. Concerning the outlier of the propyn sp carbon in Figure 1, it is not clear what the origin of the very negative charge of that atom is, and why it is an outlier whereas this is not the case for other sp carbon atoms. It is naturally not surprising that when only 7 such atoms are present, of which some are nearly redundant, the use of two parameters allows a very good fit including the outlier atom, since then one has (near) complete determination. However, it is interesting to see that the electronegativity of the sp carbon atoms is substantially higher, whereas those of the sp^2 and sp^3 atoms are roughly similar. This is coherent with the sequence of the carbon acidity in different hybridization states.[52] The negative charge that is produced on the carbon atom upon hydrogen abstraction is better catered for with increasing s-character of the carbon hybridization, which in turn is coherent with an increase of the electronegativity.

Conclusions

A large set of small organic molecules was used to test the possibility of using the electronegativity equalization scheme to predict AIM charges. To that end, the parameters were calibrated against B3LYP/6-31G* charges. It was found that good agreement could be obtained between the AIM and EEM charges.

Using EEM to predict AIM charges allows the very fast computation of atomic charges without the need of prior SCF calculations, thereby allowing the calculation of AIM charges in large molecules or large sets of molecules at very high speeds, avoiding the time-consuming AIM numerical integration.

Validation of the EEM calibration for AIM charges in amino acids and a neuroleptic revealed the very good predictive power of EEM calculations, giving correlation coefficients near 98%.

It was found that including a different potential model, namely the KOMN equation, did not yield substantial improvements. The same conclusion could be drawn for distinguishing between different hybridization states of the atoms.

