Fast Calculation of Quantum Chemical Molecular Descriptors from the Electronegativity Equalization Method

Patrick Bultinck,*,† Wilfried Langenaeker,‡ Ramon Carbó-Dorca,§ and Jan P. Tollenaere#

Department of Inorganic and Physical Chemistry, Ghent University, Krijgslaan 281 (S-3), B-9000 Gent, Belgium, Johnson & Johnson Pharmaceutical Research and Development, Molecular Design and Chemoinformatics, Turnhoutseweg 30, B-2340 Beerse, Belgium, Institute of Computational Chemistry, University of Girona, Campus Montilivi, 17071 Girona, Catalonia, Spain, and Department of Medicinal Chemistry, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, P.O. Box 80082, 3508 TB Utrecht, The Netherlands

Received August 21, 2002

The use of the Electronegativity Equalization Method (EEM) is presented for high performance calculation of molecular electrostatic descriptors, giving quite similar results to those obtained through Density Functional Theory (B3LYP/6-31G*) calculations. Molecular descriptors include atomic charges and different related descriptors as well as Fukui functions, hardness and softness.

INTRODUCTION

The number of molecules that can be synthesized is so vast that the chance that a randomly drawn molecule would exhibit a certain desired biological activity is practically zero. As a consequence, there has developed with time quite a large interest in methods to relate molecular structure to biological activity. As a consequence of this timeliness interest, more than a hundred years ago the field known now as Quantitative Structure−Activity Relationships (QSAR) emerged.[1] In this field one tries to relate elements of chemical structure of a molecule to its extent of biological activity.[2] Mathematically this translates into a very simple formula as

$$A = f(S) \tag{1}$$

expressing that the activity A is a function of the chemical structure, represented by S.

Despite the apparent simplicity of this formula, establishing such a relation is not straightforward. It raises some very basic questions such as the nature of chemical structure, how to express it in a mathematical form etc., problems which are still far from being solved. As a consequence, despite the early birth of the QSAR idea, it has taken many decades until reaching mature age. Two 1964 papers by Hansch and Fujita on one hand[3] and Free and Wilson,[4] on the other hand, can be considered as the start of modern QSAR, indicating that $f(S)$ (see eq 1) can be expressed as a linear combination of different molecular properties. The biological activity is usually expressed in terms of the concentration (C) of the molecule needed to induce some desired effect, often as the log(1/C). In "classical" QSAR, as opposed to Quantum QSAR,[5] S is represented as a set of different molecular properties such as e.g. lipophilicity, molecular weight, solubility, ... These are often called descriptors, and QSAR is based on finding a usually linear relationship between the observed activity and the statistically meaningful descriptor set.

There exists a wide range of descriptors, including solubility, partition coefficients, Hammett parameters, etc. As these descriptors cannot always be obtained experimentally, due to experimental difficulty or because one is dealing with virtual molecules, there is great interest in finding other ways to evaluate these descriptors. One option among a large collection, consists of the use of rule based or fragment approaches, where the descriptor for the molecule is obtained as a combination of different tabulated substituent or fragment contributions. In the present paper, however, the generation of molecular descriptors, considering the molecules as a whole, based on quantum chemical computations and the electronegativity equalization technique is addressed. Quantum chemical calculations can be used to obtain a priori descriptors. Such descriptors include not only the quantum chemical translation of the experimental observables but also new types of nonobservable quantities. The first case includes for instance the dipole moment and involves the calculation of expectation values using operators obtained through the correspondence principle. The latter case includes the introduction and calculation of nonobservables such as atomic charges and related descriptors. Given the fact that such molecular properties are not observable, this does entail that the way to calculate them is not uniquely defined, and as such there can be many different schemes for their calculation, none of which is fundamentally more correct than the other. Despite this theoretical problem they have proved useful in QSAR. Even though in both cases quantum chemistry is used, such descriptor seeking techniques cannot be properly named as quantum QSAR. Quantum QSAR, as introduced and defined by Carbó-Dorca and co-workers,[5] relies on molecular quantum similarity matrices computed using the electron densities of the molecules involved and integrals of the general type $z_{ab} = \langle \rho_A | \Omega | \rho_B \rangle$, which in the simplest case transforms in a kind of overlap integral. The resulting matrix elements z_{ab} are then used as descriptors in QSAR within a relationship as in eq 1, which is deduced as

* Corresponding author phone: +32/9/264.44.23; fax: +32/9/264.49.83; e-mail: Patrick.Bultinck@rug.ac.be.
† Ghent University.
‡ Johnson & Johnson Pharmaceutical Research and Development, Molecular Design and Chemoinformatics.
§ University of Girona.
# Utrecht University.

