J Mol Model (2011) 17:2217–2226
DOI 10.1007/s00894-011-1006-7

ORIGINAL PAPER

Charge sensitivity analysis in force-field-atom resolution

Anna Stachowicz · Anna Styrcz · Jacek Korchowiec

Received: 14 September 2010 / Accepted: 31 January 2011 / Published online: 3 March 2011

Abstract Charge sensitivity analysis (CSA) was extended to AMBER force-field resolution. The effective electronegativity and hardness data were found using evolutionary algorithms. Four model hardness matrices based on the classical electrostatic, Mataga–Nishimoto, Ohno, and Louwen–Vogt interpolation formulae were considered. Mulliken population analysis and electrostatically derived charges (CHELPG) were taken into account. It was demonstrated that the Ohno interpolation formula gives the best fit to Mulliken charges. For all molecules from the training set and all model hardness matrices, Mulliken charges were reproduced more accurately than CHELPG charges, indicating their good transferability from system to system. The effective electronegativities and hardnesses obtained were further verified by applying CSA to molecules from a validation set that was different from the training set. The correlation between CSA and Mulliken charges was of the same quality as that obtained for the training set.

Keywords Charge sensitivity analysis · Electronegativity equalization principle · Charge distribution
Softness/hardness data · AMBER force fields

Introduction

Chemical processes are frequently viewed from the atomic perspective. Unfortunately, there are various quantum-chemical definitions of atoms in molecules [1–6]. Some of these are based on partitioning in Hilbert or physical space. Mulliken [1], Löwdin [2], and Roby [3] population analyses are examples of the former, while Bader's [4] and Hirshfeld's [5] approaches are examples of the latter definition. Charge distributions in molecules can be obtained by fitting to ab initio electrostatic potentials. Many procedures were proposed that differ in the grid assumed and additional constraints [7–9]. Some parameterization procedures take charges from other population schemes as input and map them to reproduce charge-dependent observables obtained either from experiment or from very accurate quantum mechanical calculations on small molecules [10].

Using certain assumptions, charge distributions can be derived from experimental measurements. The analysis of core-level energy shifts from X-ray photoemission spectroscopy [11] or nuclear quadrupole coupling constants [12, 13] gives the charge distribution inside subsystems; however, it should be stressed that atomic charges as well as those of an interacting subsystem are not physical observables. The correlation between the charges of donor/acceptor systems obtained with different population analyses was investigated by Szewczyk et al. [14].

Another way to derive atomic charges is to use the electronegativity equalization (EE) principle [15]. The concept of electronegativity, originally introduced by Pauling [16], is commonly used to describe charge polarization within molecules. Electronegativity is rigorously defined within density functional theory (DFT) and is considered a negative chemical potential. Within the finite difference approximation, atomic electronegativities [17] are equivalent to Mulliken electronegativities [18]. Sanderson's EE principle [15] is also naturally justified by DFT.

The EE method is based on the second-order Taylor expansion of the system's energy with respect to the

Electronic supplementary material The online version of this article (doi:10.1007/s00894-011-1006-7) contains supplementary material, which is available to authorized users.

A. Stachowicz · A. Styrcz · J. Korchowiec (✉)
K. Gumiński Department of Theoretical Chemistry,
Faculty of Chemistry, Jagiellonian University,
Krakow, Poland
e-mail: korchow@chemia.uj.edu.pl

population variables. Its atomic resolution is used to derive the charge distributions within molecules. One of the first applications of the EE method for computing atomic charges was performed by Mortier et al. [19, 20]. The EE method was also incorporated into molecular dynamics packages in order to derive the starting charge distribution [21, 22]. Recently, there has been growing interest in EE-based approaches [23–29]. Some of these exhibit trivial-atom resolution. Other approaches distinguish atoms by hybridization or by their nearest neighbors. Except for the CHARMM fluctuating charge force-field for proteins [24], none of the abovementioned EE-based approaches have been applied to describe mutually polarized molecules. It was demonstrated using charge sensitivity analysis (CSA) that the EE method should also work for polarized reactants [30].