Acknowledgment. P.B. thanks Ghent University and the Fund for Scientific Research-Flanders (Belgium) for their grants to the Quantum Chemistry group at Ghent University. 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) Mulliken, R. S. *J. Chem. Phys.* **1962**, *36*, 3428.
(2) Löwdin, P. O. *Adv. Quantum Chem.* **1970**, *5*, 185.
(3) Besler, B. H.; Merz, K. M., Jr.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431.
(4) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129.
(5) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361.
(6) Reed, A. E.; Curtiss, L. A.; Weinhold, F. *Chem. Rev.* **1988**, *88*, 899.
(7) Hirshfeld, F. L. *Theor. Chim. Acta* **1977**, *44*, 129.
(8) Bader, R. F. W. *Atoms in Molecules*; Clarendon Press: Oxford, U.K., 1990.
(9) Bader, R. F. W. *Chem. Rev.* **1991**, *91*, 893.
(10) Popelier, P. *Atoms in Molecules: An Introduction*; Prentice Hall: Essex, U.K., 2000.
(11) Bader, R. F. W. *Phys. Rev. B* **1994**, *49*, 13348.
(12) Bader, R. F. W.; Popelier, P. L. A. *Int. J. Quantum Chem.* **1993**, *45*, 189.
(13) Bader, R. F. W.; Zou, P. F. *Chem. Phys. Lett.* **1992**, *191*, 54.
(14) Popelier, P. *Theor. Chem. Acc.* **2001**, *105*, 393.
(15) Chattaraj, P. K.; Nath, S.; Maiti, B. Reactivity Descriptors. In *Computational Medicinal Chemistry for Drug Discovery*; Bultinck, P., De Winter, H., Langenaeker, W., Tollenaere, J. P., Eds.; Dekker: New York, 2004; pp 295−322.
(16) Downs, G. M. Molecular Descriptors. In *Computational Medicinal Chemistry for Drug Discovery*; Bultinck, P., De Winter, H., Langenaeker, W., Tollenaere, J. P., Eds.; Dekker: New York, 2004; pp 515−537.
(17) Karelson, M. Quantum-Chemical Descriptors in QSAR. In *Computational Medicinal Chemistry for Drug Discovery*; Bultinck, P., De Winter, H., Langenaeker, W., Tollenaere, J. P., Eds.; Dekker: New York, 2004; pp 641−668.
(18) Breneman, C. M.; Rhem, M. *J. Comput. Chem.* **1997**, *18*, 182.
(19) Song, M.; Breneman, C. M.; Bi, J.; Sukumar, N.; Bennett, K. P.; Cramer, S.; Tugcu, N. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1347.
(20) Adam, K. R. *J. Phys. Chem. A* **2002**, *106*, 11963.
(21) Matta, C. F.; Bader, R. F. W. *Proteins: Struct. Funct. Genet.* **2003**, *52*, 360.
(22) Bader, R. F. W.; Matta, C. F.; Martín, F. J. Atoms in medicinal chemistry. In *Medicinal Quantum Chemistry*; Alber, F., Carloni, P., Eds.; Wiley-VCH: New York, 2003; pp 201−231.
(23) Bohórquez, H. J.; Obregón, M.; Cárdenas, C.; Llanos, E.; Suárez, C.; Villaveces, J. L.; Patarroyo, M. E. *J. Phys. Chem. A* **2003**, *107*, 10090.
(24) Popelier, P. L. A.; Aicken, F. M. *Chem. Phys. Chem.* **2003**, *4*, 824.
(25) Sanderson, R. T. *Science* **1951**, *114*, 670.
(26) Sanderson, R. T. *Polar Covalence*; Academic Press: New York, 1983.
(27) Mortier, W. J.; Ghosh, S. K.; Shankar, S. *J. Am. Chem. Soc.* **1986**, *108*, 4315.
(28) Bultinck, P.; Langenaeker, W.; Lahorte, P.; De Proft, F.; Geerlings, P.; Waroquier, M.; Tollenaere, J. P. *J. Phys. Chem. A* **2002**, *106*, 7887.
(29) Bultinck, P.; Langenaeker, W.; Lahorte, P.; De Proft, F.; Geerlings, P.; Van Alsenoy, C.; Tollenaere, J. P. *J. Phys. Chem. A* **2002**, *106*, 7895.
(30) Parr, R. G.; Yang W. *Density Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1989.
(31) Geerlings, P.; De Proft, F.; Langenaeker, W. *Chem. Rev.* **2003**, *103*, 1793.
(32) Baekelandt, B. G.; Janssens, G. O. A.; Toufar, H.; Mortier, W. J.; Schoonheydt, R. A. In *Acidity and Basicity in Solids: Theory, Assessement and Utility*; Fraissard, J., Petrakis, L., Eds.; NATO ASI Series C444; Kluwer Academic Publishers: Norwell, MA, 1994; p 95.
(33) Bultinck, P.; Carbó-Dorca, R. *Chem. Phys. Lett.* **2002**, *364*, 357.
(34) Bultinck, P.; Carbó-Dorca, R.; Langenaeker, W. *J. Chem. Phys.* **2003**, *118*, 4349.
(35) Bultinck, P.; Langenaeker, W.; Carbó-Dorca, R.; Tollenaere, J. P. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 422.
(36) 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.
(37) Popelier, P. L. A. *Comput. Phys. Comm.* **1996**, *93*, 212.
(38) Popelier, P. *Comput. Phys. Comm.* **1998**, *108*, 180.
(39) Menegon, G.; Shimizu, K.; Farah, J. P. S.; Dias, L. G.; Chaimovich, H. *Phys. Chem. Chem. Phys.* **2002**, *4*, 5933.
(40) Van Genechten, K. A.; Mortier, W. J.; Geerlings, P. *J. Chem. Phys.* **1987**, *86*, 5063.
(41) Mortier, W. J. *Acad. Anal., Mededelingen Koninklijke Acad. Wetenschappen, Lett. Schone Kunsten* **1990**, *52*, 29.
(42) van Duin, A. C. T.; Baas, J. M. A.; van de Graaf B. *J. Chem. Soc., Faraday Trans.* **1994**, *90*, 2881.
(43) Njo, S. L.; Fan, J.; van de Graaf, B. *J. Mol. Catal. A* **1998**, *134*, 79.
(44) Lackner, K. S.; Zweig, G. *Phys. Rev. D* **1983**, *28*, 1671.
(45) Bergman; Hinze, J. *Angew. Chem., Int. Ed.* **1996**, *35*, 150.
(46) Ohno, K. *Theor. Chim. Acta* **1964**, *2*, 219.
(47) Klopman, G. *J. Am. Chem. Soc.* **1964**, *86*, 4550.
(48) Mataga, N.; Nishimoto, K. *Z. Phys. Chem.* **1957**, *13*, 140.
(49) Saunders, M. *J. Am. Chem. Soc.* **1987**, *109*, 3150.
(50) MM3 versions 1994; 1996, obtained from QCPE.
(51) Bultinck, P.; Carbó-Dorca, R. *J. Math. Chem.* **2003**, *34*, 67.
(52) Isaacs, N. S. *Physical Organic Chemistry*; Longman Scientific & Technical: Essex, U.K., 1987.