10.1021/ci0255883 CCC: $25.00
Published on Web 11/28/2002

a consequence of the quantum mechanical expectation value definition.

As QSAR nowadays plays an important role in (bio)-pharmaceutical research, the fast calculation of molecular descriptors is of great importance. Quantum chemistry based descriptors have been shown to be quite useful,[6] and atomic charges have been used since early times for this purpose.[7] Unfortunately their calculation is often computationally too demanding for large sets of large molecules. Bultinck et al.[8,9] recently described an implementation of the Electronegativity Equalization Method (EEM) to allow very fast and chemically consistent calculation of atomic charges. In the present paper, the calculation of related electrostatic QSAR descriptors will be presented, and the agreement with the DFT calculated values investigated.

ELECTRONEGATIVITY EQUALIZATION

According to the electronegativity equalization principle described by Sanderson,[10,11] 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, such as the widely used Gasteiger-Marsili scheme.[12] Bultinck et al.[8,9] previously described the implementation of the EEM of Mortier et al.[13] with an application emphasis focused on medicinal chemistry.

For each atom in an N-atom molecule the following equation holds, (see Mortier et al.[13] for details on these formulas):

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

In eq 2, $R_{\alpha\beta}$ stands for the internuclear distance between atoms α and β. q_α is the atomic charge on atom α, and χ_{eq} is the equalized molecular electronegativity. χ_α^* and η_α^* stand for the effective electronegativity and hardness of atom α in the molecule, respectively. As such they differ from the isolated atom states and are to be considered atoms-in-molecules properties. χ_α^* and η_α^* can as a consequence not be calculated directly and have to be obtained through calibration. In the calibration process, it was assumed that for every atom of the same element in all the molecules of the calibration set, the same parameters χ^* and η^* can be used in eq 2.

The set of eq 2, one for each atom in the molecule, and the requirement that the sum of atomic charges should yield the total molecular charge Q can be combined in matrix form as

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

The calibration of the χ_α^* and η_α^* parameters based on B3LYP/6-31G* charges has been described in detail earlier,[8] and it has been found that EEM holds the possibility of yielding atomic charges at great speed, within a rate of 1.5 million medium sized molecules/hour on a personal computer, retaining very close agreement with DFT calculated charges.[8,9] As indicated above, atomic charges are not observables. Electron density itself is an observable, but the problem arises due to the lack of an operator defining an atomic basin in the molecule over which to integrate this density. As a consequence different ways to obtain atomic charges will give rise to different values of χ_α^* and η_α^*. Not every type of quantum chemical population analysis is equally well suited for EEM and because of this the present work will focus on Mulliken (MPA) and natural atomic (NPA) charges. These have previously been shown to perform quite well in the context of EEM.[8,9]

EEM also allows the calculation of other conceptual DFT quantities. One such a well-known reactivity descriptor is the Fukui function.[14] This function reflects to what extent different regions in the molecule are likely to be involved in a nucleophilic or electrophilic reaction. It is defined as the derivative of the electron density $\rho(r)$ at a point in space r with respect to the number of electrons N in the molecule under constant external potential V:

$$f(r) = \left(\frac{\partial \rho(r)}{\partial N}\right)_V \tag{4}$$

Instead of using f(r) itself, often atom condensed Fukui functions are used as introduced by Mortier et al.[13] This again requires atoms in molecules to be defined, yielding the following expression for the condensed Fukui function on atom α

$$f_\alpha = \left(\frac{\partial \epsilon_\alpha}{\partial N}\right)_V = -\left(\frac{\partial q_\alpha}{\partial N}\right)_V \tag{5}$$

where ϵ_α is the total electron density associated to atom alpha. The identification of an atomic region in the molecule may be done through the choice of the method for population analysis. In this case, the Fukui function is also minus the change of the atomic charge with changing of the number of electrons in the molecule. Another important quantity is the molecular global hardness, which is the change in the equalized molecular electronegativity with respect to a change in the number of electrons:[14]

$$\eta = -\frac{1}{2}\left(\frac{\partial \chi_{eq}}{\partial N}\right)_V \tag{6}$$