CSA was formulated in the 1990s [31]. This formalism was considered a supplementary tool to semiempirical calculations. It was based on the EE principle, but in contrast to the EE method, it was never parameterized to derive the charge distributions inside molecules. The main aim of CSA was to explore reactivity concepts that could potentially be used in qualitative structure–activity relationship (QSAR) or qualitative structure–property relationship (QSPR) models. In the present paper we would like to go beyond the original CSA formulation and parameterize CSA with force-field-atom resolution. Weiner's original AMBER84 force field [32, 33] was taken into account. The paper is organized as follows. First, the CSA formalism is described together with four model hardness matrices. Computational details are then given. Afterwards, the results obtained are presented. Finally, conclusions and future prospects are briefly discussed.

Charge sensitivity analysis

CSA at atomic resolution can be summarized as a single matrix equation [31, 34]:

$$\begin{pmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & \eta_{11} & \eta_{12} & \cdots & \eta_{1N} \\ 1 & \eta_{21} & \eta_{22} & \cdots & \eta_{2N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \eta_{N1} & \eta_{N2} & \cdots & \eta_{NN} \end{pmatrix} \begin{pmatrix} -\chi \\ q_1 \\ q_2 \\ \vdots \\ q_N \end{pmatrix} = \begin{pmatrix} q \\ -\chi_1^* \\ -\chi_2^* \\ \vdots \\ -\chi_N^* \end{pmatrix}, \quad (1)$$

where $\boldsymbol{\eta} = \{\eta_{ij} = \partial^2 E_M / \partial q_i \partial q_j\}$ is the hardness matrix, q is the total charge and $\chi = \partial E_M / \partial q$ is the global electronegativity of a molecule M composed of N atoms [17]. Vectors $\mathrm{q} = (q_1, q_2, \ldots q_N)^{\mathrm{T}}$ and $\chi = (\chi_1^*, \chi_2^*, \ldots \chi_N^*)^{\mathrm{T}}$ group the atomic charges and the electronegativities. The matrix in Eq. 1 describes a transformation from the constrained equilibrium $M = (1|2|\cdots|N)$ to the global equilibrium $M = \left(1 \vdots 2 \vdots \ldots \vdots N\right)$. The vertical solid and dotted lines indicate closed (no charge transfer among atoms) and open atoms in M, respectively. The first equation is a closure relation:

$$q = \sum_{i=1}^{N} q_i. \quad (2)$$

The remaining equations

$$\chi = \chi_i = \chi_i^* + \sum_{j=1}^{N} \eta_{ij} q_j, \quad i = 1, 2, \ldots, N \quad (3)$$

manifest the EE principle in atomic resolution ($\chi_1 = \chi_2 = \ldots = \chi_N = \chi$). Equation 3 results from the following second-order Taylor expansion of the system's energy with respect to the atomic charges:

$$dE_M = d^1E_M + d^2E_M = \sum_i \left(\frac{\partial E_M}{\partial q_i}\right) dq_i + \tfrac{1}{2} \sum_i \sum_j \left(\frac{\partial^2 E_M}{\partial q_i \partial q_j}\right) dq_i dq_j = \chi^{\mathrm{T}} d\mathbf{q} + \tfrac{1}{2} d\mathbf{q}^{\mathrm{T}} \boldsymbol{\eta} d\mathbf{q} \quad (4)$$

The charge distribution can be obtained by inverting Eq. 1 [31, 34]:

$$\begin{pmatrix} -\chi \\ q_1 \\ q_2 \\ \vdots \\ q_N \end{pmatrix} = \begin{pmatrix} -\eta & f_1 & f_2 & \cdots & f_N \\ f_1 & -\beta_{11} & -\beta_{12} & \cdots & -\beta_{1N} \\ f_2 & -\beta_{21} & -\beta_{22} & \cdots & -\beta_{2N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_N & -\beta_{N1} & -\beta_{N2} & \cdots & -\beta_{NN} \end{pmatrix} \begin{pmatrix} q \\ -\chi_1^* \\ -\chi_2^* \\ \vdots \\ -\chi_N^* \end{pmatrix}. \quad (5)$$

