Toward an Alternative Hardness Kernel Matrix Structure in the Electronegativity Equalization Method (EEM)

J. Chaves,[†] J. M. Barroso,[†] P. Bultinck,*[,‡] and R. Carbó-Dorca*[,†,‡]

Institute of Computational Chemistry, University of Girona, Campus Montilivi, 17071 Girona, Catalonia, Spain, and Department of Inorganic and Physical Chemistry, Ghent University, Krijgslaan 281, B-9000 Gent, Belgium

This study presents an alternative of the Electronegativity Equalization Method (EEM), where the usual Coulomb kernel has been transformed into a smooth function. The new framework, as the classical EEM, permits fast calculations of atomic charges in a given molecule for a small computational cost. The original EEM procedure needs to previously calibrate the different implied atomic hardness and electronegativity, using a chosen set of molecules. In the new EEM algorithm half the number of parameters needs to be calibrated, since a relationship between electronegativities and hardnesses has been found.

INTRODUCTION

Atomic charges play an important role in chemistry; very often they are used to rationalize observations, trends in molecular reactivity, and chemical properties in general. Furthermore, they can be used as descriptors in Quantitative Structure Activity Relationship (QSAR) studies to predict molecular properties such as biological activity and physicochemical parameters. Atomic charges can be obtained from ab initio calculations in a multitude of ways, e.g., through basis function based methods such as the widely used Mulliken population analysis,[1] splitting of 3-D Cartesian space in a molecule (see refs 2 and 3), and many other methods. A discussion of different types of atomic charges can be found in ref 4. The Mulliken population analysis has been shown to be derivable from a Hermitian operator (see ref 5) and has been used widely in quantum chemical studies despite some drawbacks (see ref 4). However, any kind of atomic charges requires time-consuming ab initio calculations to obtain the electron density, followed by one or more manipulations to obtain the atomic charges from the density. The Mulliken approach is in this perspective one of the fastest since it requires only simple algebraic manipulations. Still, even Mulliken charges cannot be calculated on an ab initio level when dealing with a large amount of molecules possessing a big number of atomic centers; e.g., in pharmaceutical research. This justifies searching for an alternative method for the calculation of these charges without losing the accuracy of the ab initio calculations.

One such possible approach is the Electronegativity Equalization Method (EEM), a procedure which only needs to take into account the molecular geometry and a small set of parameters, associated with well-known chemical concepts such as atomic electronegativity and hardness.

The aim of the present paper is to calibrate atomic parameters for the calculation of EEM charges which resemble very closely ab initio calculated charges. Such an approach was previously investigated in detail by Bultinck et al.[6,7] for different charge definitions and recent work on AIM charges.[8] The difference with the present work lies in the fact that a different expression has been used for the hardness kernel which allows the reduction of the number of parameters to be calibrated to only half of the original number. This allows the further extension and application of EEM.

In this paper, a brief review of EEM algorithm will be presented first. Some properties of the EEM procedures will be studied next and the nature of the chosen kernel and the consequences of this choice. Then, the algorithm is tested, and, finally, charges are computed for a well-known set of steroid molecules and results compared with ab initio data.

THEORETICAL DEVELOPMENT

Essentially, all implementations of EEM are based on Sandersons[9,10] principle stating that, when atoms unite to form a molecule or crystal, the attached atomic electronegativities χ_α^* will become equalized to the same common molecular electronegativity, χ_{eq}. The algorithm based on the Sanderson principle, which will be used as starting point here, is the EEM described by Mortier et al.[11] These authors have shown that for each atom α of an N-atom molecule, M, one can write

$$\forall\, \alpha \in M{:}\chi_\alpha = \chi_{eq} = \chi_\alpha^* + 2\eta_\alpha^* q_a + \sum_{\beta\neq\alpha}^{N} q_\beta R_{\beta\alpha}^{-1} \quad (1)$$

In this equation $R_{\alpha\beta}$ represents the interatomic distance between atoms α and β; q_α is the atomic charge on atom α; χ_{eq} is the equalized molecular electronegativity, and, finally, χ_α^* and η_α^* are the effective electronegativity and hardness of atom α in molecule M.

In the present work, instead of the classical eq 1 the following alternative equation form will be considered

$$\forall \alpha \in M{:}\ \chi_\alpha = \chi_{eq} = \chi_\alpha^* + \sum_{\beta\neq\alpha}^{N} q_\beta \varphi_{\beta\alpha} \quad (2)$$

* Corresponding author e-mail: Patrick.Bultinck@UGent.be (P.B.) and quantumqsar@hotmail.com (R.C.D.).
† University of Girona.
‡ Ghent University.

where the hardness kernel function $\varphi_{\beta\alpha}$ when applied as in eq 1 for the diagonal terms was originally associated to $2\eta_\alpha^*$ and the Coulomb term $R_{\alpha\beta}^{-1}$ was the off-diagonal hardness kernel. Because of the purpose of this study, from now on, the corresponding hardness kernel function $\varphi_{\beta\alpha}$ term will be linked to a convenient well behaved function, which can be chosen as producing a finite result at zero interatomic distances. In this way, using eq 2 all elements of the hardness kernel matrix can be fused into a unique function expression. Bultinck et al.[8] and Menegon et al.[12] have previously also studied such formalism, where the Klopman–Ohno–Mataga–Nishimoto[13] formula was used. It was found by Bultinck et al.[8] that this approach did not yield sufficient computational advantages to justify the extra cost of using the Klopman–Ohno–Mataga–Nishimoto expression instead of the electrostatic approach of eq 1. In an interesting study, Njo et al.[14] also aimed at extending and simplifying the EEM. Their attempt was to reconcile the advantage of the Qeq[15] scheme, namely the possibility to take shielding between overlapping densities into account, with the computational efficiency of the EEM scheme by Mortier et al.[11] The Qeq scheme is less efficient because of its iterative nature and because of the necessity to have parameters for each combination of two atom types or elements. If one wants to develop a Qeq scheme for N elements, the number of parameters becomes $^1/_2N(N+1)$. To avoid this extra complexity, Njo et al. replaced the explicit calculation of the parameters describing the interaction between two atomic charge clouds by a simple geometrical mean of monoatomic terms. In this sense, their approach is somewhat similar in philosophy to ours, although here a different kernel is used which is computationally more efficient.