Although often a very similar definition is found, but without the factor $^1/_2$, to be consistent with the definition used in the derivation of the original EEM equations, the factor 1/2 is retained here. Since the EEM equations hold the electronegativities as well as atomic charges, which are directly related to the atomic integrated electron density, new derivations can be made to obtain the molecular hardness and atomic Fukui functions. In principle, one could adopt the usual finite difference approximation and solve the EEM equations for Q, $Q+\delta$, and $Q-\delta$, and afterward deduce the Fukui functions. However, the EEM equations as given above can be differentiated with respect to N. Similar derivations have also been reported by Baekelandt et al.[16] To allow the calculation of the atomic Fukui functions and molecular hardness, the effective hardness and electronegativity values

are assumed constant under the change of N. The validity of this assumption in medicinal chemistry related molecules has been shown earlier by Bultinck et al.[8,9] Combining eqs 2 and 3 as well as 5 and 6, the following expression relating Fukui functions and the molecular hardness with the calibrated atomic hardness parameters is found:

$$0 = 2\eta_\alpha^* f_\alpha + \sum_{\beta \neq \alpha}^{N} \frac{f_\beta}{R_{\alpha\beta}} - 2\eta_{mol} \tag{7}$$

Such a result is quite similar to the EEM set of equations. Adding the constraint of normalization of the Fukui function,[14] one finds the following matrix equation:

$$\begin{bmatrix} 2\eta_1^* & 1/R_{12} & \cdots & 1/R_{1N} & 1 \\ 1/R_{21} & 2\eta_2^* & \cdots & 1/R_{2N} & 1 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 1/R_{N1} & 1/R_{N2} & \cdots & 2\eta_N^* & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \\ -2\eta_{mol} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \tag{8}$$

Another set of descriptors is the global molecular softness and the local atomic softness.[14] The total softness S is defined as half the inverse of the global molecular hardness. Using the relationship between local softness and the Fukui function, while adding the constraint that the local atom condensed softnesses should add up to the total softness, one finds

$$\begin{bmatrix} 2\eta_1^* & 1/R_{12} & \cdots & 1/R_{1N} & 0 \\ 1/R_{21} & 2\eta_2^* & \cdots & 1/R_{2N} & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 1/R_{N1} & 1/R_{N2} & \cdots & 2\eta_N^* & 0 \\ 1 & 1 & \cdots & 1 & -1 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_N \\ S \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 0 \end{bmatrix} \tag{9}$$

Then these three EEM equations allow the calculation of the following molecular descriptors: atomic charges and equalized molecular electronegativity through eq 3, atom condensed Fukui functions and molecular hardness via eq 8, and finally atom condensed softness and the total softness by means of eq 9. Next to these basic descriptors, Karelson[6] has described many derived QSAR parameters such as the globally most positive and negative atomic charge in the molecule and, similarly for each element present in the molecule, a polarity-like parameter as the difference between the most positive and most negative charge as well as various sums of absolute or squared values of partial charges. Other common charge based descriptors are averages of atomic partial charges absolute values. Fukui functions, softness and hardness and derived quantities have been found to be useful molecular descriptors by different authors.[6,17−23]

COMPUTATIONAL METHODS

To test the quality of the EEM derived molecular descriptors, the Fukui, hardness and softness parameters were calculated for a set consisting of 138 molecules, representing a wide range of the medicinal chemical space, involving the elements C, H, N, O, and F. The agreement between atomic charges computed by quantum chemical calculations and EEM has previously been investigated for this same set.[8,9] The actual molecules contained in the calibration set may be found as Supporting Information to Bultinck et al.[8,9]

Next to this set, which has been used in the calibration of the χ^* and η^* element parameters, a second set of molecules, which have not been used in the calibration set, has also been considered. This second set of molecules consists of actual drug molecules of Janssen Pharmaceutica, thereby forming a good test case for the application of EEM in medicinal QSAR. The structures of the drug molecules are available as Supporting Information.