Here, $\boldsymbol{\beta} = \{\beta_{ij} = \partial^2 E_M / \partial v_i \partial v_j = -\partial q_j / \partial v_i = -\partial q_i / \partial v_j\}$ is the polarization matrix (the linear response matrix or the negative internal softness matrix) [31]. It fulfills the following normalization condition: $\sum_{i=1}^{N} \beta_{ij} = \sum_{j=1}^{N} \beta_{ij} = 0$, since a perturbation in the external potential (v) at the position of a given atom cannot change the number of electrons in the system. The remaining undefined quantities are the global hardness [35] $\eta = \partial^2 E_M / \partial q^2$ and the Fukui function (FF) [36] vector $\boldsymbol{f} = (f_1, f_2, \ldots f_N)^{\mathrm{T}}$. The FF indices are normalized to unity ($\sum_{i=1}^{N} f_i = 1$). Each component of $\boldsymbol{f}$ represents a response of a given atom to a perturbation in the total number of electrons in the system: $f_i = (\partial q_i / \partial q)_v = -(\partial \mu / \partial v_i)_q$. It should be stressed that FF describes the global equilibrium when there is no barrier to charge flows between atoms. Apart from describing two extreme equilibrium cases (global and constrained), CSA can be applied to intermediate situations: the division of molecular system M into mutually closed subsystems/reactants (molecule–environment, mutually polarized reactants, active center and environment, catalyst and its support, etc.).

Hardness matrix

It is obvious from Eq. 1 that in order to derive atomic charges, the effective electronegativities $\boldsymbol{\chi}$ and hardness

matrix $\boldsymbol{\eta}$ must be known. In addition, the hardness matrix is the most important second-order sensitivity. Other sensitivities (e.g., polarization matrix, softness matrix and FF vector) are derived from $\boldsymbol{\eta}$. Assuming the frozen core approximation, one can relate the off-diagonal elements of the hardness matrix to the Coulomb and exchange integrals between valence shell *s*-electrons (rotational invariance):

$$\eta_{ij} \approx \left(s_i s_i || s_j s_j\right) = \left(s_i s_i | s_j s_j\right) - \tfrac{1}{2}\left(s_i s_j | s_j s_i\right) \equiv J_{ij} - \tfrac{1}{2}K_{ij}. \quad (6)$$

The zero-differential overlap criterion, adopted in semiempirical methods, further reduces η_{ij} to the Coulomb integral J_{ij}. The diagonal elements are approximated by the finite difference method $\eta_{ii} \approx I_i - A_i. \approx J_{ii}$ (the Pariser formula), where I_i and A_i are the ionization potential and the electron affinity of atom *i*. The crudest approximation for off-diagonal elements,

$$\eta_{ij} \approx 1 / R_{ij}, \quad (7)$$

results from the point-charge formula $E_{ij} = (Z_i - N_i) \times \left(Z_j - N_j\right)/R_{ij} = q_i q_j / R_{ij}$ that is used in classical molecular dynamic simulations. Here, R_{ij} is the distance between atoms *i* and *j*, Z_i is the nuclear charge and N_i is the electron population of the *i*th atom. Such an approximation is correct in the limit of separated atoms; however, its behavior in the limit of united atoms is incorrect. A more realistic model of off-diagonal elements of the hardness matrix can be obtained from the diagonal integrals via the combination formulae of semiempirical theories, such as those of Mataga–Nishimoto (M-N) [37]

$$\eta_{ij} = \frac{1}{a_{ij} + R_{ij}}, \quad (8)$$

Ohno [38]

$$\eta_{ij} = \frac{1}{\sqrt{a_{ij}^2 + R_{ij}^2}}, \quad (9)$$

and Louwen–Vogh (L-V) [39]

$$\eta_{ij} = \frac{1}{\sqrt[3]{R_{ij}^3 + d_{ij}^3}}. \quad (10)$$

The constants a_{ij} and d_{ij} depend on diagonal hardness data and are equal to $a_{ij} = 2/\left(\eta_{ii} + \eta_{jj}\right)$ and $d_{ij} = 1/\sqrt{\eta_{ii}\eta_{jj}}$, respectively. The construction of the hardness matrix only requires that the diagonal hardnesses are known. These, together with the effective electronegativities, should be optimized for each of the atom types.