Following the work of Bultinck et al.[16,17] it can be seen that it is easily feasible to transform eqs 1 or 2 into a compact matrix form

$$\begin{bmatrix} \mathbf{E}_0 & |\mathbf{1}\rangle \\ \langle\mathbf{1}| & 0 \end{bmatrix}\begin{bmatrix} |\mathbf{q}\rangle \\ -\chi_{eq} \end{bmatrix} = \begin{bmatrix} |\chi\rangle \\ Q \end{bmatrix} \tag{3}$$

where $|\mathbf{1}\rangle$ is the unity column vector, whose elements are the real unit, the vector $|\mathbf{q}\rangle$ contains the molecular charges, Q is the total molecular charge, the vector $|\chi\rangle$ bears the electronegativities, and, finally, the $N \times N$ matrix $\mathbf{E}_0$, called the hardness kernel matrix, with the following structure

$$\mathbf{E}_0 = \{e_{0;\alpha\beta}\} = \{\delta(\alpha = \beta)2\eta_\alpha^* + \delta(\alpha \neq \beta)\varphi(R_{\alpha\beta})\} \tag{4}$$

where logical Kronecker delta symbols[18-20] have been used for convenience.

From eq 3, a relationship can be obtained, starting from effective electronegativities and hardnesses and leading to atomic charges. The following expressions have been developed by Bultinck et al.,[16-17] for producing the charge vector $|\mathbf{q}\rangle$ and the molecular electronegativity χ_{eq}, in terms of the hardness kernel matrix $\mathbf{E}_0$ and the electronegativity vector $|\chi\rangle$:

$$|\mathbf{q}\rangle = [\langle\mathbf{E}_0^{-1}\rangle^{-1}(\mathbf{E}_0^{-1}\mathbf{1}_0\mathbf{E}_0^{-1}) - \mathbf{E}_0^{-1}]|\chi\rangle +$$
$$Q\langle\mathbf{E}_0^{-1}\rangle^{-1}\mathbf{E}_0^{-1}|\mathbf{1}\rangle \tag{5}$$

$$\chi_{eq} = \langle\mathbf{E}_0^{-1}\rangle^{-1}(Q + \langle\mathbf{1}|\mathbf{E}_0^{-1}|\chi\rangle) \tag{6}$$

For convenience, the special notation for the summation of the matrix elements[18-20] is used, that is

$$\forall \mathbf{A} = \{a_{IJ}\}:\langle\mathbf{A}\rangle = \sum_I \sum_J a_{IJ} \tag{7}$$

To introduce the new hardness kernel expression, the application of an inward product function over a $N \times N$ matrix is defined as[18-20]

$$f\{\mathbf{A}\} = \{f_{\alpha\beta}(a_{\alpha\beta})\} \tag{8}$$

where the function acts on each element of the matrix producing in this way another $N \times N$ matrix.

Introducing the interatomic distance matrix for the molecule

$$\mathbf{R} = \{R_{\alpha\beta}\} \tag{9}$$

one can choose as a function in the inward matrix product the hyperbolic secant function

$$\mathbf{S} = \text{sech}\{\sigma\mathbf{R}\} = \{S_{\alpha\beta} = \text{sech}(\sigma R_{\alpha\beta})\} \tag{10}$$

The resulting matrix (10) will be called the hyperbolic secant distance matrix (HSDM). σ is a parameter, whose value is obtained by comparing the hyperbolic secant function with the inverse distance function as shown in Figure 1, within a given interatomic range between 1.25 and 2.75 au. The best σ parameter value was found to be 0.65 au^{-1}.

The kernel function used here was chosen as the product of the hyperbolic secant and the geometric mean of the atomic hardnesses η_α^* and η_β^*, in the following way:

$$\mathbf{E}_0 = \{e_{0;\alpha\beta} = 2\sqrt{\eta_\alpha^*\eta_\beta^*}\text{sech}(\sigma R_{\alpha\beta})\} \tag{11}$$

Such a kernel possesses good properties. The first one is that this kernel takes into account the overlapping effects of the neighboring atoms, and the second is that it decreases faster than the Coulombic kernel and that it can express the shielding of the nearest atoms. Other works that try to take into account the effect of the shielding are one made by Menegon et al.[12] and another by Njo et al.[14] Equation 11 is constructed with a mathematical aim, i.e., good mathematical behavior in three ranges of distances. First, it is required that in zero it does not tend to infinity. Between 1.5 and 3.0 au, it is chosen so that the hyperbolic secant function and the Coulomb kernel give the same values. Finally, it is required that for large distances, the hyperbolic secant function decreases stronger than an inverse one due to the effect of shielding. Basically, it is based on a heuristic foundation.

Another comment is that by using eq 11 in an inverse way it may be possible to obtain geometries from the atomic hardnesses. However, it must be considered that the necessary parameters are obtained by a calibration of charges only. Although the parameters allow obtaining very good charges compared to actual DFT calculations, this does not necessarily mean that all other molecular properties are obtained with good accuracy. As a consequence, expression 11 can only be expected to give crude molecular geometries.
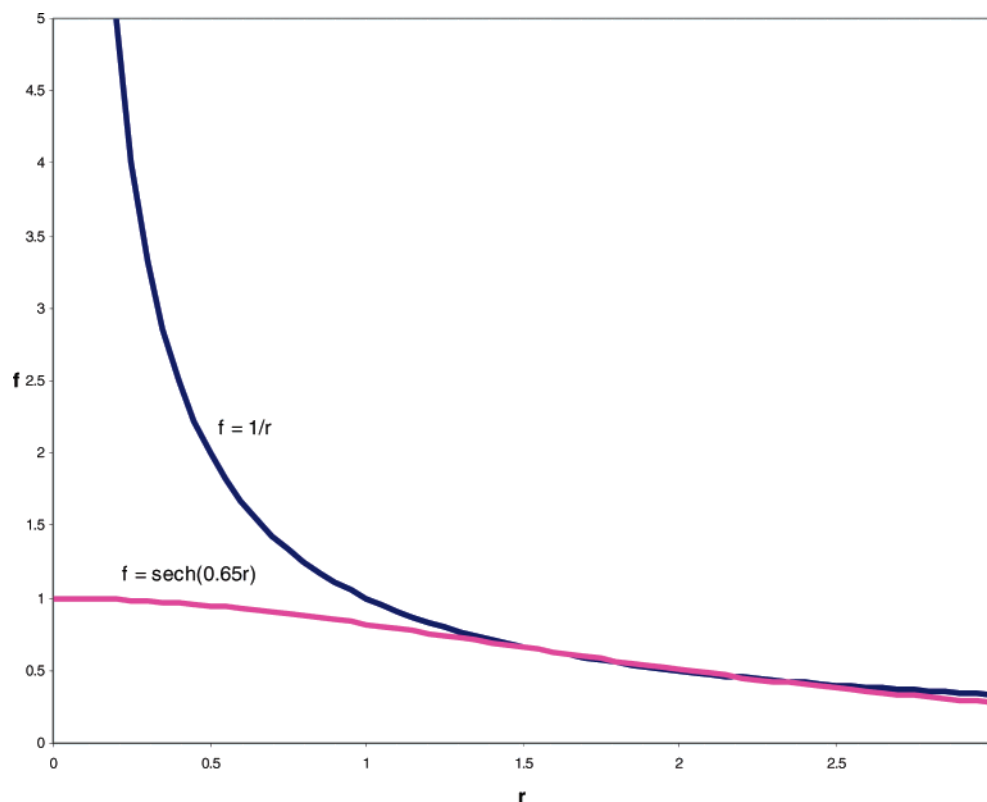
Figure 1. Comparison between hyperbolic secant and inverse distance.