Quantum chemical calculations of the different descriptors were performed using the B3LYP/6-31G* level of calculation and the Gaussian98 program.[24] Geometries of the molecules of the calibration set were taken from our previous studies. For the actual drug molecules experimental structures were used, or in case molecular geometries were not available, they were obtained using the CORINA program by Gasteiger et al.[25]

MPA and NPA charges were calculated at the B3LYP/6-31G* level. Fukui functions were calculated at this same level by finite differences between the neutral and cationic and anionic states of the molecule. The molecular hardness was calculated from the ionization energy (I) and electron affinity (A) as (I−A)/2.[14] The total softness is just half the inverse of the molecular hardness, and knowledge of the Fukui functions is then used to obtain the atomic softnesses. The finite difference approximation is, however, a quite crude approximation of the derivative in eq 5. Furthermore, other factors such as relaxation may have an influence on the calculated quantities such as Fukui functions. Therefore, DFT finite difference calculated values are considered as a reference, without assumption of absolute correctness. The EEM procedure using the parameters by Bultinck et al.[8,9] for MPA and NPA population analysis is then used to calculate this same set of descriptors, and both sets are compared. Different electrostatic molecular descriptors reported by Karelson et al.[6] are constructed from both the DFT and EEM charges and compared.

RESULTS AND DISCUSSION

1. Atomic Charges and Related Quantities. The agreement between the atomic charges produced by the DFT and EEM calculations was previously discussed[8,9] for the set of molecules used in the calibration. It was found that the agreement was very good, both for MPA and NPA. For the second set of molecules, consisting of actual drug molecules, again good agreement is found between the B3LYP/6-31G* and EEM charges, both for MPA and NPA. Table 1 gives the average absolute deviation between the DFT and EEM charges for the five different elements and for both types of population analysis. The agreement is seen to be good for most elements, especially for MPA. For nitrogen the NPA based agreement is somewhat less, in accordance with previous findings for neuroleptics.[8,9] Still, as will be discussed below, EEM with the NPA based parameters succeeds at ranking the nitrogen atoms correctly, for instance properly indicating where the most positive nitrogen atom is found in each molecule, and over all molecules where the most positive nitrogen atom is found.

It is important, in this context, to realize that calculation of atomic charges using B3LYP/6-31G* takes several CPU

Table 1. Average Absolute Difference (au) between the DFT and EEM Calculated Charges in Actual Drug Molecules[a]

	average absolute difference
H	0.01
	0.02
C	0.06
	0.07
N	0.06
	0.15
O	0.04
	0.05
F	0.03
	0.07

[a] The first numbers refer to MPA, the ones in italics to NPA.

hours on a powerful server for the 138 fairly small molecules in the calibration set but that all EEM atomic charges for the set of molecules in the calibration set have been calculated in less than one second on a medium range personal computer. For the actual drug molecules, the situation is even more outspoken. The increase in computational cost for the DFT calculations is several orders of magnitude higher than for the solution of the EEM matrix equation. Solving the EEM equations for the different molecules takes less than 100 milliseconds on a personal computer.

Karelson et al.[6] have described a number of molecular electrostatic descriptors derived from atomic charges. The ones used in the present study for the set of medicinal molecules are given in Table 2.

For the different drug molecules it has first been examined whether the same atoms were identified as the most positive or negative one in the DFT and EEM calculations. It was found that EEM succeeds very well in doing so. The most important cases are very well resolved, for example for the most positively charged carbon atom, often involved in a nucleophilic reaction. This is, however, only achieved in those cases where there is a sufficiently large difference in chemical environment and in resulting atomic charge between atoms of the same element. This means that in the case where there are several atoms in a very similar chemical environment and with very similar DFT atomic charges, EEM cannot fully rank the atoms correctly. This is of no consequence for the identification of reactive centers, since such fuzzy cases are mainly associated with discriminating hydrogen atoms bound to carbon in aliphatic carbons. The DFT calculated differences in charges are less than 0.01 in those cases, and consequently it becomes difficult for EEM to reproduce the ranking.