Methods

The effective electronegativity and diagonal hardness data for a few selected atoms can be obtained via a systematic search of the variation space. However, if one additionally differentiates atoms according to their nearest chemical neighborhood, the optimization is no longer an easy task. A systematic search of the variation space is meaningless due to a huge number of optimization parameters. For this reason, an evolutionary algorithm (EA) was used to find the optimal values of the effective electronegativities and diagonal hardnesses. The GAUL [40] library was used to perform EA calculations. It was coupled with the CSA package developed in our group. Calculations were carried out for a set of 101 small organic molecules. These molecules are called the training set. The total number of atoms in this set was 1639. Most of the molecules in the training set were of biological importance. In particular, standard amino acids and DNA/RNA bases were included. All of the molecules in the training set are listed in the "Electronic supplementary material" (ESM). The structures of all molecules and the charge distribution data were calculated at the HF/6-31G(d) level of theory using the GAMESS package [41].

Evolutionary algorithms mimic the natural concept of evolution in order to tackle many dimensional optimization problems. In this approach, the possible solutions are regarded as specimens of a population. They are characterized by a fitness function that describes the extent to which a given solution is different from the optimal one. For our needs, it was defined as:

$$\mathrm{S}^2 = -\sum_i \sum_{\alpha \in i} \left(q_{i,\alpha}^{HF} - q_{i,\alpha}^{CSA}\right)^2, \quad (11)$$

where $\boldsymbol{q}_i^{\mathrm{HF}} = \left\{q_{i,\alpha}^{HF},\ \alpha = 1,\ 2, \ldots\right\}$ and $\boldsymbol{q}_i^{\mathrm{CSA}} = \left\{q_{i,\alpha}^{CSA}, \alpha = 1,\ 2, \ldots\right\}$ denote vectors holding the *i*th molecule's atomic charges calculated with the HF and CSA methods, respectively. The sum in Eq. 11 is performed across the training molecules. S^2 tends to zero in the ideal case where q_i^{CSA} is identical to q_i^{HF}. The function is given a negative sign to ensure that it is increasing. For trivial and hybridized atoms, the population employed consisted of 50 entities. When the force-field atoms were introduced, it was necessary to increase the size of the population to 70 entities due to a large number of optimization parameters. The genome of each entity was a matrix of effective electronegativities (χ_i^*) and diagonal hardnesses ($\eta_{ii} \equiv \eta_i^*$). The specimens in the population evolved in

two ways: via recombination or mutation. The former mixed the corresponding allele of two entities' genomes, while the latter changed the contents of a single entity's genome. Both recombination and mutation could appear at multiple points in the genomes. The crossover took place at the beginning of each generation. Selection of the entities for recombination was conducted as follows: in each pair, the first entity was chosen to have a small absolute value of the fitness function while the other was chosen randomly. Then, all of the entities —parents and children— were mutated. Full elitism was introduced, which meant that the entities that survived to the next generation were the ones with the smallest absolute values of the fitness function, regardless of whether they were parents, children or mutants. The evolution scheme employed in this work corresponded to Darwinian evolution (i.e., no local optimization was performed on the entities during the evolution). Depending on the number of optimization parameters, the evolution lasted for 80 (trivial atoms) or 250 generations (hybridized atoms). In the case of force-field atoms, the calculations had to be split. Parameters for atom types of different elements were optimized in turn. Evolution lasted for 80, 250, 150, 80 and 50 generations for hydrogen, carbon, nitrogen, oxygen and sulfur atoms, respectively.