The algorithm with the hardness kernel (10) is efficient and fast, because the hardness kernel matrix admits a useful decomposition into a diagonal matrix Γ and an $N \times N$ matrix **S**. That is

$$\mathbf{E}_0 = \Gamma \mathbf{S} \Gamma \qquad (12)$$

where

$$\Gamma = \sqrt{2}\,\text{Diag}(\sqrt{\eta_1^*};\sqrt{\eta_2^*};...\sqrt{\eta_N^*}) \qquad (13)$$

and **S** is the HSDM as described in eq 10.

Another interesting property of the use of the hardness kernel matrix form (11) is that all values of electronegativity and hardness can be multiplied by a constant, while the EEM charges remain invariant. This is easily seen looking at eqs 2 and 11. Whenever electronegativities and hardnesses are multiplied by the same constant relationship 2 holds in the same form, and the equalized electronegativity is scaled by the same constant.

The kernel decomposition (12) permits to enunciate a useful property, which can be described as follows: The matrix $\mathbf{E}_0$ is positive definite if and only if, **S** *is positive definite too.*

If **S** is positive definite, a circumstance which is accomplished whenever the Gerschgorin[21] radii fulfill

$$\forall i: \sum_{j \neq i} \text{sech}(\sigma R_{ij}) < 1 \qquad (14)$$

then a Cholesky decomposition[22] can be obtained as

$$\exists \mathbf{T}: \mathbf{S} = \mathbf{T}^T \mathbf{T}$$

and eq 12 can be written in the following manner:

$$\mathbf{E}_0 = \Gamma \mathbf{T}^T \mathbf{T} \Gamma$$

As Γ is diagonal, it can be also written

$$\mathbf{E}_0 = (\Gamma \mathbf{T})^T \Gamma \mathbf{T}$$

so in this way, another matrix can be defined in the form

$$\mathbf{B} = \Gamma \mathbf{T}$$

Finally, **B** is a triangular matrix consequently

$$\mathbf{E}_0 = \mathbf{B}^T \mathbf{B} \qquad (15)$$

This means that $\mathbf{E}_0$ is positive definite as there exists a Cholesky decomposition[22] (15), thus $\mathbf{E}_0$ is regular.

Atom condensed Fukui functions can be valuable tools in order to assess the good behavior of the parameters to be calibrated in EEM procedures. A résumé of the Fukui functions within EEM definitions and properties will be given now. Consider the algebraic expression for the Fukui function vector as deduced by Bultinck et al.:[23]

$$|\mathbf{f}\rangle = \langle \mathbf{E}_0^{-1}\rangle^{-1} \mathbf{E}_0^{-1}|1\rangle \qquad (16)$$

The Fukui functions can also be calculated using the hardness kernel matrix (11). It can be seen that the values of the condensed Fukui function vector (16) belong to the open interval: (0,1).

Using the definition in eq 16, eqs 5 and 6 become

$$|\mathbf{q}\rangle = \langle \mathbf{E}_0^{-1}\rangle \chi_{eq}|\mathbf{f}\rangle - \mathbf{E}_0^{-1}|\chi\rangle \qquad (17)$$

and

$$\chi_{eq} = \langle \mathbf{E}_0^{-1} \rangle^{-1} Q + \langle \mathbf{f} | \chi \rangle \qquad (18)$$

Equation 17 could give a simple insight of the contribution of each atom and their neighbors to the total charge of every atom inside the molecule, since Fukui functions can be considered a discrete probability distribution.

Calculations of EEM based Fukui functions have been performed by Bultinck et al.,[23−24] showing the importance of diagonal dominance of the hardness kernel matrix.

For neutral molecules, meaning that $Q = 0$ in eqs 5 and 6, one has[16]

$$|\mathbf{q}\rangle = \langle \mathbf{E}_0^{-1} \rangle \chi_{eq} |\mathbf{f}\rangle - \mathbf{E}_0^{-1} |\chi\rangle \qquad (19)$$

and

$$\chi_{eq} = \langle \mathbf{f} | \chi \rangle \qquad (20)$$

Equation 19 is essentially a function of the hardness kernel matrix inverse, as one can obviously rename the involved inverse as

$$\mathbf{Z} = \langle \mathbf{E}_0^{-1} \rangle^{-1} (\mathbf{E}_0^{-1} \mathbf{1}_0 \mathbf{E}_0^{-1}) - \mathbf{E}_0^{-1} \qquad (21)$$