All descriptors from Table 2 were calculated for the different molecules and examined whether the same trends as in DFT were found. Given the vast set of numbers that need to be compared, no quantitative information is presented, but rather trends are given in a table showing the EEM performance for different descriptors. For every descriptor the molecules were ranked according to their descriptor values, calculated using both DFT and EEM. Both rankings may then be compared. The results show that for many descriptors almost the same ranking is obtained on the basis of the DFT calculations and the EEM calculations. EEM is thus a reliable and fast method to rank molecules for the different descriptors employed here. It is especially useful to identify the extremes. As an example, one can take the following descriptors: most positive, most negative carbon atom, and the carbon charge separation. The most positive atom in every molecule has been identified, and molecules are ranked according to this most positive charge. It was found that the ranking was identical whether DFT or EEM charges in the NPA framework were used. A similar performance has been observed for the most negative carbon atom, except that again EEM is unable to reproduce very small charge differences. As a consequence there appears one inversion in the ranking between DFT and EEM. This inversion occurs at the middle of the ranking, and in fact the ranking of the DFT results is based on a charge difference of only a few thousandths. However, the extremes are again correctly assigned to the correct molecules and to the correct atom in these molecules. It is worth noting that the very small differences in DFT calculated NPA atomic charges sometimes are meaningless if one considers that, due to for instance the use of cutoffs in the geometry optimizations, a different calculation may reverse the order. Looking at the charge separation for a specific element, it was found that the resulting ranking does not agree so well between both techniques to yield charges. This is due to the fact that there are some differences in the charges themselves, which in a difference like in $C_{Q,sep}$ influence the ranking. Charge separations per element should therefore not be calculated using EEM. For the maximum charge separation a similar effect is found, although the EEM and DFT based maximum charge separation values both rank the same molecule as the most polar one. The maximum negative charge separation is again a problem due to the reason given above. The sums of squared charges over all atoms correlate nicely between the DFT and EEM methods. For the mean of the positive charges good agreement is found; worse agreement on the other hand is found for the negative charges. For the average absolute charge good agreement is again found. For the descriptors involving relative charges, EEM performs adequately at identifying the extremes. Summarizing the results, Table 3 gives an overview of the agreement between DFT and EEM for the calculation of molecular electrostatic descriptors in several classes, ranging from ++ to −. It should also be noted that the selection of the molecular descriptors to be used depends on the QSAR problem at hand and that the table should not be considered the definite construct for employing certain descriptors. Also, the subsequent study of other problems in QSAR might reveal that EEM is a powerful and useful technique for the calculation of molecular descriptors, even though in the present study some of these may have received a "−" mark in Table 3. The choice of the descriptor to use or the sets of descriptors, from which some are selected, in a specific QSAR study is however user-based and connected to some choice usually based on statistical techniques. EEM cannot alleviate the associated technical problems. This matter is largely solved in quantum QSAR, although even there some steps need user expertise, for example in quantum similarity the choice of the adequate operator used in $z_{ab} = \langle \rho_A | \Omega | \rho_B \rangle$.

2. Fukui Functions, Molecular Hardness and Softness. Fukui functions and related quantities have previously been used in QSAR studies.[17−23] They are also considered among the most illustrative reactivity descriptors in conceptual density functional theory.[14,26] Usually, in quantum chemical

Table 2. Different Derived Molecular Electrostatic Descriptors and Notation

descriptor	notation
most positive (least negative) charge on element X in the molecule	$H_{Q,MAX}$, $C_{Q,MAX}$, $N_{Q,MAX}$, $O_{Q,MAX}$, $F_{Q,MAX}$
most negative (least positive) charge on element X in the molecule	$H_{Q,min}$, $C_{Q,min}$, $N_{Q,min}$, $O_{Q,min}$, $F_{Q,min}$
most positive (least negative) atom	$A_{Q,MAX}$
most negative (least positive) atom	$A_{Q,min}$
maximum charge separation per element	e.g. $H_{Q,sep}=H_{Q,MAX}-H_{Q,min}$
maximum charge separation	$A_{Q,MAX}- A_{Q,min}$
sum of squares of charges per element	$\sum Q_C^2$, $\sum Q_H^2$, $\sum Q_N^2$, $\sum Q_O^2$, $\sum Q_F^2$
sum of squares over all atomic charges	$\sum Q_A^2$
mean of positive charges (N=total number of positive atomic charges)	$MPC = \frac{[\sum Q_A]_{QA>0}}{N}$
mean of negative charges (N=total number of negative atomic charges)	$MNC = \frac{[\sum Q_A]_{QA<0}}{N}$
mean absolute charge (N=total number of atoms)	$MAC = \frac{\sum \lvert Q_A \rvert}{N}$
relative positive charge	$RPC = \frac{Q^+_{max}}{\sum_A Q_A} A \in \{Q_A > 0\}$ Q^+_{max} maximum positive charge in the molecule
relative negative charge	$RNC = \frac{Q^-_{min}}{\sum_A Q_A} A \in \{Q_A < 0\}$ Q^-_{min} minimum negative charge in the molecule

Table 3. Performance of EEM for the Calculation of Atomic Charge Based Electrostatic Descriptors[a]

descriptor	EEM performance
$H_{Q,MAX}$, $C_{Q,MAX}$, $N_{Q,MAX}$, $O_{Q,MAX}$, $F_{Q,MAX}$	++
$H_{Q,min}$, $C_{Q,min}$, $N_{Q,min}$, $O_{Q,min}$, $F_{Q,min}$	+
$A_{Q,MAX}$	+
$A_{Q,min}$	−
$H_{Q,sep}$, $C_{Q,sep}$, $N_{Q,sep}$, $O_{Q,sep}$, $F_{Q,sep}$	−,−,−,−,−
$A_{Q,MAX}- A_{Q,min}$	0
$\sum Q_C^2$, $\sum Q_H^2$, $\sum Q_N^2$, $\sum Q_O^2$, $\sum Q_F^2$	+
$\sum Q_A^2$	+
MPC	+
MNC	−
MAC	+
RPC	+
RNC	+

[a] "++" means nearly quantitative agreement between EEM and DFT calculated quantities, "+" good agreement, "0" neutral, and "−" poor agreement.

calculations, as it is the case here, they are obtained as finite differences involving the cation and anion of the molecule through Koopmans' theorem or adiabatically. In the present study the latter method is used for the DFT calculation of the Fukui functions. As a consequence, different values are calculated depending on whether the difference is made between the neutral molecule and the cation (f_a^-) or the neutral molecule and the anion (f_a^+). In EEM only one Fukui function is calculated. The calculation of Fukui functions through EEM is comparable in speed to the calculation of the atomic charges, since the equations are of the same dimensionality and involve the same algebraic manipulations. In case of finite difference DFT calculations of Fukui functions, the computational demand increases considerably, since three DFT calculations have to be performed. The EEM advantage in speed is therefore even more evident.

The agreement between the DFT calculated and EEM Fukui functions is poor. It is impossible to identify exactly whether this is due to the EEM scheme or to DFT related factors. In the DFT finite difference approximation a number of difficulties arise, such as the necessity to reach the same accuracy for the neutral singlet and ionic doublet states. There is also the relaxation of the orbitals when performing separate DFT calculations that may play a role. There are also a number of problems associated with the use of different types of population analysis. It is important to note that there are several instances where negative Fukui functions are found when using the DFT finite difference approach using MPA and NPA. This is a physically unlikely situation as Fukui functions are obtained as manipulations of probability density functions. The EEM derived Fukui functions, on the other hand, are always positive, both for the MPA and NPA based EEM parameter sets. This is a major advantage, since it relieves the calculation of Fukui functions from the physically unlikely situations.

A classical test for Fukui functions is the protonation of aniline and its mono-ortho, meta, and parasubstituted fluoroanilines.[27] Two competitive sites for protonation exist, namely the protonation of the nitrogen atom and that of the carbon atom in the para-position. Both of them are negatively charged, for example in aniline NPA charges are −0.84 and −0.27 for N and C4, respectively. Based on NPA derived finite difference Fukui functions, protonation of aniline would

occur on the ring but in the ortho position of the amine function. This is very unlikely and does not agree with experiment.[28] MPA indicates amine protonation. This indicates that depending on the type of population analysis used, different conclusions can be reached. When using EEM, independently of which set of parameters is used (MPA or NPA calibrated), the Fukui functions always indicate protonation of the amine nitrogen atom. This agrees with experimental findings.[28]

As indicated by Baekelandt et al.,[16] the following rules can be derived: the atom with the largest Fukui function among the set of positively charged atoms will be the preferred locus for nucleophilic attack, and the atom which of all negatively charged atoms has the largest Fukui function value is the preferred location for an electrophilic reaction. The Fukui function, as calculated here, can be used as a means to decide which atom, among a set of atoms that are likely to react in a similar way, will participate in a given reaction.

The calculation of molecular hardness in a finite difference scheme is based on a calculation of the ionization energy and electron affinity. This again implies that the method used should be of equal quality at describing singlet states and derived states of doublet multiplicity for the same molecule, with different numbers of electrons. It is therefore very hard to compare results of the DFT and EEM results. The rather poor correlation for the calibration set of 138 molecules is as follows: R^2 is approximately 40% for both the NPA and MPA based DFT-EEM correlations and may not necessarily be seen as an indication of the poor performance of EEM, since the DFT methods to calculate hardness contain many possible sources of error themselves. Obviously, similar remarks can be made for the total and atom condensed softnesses.

It may appear as EEM is only suitable for the calculation of atomic charges, because in that situation there is good agreement with the DFT calculated charges. It is necessary to stress now that this is also the only situation where DFT allows obtaining the parameters through one single calculation. EEM allows the direct calculation of the equalized electronegativity, Fukui, hardness and softness, whereas in DFT the same molecular descriptors are obtained though performing several independent calculations. This may introduce errors that could make the quantitative comparison useless. Some validation of the EEM derived quantities will therefore need to be found in their QSAR application.