Thus, in the neutral molecular case the following linear equation holds

$$|\mathbf{q}\rangle = \mathbf{Z} |\chi\rangle \qquad (22)$$

The molecular charges vector given in eq 22 for neutral molecules could be very interesting if it could be used to express electronegativities as

$$|\chi\rangle = \mathbf{Z}^{-1} |\mathbf{q}\rangle \qquad (23)$$

meaning that the matrix $\mathbf{Z}$ must be nonsingular. However, practical calculations have shown that this matrix was always singular. Thus, to prove that matrix $\mathbf{Z}$ is singular for every molecular structure, irrespective of the nature of the kernel matrix, one should show the existence of a zero eigenvalue for any matrix $\mathbf{Z}$. To prove that matrix $\mathbf{Z}$ always has a zero eigenvalue, one can employ the unity vector $|\mathbf{1}\rangle$ and use the projector form of the unity matrix appearing in the definition (22), that is

$$\mathbf{1}_0 = |\mathbf{1}\rangle\langle\mathbf{1}| \qquad (24)$$

then eq 21 can be rewritten as

$$\mathbf{Z} = \langle \mathbf{E}_0^{-1} \rangle \mathbf{E}_0^{-1} |\mathbf{1}\rangle\langle\mathbf{1}| \mathbf{E}_0^{-1} - \mathbf{E}_0^{-1} \qquad (25)$$

Matrix (25) can be transformed taking into account the expression of Fukui functions (16)

$$\mathbf{Z} = \langle \mathbf{E}_0^{-1} \rangle |\mathbf{f}\rangle\langle\mathbf{f}| - \mathbf{E}_0^{-1} \qquad (26)$$

using the unity vector $|\mathbf{1}\rangle$ as a probe eigenvector of the matrix $\mathbf{Z}$ as given in expression 26, it is readily shown that it really corresponds to a zero eigenvalue, that is

$$\mathbf{Z}|\mathbf{1}\rangle = 0|\mathbf{1}\rangle \qquad (27)$$

Condition (27) is always fulfilled by any matrix $\mathbf{Z}$, thus this kind of matrices is always singular and eq 23 cannot be set up.

Taking into account the first relationship provided by eq 3 and considering that eq 22 can never be invertible, then the electronegativity vector can be obtained from known DFT charges by the expression

$$\mathbf{E}_0 |\mathbf{q}\rangle = -|\chi\rangle + \chi_{eq} |\mathbf{1}\rangle = -|X\rangle \qquad (28)$$

where $|X\rangle$ are atomic shifted electronegativities, that is, atomic electronegativities displaced from the origin by the global equalized electronegativity χ_{eq}. Then, as the global equalized electronegativity is not known, necessarily some origin of electronegativities has to be defined. But, in fact, the sum or subtraction of a constant in the electronegativity values cannot affect the EEM charge values because if one adds a constant to all effective electronegativities, the equalized electronegativity is simply affected by the same constant, as can be seen from eq 20. These changes are compensated in eq 28, due to the different sign of the electronegativity vector and the global electronegativity, the value of $|X\rangle$ remains invariant, so the charge vector is invariant with respect to an arbitrary displacement of the origin of the electronegativities. The values of atomic shifted electronegativities are taken as constant parameters for each atomic species independently of the molecule. The transferability of some properties of atoms or functional groups inside of any molecule is used as proven, for example by Ayers[25] who studied the concept of transferability of atoms in molecules. Although the atom in a molecule electronegativities certainly vary from molecule to molecule, the transferability means that a constant value can be used. The variance of the electronegativity of every atom is related to the spread of the valencies of that atom in the calibration set. In the present calibration set the variance is the largest for carbon, consistent with the fact that the carbon atoms show the widest range of charges and molecular surroundings. No general conclusion can, however, be drawn concerning the variance of atomic electronegativities. A different calibration set may result in different conclusions, so general chemical conclusions should not be drawn from EEM calibrations, except if a very large and very diverse calibration set would be available.

In any molecular case, the charge vector is subject for every molecular set to an obvious constraint, namely

$$\langle \mathbf{1} | \mathbf{q} \rangle\rangle = Q \qquad (29)$$

This property is extensible to any molecular effective charge vector $|\mathbf{q}'\rangle$ computed within the same situation, irrespective to the charge definition used. That is, the following equivalent constraint will also hold:

$$\langle \mathbf{1} | \mathbf{q}' \rangle\rangle = Q \qquad (30)$$

As a consequence of these constraints imposed to any molecular effective charge vector, the following result will always hold for neutral molecules ($Q=0$):

Any approximate linear relationship between two effective charge vectors $|\mathbf{q}\rangle \wedge |\mathbf{q}'\rangle$ for neutral molecules, $\langle\mathbf{1}|\mathbf{q}\rangle\rangle = \langle\mathbf{1}|\mathbf{q}'\rangle\rangle = 0$, and computed by means of different theoretical sources can only be written as $|\mathbf{q}\rangle = \alpha|\mathbf{q}'\rangle$.

To see this, when seeking for a linear relationship between both discrete effective charge distributions, of the sort

$$|\mathbf{q}\rangle = \alpha|\mathbf{q}'\rangle + \beta|\mathbf{1}\rangle$$

ELECTRONEGATIVITY EQUALIZATION METHOD

J. Chem. Inf. Model., Vol. 46, No. 4, 2006 **1661**

where $\{\alpha,\beta\}$ are the parameters associated with the linear function to be obtained and $|1\rangle$ is the auxiliary unity vector; that is, in a N atom molecular problem, this is the same as to write $\langle|1\rangle\rangle = N$. A least squares procedure produces

$$\langle|\mathbf{q}\rangle\rangle = \alpha\langle|\mathbf{q}'\rangle\rangle + \beta N$$

The constraints 29 and 30, in the neutral case, yield $\beta N = 0 \rightarrow \beta = 0$, and, thus, the linear relationship between two effective charge vectors could only be written as $|\mathbf{q}\rangle = \alpha|\mathbf{q}'\rangle$.

With this linear relation, for neutral molecules, EEM charges differ from ab initio charges only by a scale factor.

COMPUTATIONAL METHODS

Invariant Transformations. It has already been commented that two kinds of transformations of the EEM parameters to be calibrated leave the charge vector invariant. The first one, a scaling transformation, means that the set of electronegativities and hardnesses can be multiplied by a constant without affecting the charges. Then, the values of hardnesses and electronegativities can be referred to any atom as unit. In this paper the hydrogen hardness has been chosen as reference having a unity value.

The second transformation is an origin shift transformation, meaning that one can add a constant to the electronegativities, leaving the charges invariant.

Using the proposed hardness kernel (11), a reduction in the number of EEM parameters to be calibrated is implicit in this choice. The calibration requires only one parameter per atom instead of two, as usual in preceding works. Finally another parameter is removed by the scale transformation, and the hydrogen hardness remains equal to one.

Training Set. The training set chosen for EEM parameter calibration consists of 138 noncharged molecules containing H, C, N, O, and F, which is the same set as used by Bultinck et al.[6-8] for the same purpose within a Coulomb kernel. This set has been chosen as it includes the most important functional groups used in medicinal chemistry. This fact ensures a good calibration in order to predict charges with some QSAR usefulness in pharmaceutical applications. The molecules in Bultinck's set provide a calibration with 930 H, 602 C, 105 N, 101 O, and 65 F atoms.

Optimization Algorithms. A Genetic Algorihtm[26,27] (GA) has been used to fit the four remnant parameters in Bultinck's set when using the hardness kernel (11); the calibration is thus performed for the four relative hardnesses of C, N, O, and F. The optimization calculated by the GA involves 20 individuals for at least 500 generations. Individuals are taken as the four hardnesses which were allowed to vary within an interval from 0 to 100. Children are generated by uniform parental crossover. Probability of mutation is placed to 0.02. Elitism (selection of the best individuals) and niching (sharing) are activated. GA is an efficient way to find the global maximum but converges slowly to it. Then, after an optimal calibration set was found, a Simplex Algorithm[28] was used to refine the result. This is the same procedure that Menegon et al.[12] have used, but with another hardness kernel. The Simplex procedure converges in few iterations after the GA optimization step.

Calibration Procedure. The first step consists of the calibration of atomic hardness, because the electronegativity will be determined using relationship 28 between $|\mathbf{q}\rangle$ and

Table 1. Relative Parameters Calibrated Using SFKEEM

atom	χ_{rel}	η_{rel}
H	−0.2131	1.0000
C	0.0056	0.3481
N	0.2378	0.4840
O	0.5318	0.8384
F	0.5808	1.1929

Table 2. Transformation of SFKEEM Calibrated Parameters in Order To Compare Them with Bultinck et al.[6] Values[a]

atom	present work		Bultinck et al.	
	χ(eV)	η(eV)	χ(eV)	η(eV)
H	1.00	17.95	1.00	17.95
C	4.92	6.25	5.25	9.00
N	9.10	8.69	8.80	9.39
O	14.37	15.04	14.72	14.32
F	15.25	21.36	15.00	19.77

[a] Table 1 parameters are multiplied by 17.95 eV (H atomic hardness of Bultinck et al.), and the electronegativities are afterward shifted taking into account that H electronegativity becomes unity.
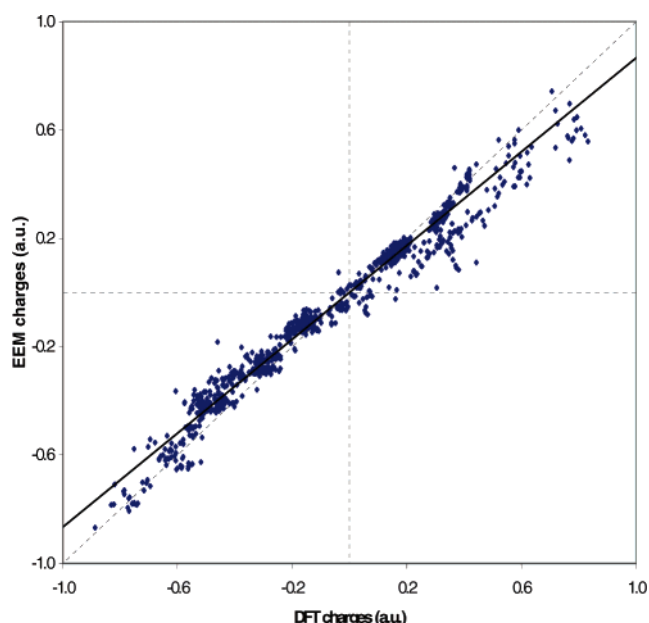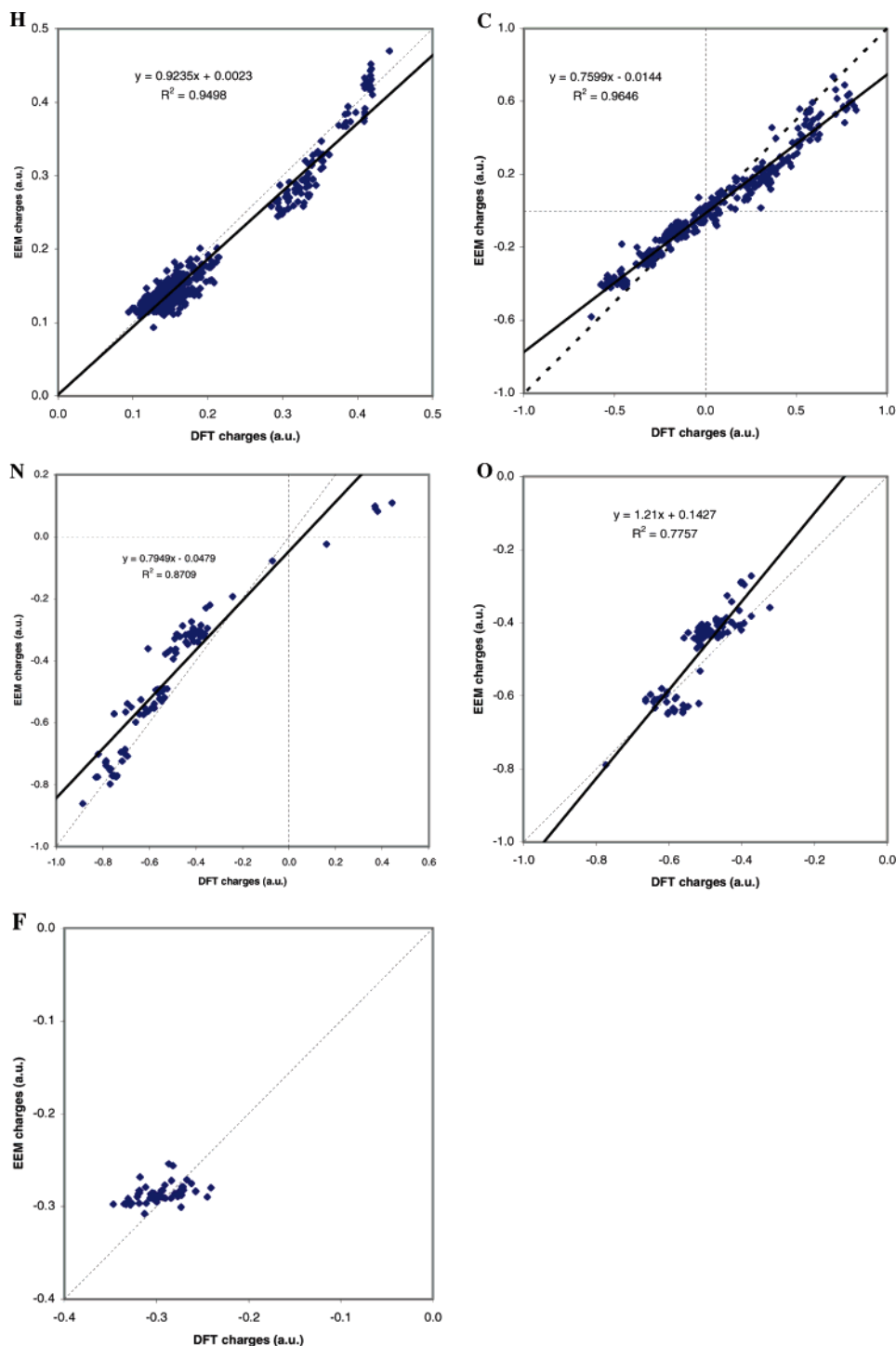