CONCLUSIONS

The electronegativity equalization method (EEM) has been tested for the calculation of molecular electrostatic descriptors for use in QSAR and related fields. These descriptors are based on fast calculation of atomic charges but also include descriptors from further manipulations of the EEM equations, yielding Fukui functions, hardness and local and total softness. Calculation of all these descriptors through EEM costs less than one millionth of the time needed to obtain them from actual DFT calculations.

The EEM atomic charges were found to agree very well with those calculated using B3LYP/6-31G* calculations. For the other molecular electrostatic descriptors, the performance is found to be dependent on the nature of the descriptor. Usually the agreement is fair to good, although when charge differences are involved some descriptors were less well reproduced.

The calculation of atom condensed Fukui functions, molecular hardness and softness, and atom condensed softness is possible extending the EEM methodology via derivations presented in this work and analogously by Baekelandt et al.[16] Both for Mulliken and NPA based calculations, only positive values for the EEM Fukui functions were found. This relieves EEM Fukui functions of the physically less likely negative values obtained by finite difference DFT calculations on molecules in the neutral, cationic, and anionic situations. No quantitative agreement could be investigated between the EEM and corresponding DFT Fukui indices, mainly due to shortcomings in the finite difference DFT calculations.

The present study shows EEM to be a powerful method for generating atomic charges resembling the DFT calculated ones very closely and allowing the fast calculation of several electrostatic molecular descriptors. In some cases, the small deviations between EEM and DFT cause less good performance for some descriptors. In other cases some DFT drawbacks are avoided, like in the case of EEM Fukui functions which exhibit only positive values.

ACKNOWLEDGMENT

P. Bultinck wishes to thank Ghent University and the *Fund for Scientific Research-Flanders* (Belgium) for their grants to the Computational Chemistry group at Ghent University and acknowledges the *European Community − Access to Research Infrastructure action of the Improving Human Potential Program*, allowing the use of the CEPBA infrastructure at the PolyTechnical University of Catalonia (Spain) and the visitors fellowship with the Institute of Computational Chemistry at the University of Girona (Catalonia, Spain). *Janssen Research Foundation* is gratefully acknowledged for financial support for the postdoctoral stay of P.B. R. Carbó-Dorca acknowledges the Foundation M. F. de Roviralta as well as the CICYT project #SAF2000-223, which have supported this work.

Supporting Information Available: 2-D representations of the drug molecules used as EEM test set. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Crum Brown, A.; Fraser, T. R. On the connection between chemical constitution and physiologic action. Part 1. On the physiological action of salts of the ammonium bases, derived from strychnia, brucia, thebia, codeia, morphia and nicotia. *Trans. R. Soc. Edinburgh* **1868**, *25*, 151−203.

(2) Kubinyi, H. Quantitative Structure−Activity Relationships in Drug Design. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III, Schreiner, P. R., Eds.; John Wiley and Sons Ltd.: Chichester, UK, 1998; Vol. 4, pp 2309−2319.

(3) Hansch, C.; Fujita, T. $\rho-\sigma-\pi$ Analysis. A method for the correlation of biological activity and chemical structure. *J. Am. Chem. Soc.* **1964**, *86*, 1616−1626.

(4) Free, S. M., Jr.; Wilson, J. W. A mathematical contribution to structure−activity studies. *J. Med. Chem.* **1964**, *7*, 395−399.

(5) Carbó-Dorca, R.; Robert, D.; Amat, L.; Gironés, X.; Besalú, E. Molecular Quantum Similarity in QSAR and Drug Design. *Lecture Notes Chem.* **2000**, *73*.

(6) Karelson, M.; Lobanov, V. S.; Katritzky, A. R. Quantum-Chemical Descriptors in QSAR/QSPR Studies. *Chem. Rev.* **1996**, *96*, 1027−1043.