Figure 2. Bultinck's molecular set. EEM charges using SFKEEM vs DFT Mulliken charges. Linear fit: $y = 0.87x$ with $R = 0.986$ and $\sigma^2 = 0.076$.

$|\chi\rangle$, so the iteration process for the nth step consists of the following steps. First, guess atomic hardnesses and calculate their respective electronegativities using the relationship 28 with Mulliken charges calculated at the B3LYP/6-31G* level using Gaussian 98 program.[29] These will be called DFT charges.

Once all electronegativities are calculated, an arithmetic mean over the obtained values is done in order to get a unique value of electronegativity for each atom type.

Averaged atomic electronegativities are used with the optimized atomic hardnesses to calculate a EEM charge vector using eq 22. The correlation of the EEM charges versus DFT charges is computed, and when the GA stops, the parameters are refined using a Simplex Algorithm in order to reach a refined maximum correlation.

In eq 16 the relationship between hardnesses and the Fukui functions is obvious, so the hardness calibration process has

Figure 3. Comparison of EEM and DFT Mulliken charges for all atoms in the Bultinck's molecular set (separate correlations are shown for H, C, N, and O; F is considered as a point with dispersion).

to take into account this point. To introduce this relationship in the evaluation and optimization of atomic hardnesses, the correlation between the EEM charges, calculated employing the hardness matrix and DFT charges, is divided by the number of negative Fukui functions appearing at each step. Such a correction is implemented in this manner because results with negative Fukui functions can be seen as not adequate to the previous definitions (see above) and in that way, when reaching the optimal calibration step, all Fukui functions are positive. If the sign of the electronegativities are left free, the Pearson's correlation increases to 0.9911, but the number of negative Fukui functions is 410 among

1840. Then one can decide if the fit is more important than the sign of the atomic condensed Fukui functions. There are many papers about the discussion of the sign of Fukui functions, but the convexity of the Fukui vector is adopted in this paper, although one can easily relax this condition in the implementation.

All these procedures will be called the Selfconsistent Functional Kernel Equalized Electronegativity Method (SF-KEEM) since the calibration is performed using ab initio charges and hardnesses as parametrization, electronegativities are calculated using both these sets of values ones in a self-consistent way.

ELECTRONEGATIVITY EQUALIZATION METHOD

J. Chem. Inf. Model., Vol. 46, No. 4, 2006 **1663**
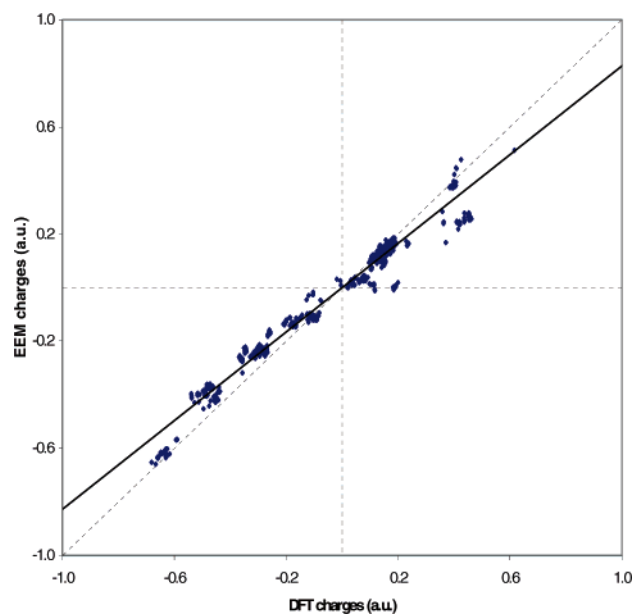
Figure 4. Cramer steroid set. EEM charges using SFKEEM vs DFT Mulliken charges. Linear fit: $y = 0.83x$ with $R = 0.984$ and $\sigma^2 = 0.057$.

RESULTS AND DISCUSSION

Once the calibrated parameters are obtained, and the values of the EEM charges, we proceed to discussing the fitness of these new charges. One of the main properties of EEM

consists of the fact that charges are computed, at least, a factor 1000 faster than DFT calculations using an ab initio program. Another point is that EEM charges exhibit the same trends as DFT charges, so they can be used in the same way as the DFT charges for recognizing the active molecular zones or to predict other kinds of properties as in QSAR.

The calibrated EEM parameters obtained in this work (see Table 1) seem to be, at a first glance, different from the ones reported before[6,7] (see Table 2). However, present EEM parameters are all related to the reference value of the H atomic hardness. When comparing the optimal parameters with other results, electronegativities and hardnesses must be scaled with the same factor, and electronegativities may be shifted to some common hydrogen electronegativity value.

Figure 2 represents a linear regression of the EEM charges versus DFT charges for all atoms in the molecular set. The distribution of all present points is shown in the first and third quadrants of the figure; there EEM charges have approximately the same values as the DFT ones. Charges which differ in sign have almost zero values, so results can be considered accurate enough.

With respect the work of Bultinck et al.,[6,7] the correlations of separated atoms are similar to the present work. They are improved for hydrogen and carbon and remain equal for nitrogen and oxygen. The linear fit for the fluorine species has not been made because it can be considered as a dot with dispersion, having a value of approximately −0.35 u.a. The main contribution to the global correlation is made by
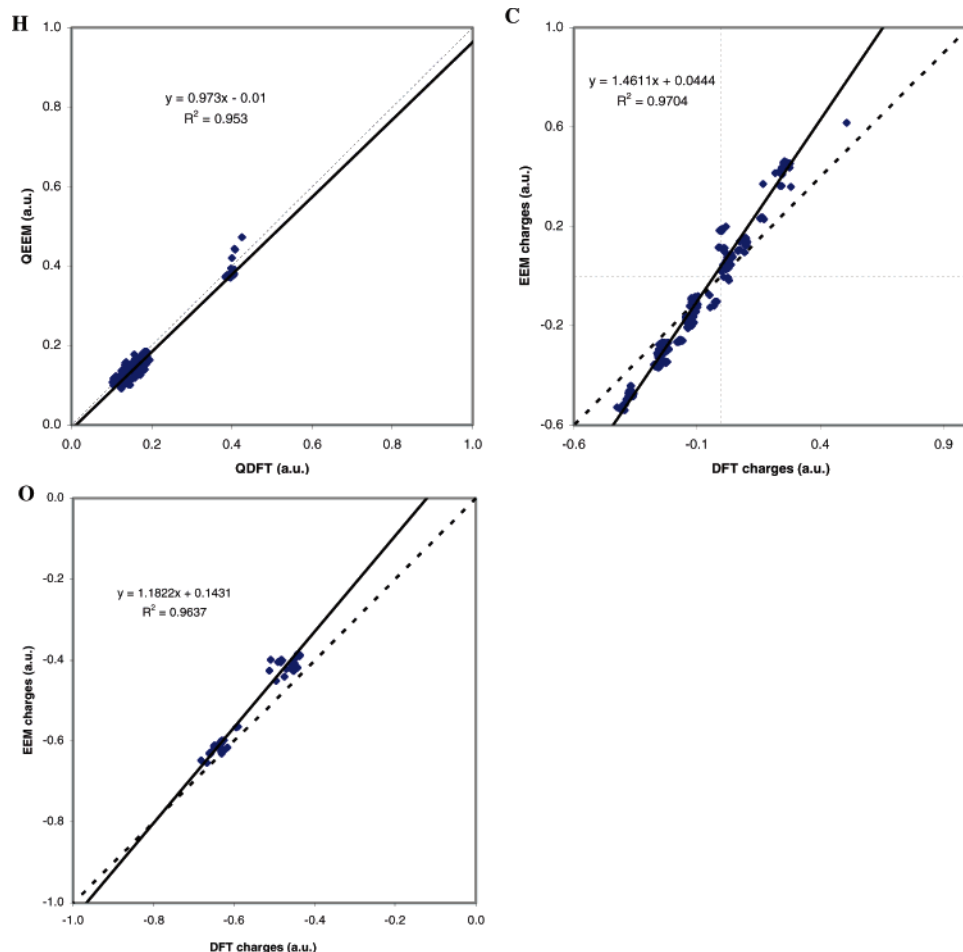
Figure 5. EEM charges using SFKEEM versus DFT Mulliken charges for all atoms in the Cramer steroid set (separate correlations are shown for H, C, and O; there is only one point for F with a value of −0.357 au).

1664 *J. Chem. Inf. Model., Vol. 46, No. 4, 2006*

CHAVES ET AL.

the carbon species, since it has the biggest number of individuals and can get any sign. The linear fit for hydrogen is good enough, but it seems to be localized in different sets, depending on the atom it is bound to. In the case of nitrogen, the charges are near the linear fit for negatives charges, but positive ones deflect somewhat more from the tendency of the majority. The spread of oxygen charges displays an acceptable trend too. Basically, one can conclude that the linear fit is globally good and, for each species, depends on the number of individuals.

After the described calibration, a prediction was set up in order to check the applicability of the SFKEEM as well as its accuracy. The nature of the chosen molecular test set, called the Tripos or Cramer set[30] with refined structures in our laboratory,[31] is different enough from the training set in order to test the behavior of the calibrated parameters. Nitrogen is not present in this set, so this contribution to the correlation was not taken into account. At first sight, this could seem to be a problem, but the results show that the prediction is correct even in the absence of one species. There is only one fluorine, but its value is approximately the same obtained in the calibration (−0.35 u.a.), so this charge has been well predicted. The results are presented in Figure 4, the correlation between EEM and DFT charges is quite good, and the dispersion sufficiently low. Figure 4 shows that the prediction for the hydrogen, carbon, and oxygen atoms is not a crude approximation; in fact it is an excellent prediction in the sense of fast calculations. This is especially true when one needs the trend of the charges and not the exact numerical value. Furthermore, there are a lot of definitions for atomic charges in the ab initio realm. The differences between EEM and DFT charges are of the same order as the differences in charges between different atomic charge definitions.

The results of the test set prove that the parameters obtained in the present work are applicable to another set of molecules, obtaining accurate results while avoiding the use of large amounts of computer time to do actual ab initio calculations.

CONCLUSIONS

In this work the SFKEEM algorithm gives a similar parameter set as in the Bultinck et al.[6,7] procedure. The main difference consists of the fact that the number of free calibration parameters is reduced from 10 to 4 as the electronegativities and hardnesses are expressed in units of H hardness. The described EEM formalism is quite different as well, showing that it has been possible to integrate the whole kernel matrix into a unique expression. Finally, the actual values of the atomic electronegativity depend on a global equalized electronegativity of the molecule, but the electronegativity values can be shifted when it is required, without changing the charge vector. All these aspects configure the described procedure as a useful tool to find atomic electronegativities and hardnesses. Such calibrated parameters are useful for ultrafast calculations of molecular charge vectors which do not differ from ab initio values.