(7) Carbó, R.; M. Martin, M.; Pons, V. Application of quantum mechanical parameters in QSAR: Development of an electrostatic interaction hypothesis. *Afinidad* **1977**, *34*, 348−353.
(8) Bultinck, P.; Langenaeker, W.; Lahorte, P.; De Proft, F.; Geerlings, P.; Van Alsenoy, C.; Tollenaere, J. P. The electronegativity equalization method I: parametrization and validation for atomic charge calculations. *J. Phys. Chem. A.* **2002**, In press.
(9) Bultinck, P.; Langenaeker, W.; Lahorte, P.; De Proft, F.; Geerlings, P.; Waroquier, M.; Tollenaere, J. P. The electronegativity equalization method II: applicability of different atomic charge schemes. *J. Phys. Chem. A* **2002**, In press.
(10) Sanderson, R. T. *Science* **1951**, *114*, 670.
(11) Sanderson, R. T. *Polar Covalence*; Academic Press: New York, 1983.
(12) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity − A Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219−3228.
(13) Mortier, W. J.; Ghosh, S. K.; Shankar, S. Electronegativity Equalization Method for the Calculation of Atomic Charges in Molecules. *J. Am. Chem. Soc.* **1986**, *108*, 4315−4320.
(14) Parr, R. G.; Yang W. *Density Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1989.
(15) Reed, A. E.; Weinstock, R. B.; Weinhold, F. Natural population analysis. *J. Chem. Phys.* **1985**, *83*, 735−746.
(16) Baekelandt, B. G.; Janssens, G. O. A.; Toufar, H.; Mortier, W. J., Schoonheydt, R. A. Acidity and Basicity concepts: The Density Functional viewpoint based on an atoms-in-molecules sensitivity analysis − A tutorial. In *Acidity and Basicity in Solids: Theory, Assessement and Utility*; Fraissard, J., Petrakis, L., Eds.; NATO ASI Series C444; Kluwer Academic Publishers: 1994; pp 95−126.
(17) Zhou, Z.; Parr, R. G. Activation hardness. New Index for Describing the Orientation of Electrophilic Aromatic Substitution. *J. Am. Chem. Soc.* **1990**, *112*, 5720−5724.
(18) Pearson, R. G. Absolute Electronegativity and Hardness: Applications to Organic Chemistry. *J. Org. Chem.* **1989**, *54*, 1423−1432.
(19) Tuppurainen, K.; Lötjönen, S.; Laatikainen, R.; Vartiainen, T.; Maran, U.; Strandberg, M.: Tamm, T. About the mutagenicity of chlorine-substituted furanones and halopropenals. A QSAR study using molecular orbital indexes. *Mutat. Res.* **1991**, *247*, 97−102.
(20) Gaudio, A. C.; Korolkovas, A.; Takahata, Y. Quantitative Structure−Activity Relationships for 1,4-Dihydropyridine Calcium Channel Antagonists (Nifedipine Analogues): A Quantum/Classical Approach. *J. Pharm. Sci.* **1994**, *83*, 1110−1115.
(21) Pearson, R. G. *Chemical Hardness*; Wiley-VCH: Weinheim, 1997.
(22) Langenaeker, W.; Demel, K.; Geerlings, P. Quantum-chemical study of the Fukui function as a reactivity index; Part 2. Electrophilic substitution on monosubstituted benzenes. *J. Mol. Struct. (THEOCHEM)* **1991**, *234*, 329−342.
(23) Baeten, A.; Tafazoli, M.; Kirsch-Volders, M.; Geerlings, P. Use of the HSAB Principle in Quantitative Structure−Activity Relationships in Toxicological Research: Application to the Genotoxicity of Chlorinated Hydrocarbons. *Int. J. Quantum Chem.* **1999**, *74*, 351−355.
(24) 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.; Baboul, A. G.; 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.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, Revision A.7; Gaussian, Inc.; Pittsburgh, PA, 1998.
(25) Gasteiger, J.; Rudolph, C.; Sadowski, J. Automatic Generation of 3D-Atomic Coordinates for Organic Molecules. *Tetrahedron Comput. Methodol.* **1990**, *3*, 537−547.
(26) Geerlings, P.; De Proft, F.; Langenaeker, W. Conceptual Density Functional Theory. *Chem. Rev.* **2002**, submitted for publication.
(27) Olah, J.; Van Alsenoy, C.; Sannigrahi, A. B. Condensed Fukui Functions Derived from Stockholder Charges: Assessment of Their Performance as Local Reactivity Descriptors. *J. Phys. Chem. A* **2002**, *106*, 3885−3890
(28) Pollack, S. K; Devlin, J. L.; Summerhays, K. D.; Taft, R. W.; Hehre, W. J. The site of protonation in aniline. *J. Am. Chem. Soc.* **1977**, *99*, 4583−4584.

CI0255883