ACKNOWLEDGMENT

REFERENCES AND NOTES

(1) Mulliken, R. S. Criteria for the Construction of Good Self-Consistent-Field Molecular Orbital Wave Functions, and the Significance of LCAO-MO Population Analysis. *J. Chem. Phys.* **1962**, *36*, 3428−3439.

(2) Bader, R. F. W. *Atoms in Molecules*; Clarendon Press: Oxford, 1990.

(3) Bader, R. F. W. A Quantum Theory of Molecular Structure and its Applications. *Chem. Rev.* **1991**, *91*, 893−928.

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

(5) Carbó-Dorca, R.; Bultinck, P. Quantum Mechanical Basis for Mulliken Population Analysis. *J. Math. Chem.* **2004**, *36*, 201−210.

(6) Bultinck, P.; Langenaeker, W.; Lahorte, P.; De Proft, F.; Geerlings, P.; Waroquier, M.; Tollenaere, J. P. The Electronegativity Equalization Method I: Parametrization and Validation for Atomic Charge Calculations. *J. Phys. Chem. A* **2002**, *106*, 7887−7894.

(7) Bultinck, P.; Langenaeker, W.; Lahorte, P.; De Proft, F.; Geerlings, P.; Van Alsenoy, C.; Tollenaere, J. P. The Electronegativity Equalization Method II: Applicability of Different Atomic Charge Schemes. *J. Phys. Chem. A* **2002**, *106*, 7895−7901.

(8) Bultinck, P.; Popelier, P.; De Proft, F.; Geerlings, P. High-Speed Calculation of AIM Charges Through the Electronegativity Equalization Method. *J. Phys. Chem. A* **2004**, *108*, 10359−10366.

(9) Sanderson, R. T. An Interpretation of Bond Lengths and a Classification of Bonds. *Science* **1951**, *114*, 670−672.

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

(11) 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.

(12) Menegon, G.; Shimizu, K.; Farah, J. P. S.; Dias, L. G.; Chaimovich, H. Parameterization of the Electronegativity Equalization Method Based on the Charge Model 1. *Phys. Chem. Chem. Phys.* **2002**, *4*, 5933−5936.

(13) Mataga, N.; Nishimoto, K. Electronic Structure and Spectra of Nitrogen Heterocycles. *Z. Phys. Chem.* **1957**, *13*, 140−144.

(14) Njo, S. L.; Fan, J.; van de Graaf, B. Extending and Simplifying the Electronegativity Equalization Method. *J. Mol. Catal. A* **1998**, *134*, 79−88.

(15) Rappé, A. K.; Goddard, W. A., III Charge Equilibration for Molecular Dynamics Simulations. *J. Phys. Chem.* **1991**, *95*, 3358−3363.

(16) Bultinck, P.; Carbó-Dorca, R. Algebraic Relationships Between Conceptual DFT Quantities and the Electronegativity Equalization Hardness Matrix. *Chem. Phys. Lett.* **2002**, *364*, 357−362.

(17) Bultinck, P.; Langenaeker, W.; Carbó-Dorca, R.; Tollenaere, J. P. Fast Calculation of Quantum Chemical Molecular Descriptors from the Electronegativity Equalization Method. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 422−428.

(18) Carbó-Dorca, R. Inward Matrix Products: Extensions and Applications to Quantum Mechanical Foundations of QSAR. *J. Mol. Struct. (THEOCHEM)* **2001**, *537*, 41−54.

(19) Carbó-Dorca, R. Shell Partition and Metric Semispaces: Minkowski Norms, Root Scalar Products, Distances and Cosines of Arbitrary Order. *J. Math. Chem.* **2002**, *32*, 201−223.

(20) Carbó-Dorca, R. Applications of Inward Matrix Products and Matrix Wave Functions to Hückel MO Theory, Slater Extended Wave Functions, Spin Extended Functions, and Hartree Method. *Int. J. Quantum Chem.* **2003**, *91*, 607−617.

(21) Cullen, C. G. *Matrices and linear transformations*; Addison-Wesley Pub. Co.: 1966.

(22) Wilkinson, J. H.; Reinsch, C. *Linear Algebra*; Springer-Verlag: New York, 1971; Vol. II.

(23) Bultinck, P.; Carbó-Dorca, R. Negative and Infinite Fukui Functions: The Role of Diagonal Dominance in the Hardness Matrix. *J. Math. Chem.* **2003**, *34*, 67−74.

(24) Bultinck, P.; Carbó-Dorca, R.; Langenaeker, W. Negative Fukui Functions: New Insights Based on Electronegativity Equalization. *J. Chem. Phys.* **2003**, *118*, 4349−4356.

(25) Ayers, P. W. Atoms in Molecules, An Axiomatic Approach. I. Maximum Transferability. *J. Chem. Phys.* **2000**, *113*, 10886−10898.

ELECTRONEGATIVITY EQUALIZATION METHOD

J. Chem. Inf. Model., Vol. 46, No. 4, 2006 **1665**

(26) Carroll, D. L. *Fortran Genetic Algorithm (Ga) Driver.* http://cuaerospace.com/carroll/ga.html (accessed 05/05/2006).

(27) Goldberg, D. E. *Genetic Algorithms in Search, Optimization and Machine Learning*; Addison-Wesley: 1989.

(28) Press, W. H.; Teulosky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in Fortran 77: The Art of Scientific Computing*, 2nd ed.; Cambridge University Press: New York, 1992.

(29) 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., R. E. S.; 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.; Salvador, P.; Dannenberg, J. J.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; iashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; aham, M. A. A.; Peng, C. Y.; Nanayakkara, A.; 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.11)*; Gaussian, Inc.: Pittsburgh, PA, 2001.

(30) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (Comfa). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.

(31) Robert, D.; Amat, L.; Carbó-Dorca, R. Three-Dimensional Quantitative Structure−Activity Relationships from Tuned Molecular Quantum Similarity Measures: Prediction of the Corticosteroid-Binding Globulin Binding Affinity for a Steroid Family. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 333−344.