Results and discussion

The optimized effective electronegativity (χ_i^*) and hardness (η_i^*) data to reproduce Mulliken charges are collected in Tables 1, 2, 3. We have considered only hydrogen, carbon, oxygen, nitrogen, and sulfur atoms. Three different atomic resolutions were taken into account. First, we only distinguished the elements (Table 1). Next, we additionally distinguished between hybridizations for each atom (Table 2). Finally, AMBER force-field-atom types were considered (Table 3). All of the tables contain four sets of parameters, each for a different hardness matrix. The reported values are in eV/e or eV/e^2 for electronegativities and hardnesses, respectively. In the case of electronegativity, we assumed that χ_{H}^* and χ_{HC}^* (the first row in Tables 1–3) are equal to 10 eV/e. This constant value for one electronegativity parameter was introduced in order to simplify the optimization procedure, and has no effect on the atomic charges obtained. The only consequence of this assumption is a shift in the reference chemical potential (electronegativity) level by a constant value. The following matching rule: $\chi = (\partial E_M/\partial q) = \sum_i (\partial E_M/\partial q_i)(\partial q_i/\partial q) = \sum_i \chi_i^* f_i$ clearly demonstrates this effect. By adding a constant value to each χ_i^*, the global electronegativity is shifted by the same value, as the FF indices are normalized to 1: $\sum_i (\chi_i^* + \mathrm{const}) f_i = \chi + \mathrm{const}$. In order to adjust this level, it is necessary to either reproduce the system's electronic energy or to take into account the known relation between the first (d^1E) and the second (d^2E) differential of the system's energy for the equilibrium charge distribution. In the first case, a set of effective atomic energies E_i^* must be found in order to reproduce the system energy. In the second case, the following relation holds: $|d^2E| = \frac{1}{2}|d^1E|$.

The data reported in Table 1 qualitatively agree with the isolated atom electronegativity and hardness sequences. The order of the effective electronegativities is as follows $\chi_{\mathrm{H}}^* < \chi_{\mathrm{S}}^* < \chi_{\mathrm{C}}^* < \chi_{\mathrm{N}}^* < \chi_{\mathrm{O}}^*$. This is in agreement with the Pauling electronegativities. In the Mulliken scale, the electronegativity of a hydrogen atom is higher than electronegativity of a carbon atom. The order of the hardness data (except for the M-N scheme) is: $\eta_{\mathrm{O}}^* > \eta_{\mathrm{N}}^* > \eta_{\mathrm{C}}^* > \eta_{\mathrm{S}}^*$. The value of η_{H}^* is close to that of η_{O}^*, and, depending on the hardness matrix, η_{H}^* is located either below η_{O}^* (Ohno and L-V) or above it (1/R). This is the isolated atom perspective. The chemical potential difference indicates the direction of charge transfer (CT). The electrons flow from an atom of lower electronegativity (higher chemical potential) to an atom of higher electronegativity (lower chemical potential). The hardness reflects the resistance of a given atom to charge transfer. The electronegativity of a closed atom in a molecule includes the electrostatic correction (see Eq. 3), which can change the isolated atom picture. For example, it is well known that carbon atoms bonded to electronegative atoms can form hydrogen bonds.

In order to compare the optimized parameters with the values reported in the literature, the effective hardness should be halved. This is connected with the Taylor expansion coefficient, which is not included in the hardness

Table 1 The optimized electronegativity (χ_i^*) and hardness (η_i^*) data for trivial-atom resolution

	1/R		M-N		Ohno		L-V	
	η_i^*	χ_i^*	η_i^*	χ_i^*	η_i^*	χ_i^*	η_i^*	χ_i^*
H	29.94	10.00	12.07	10.00	24.75	10.00	27.37	10.00
C	18.08	14.14	7.522	12.10	13.68	13.40	16.05	13.58
N	28.35	25.04	16.84	18.97	25.15	23.00	23.92	21.54
O	29.14	26.48	28.56	24.32	26.11	23.87	29.60	25.93
S	14.63	12.21	16.26	10.48	11.97	11.77	13.37	11.86

Table 2 The optimized electronegativity (χ_i^*) and hardness (η_i^*) data for hybridized-atom resolution

	1/*R*		M-N		Ohno		L-V	
	η_i^*	χ_i^*	η_i^*	χ_i^*	η_i^*	χ_i^*	η_i^*	χ_i^*
H	29.79	10.00	17.66	10.00	23.37	10.00	29.57	10.00
C sp^2	17.17	14.17	8.39	13.01	12.59	13.13	15.17	14.16
C sp^3	19.31	14.40	10.38	13.56	14.19	13.16	17.00	14.12
N sp^2	24.35	21.60	25.02	24.69	20.11	19.40	20.18	19.56
N sp^3	28.26	24.85	27.36	27.36	18.16	17.99	22.71	21.15
O sp^2	25.29	23.07	28.77	25.55	22.29	20.71	27.03	24.03
O sp^3	23.26	22.26	29.17	28.00	28.72	25.08	29.34	26.62
S	14.21	12.05	9.65	11.44	11.11	11.66	23.63	10.97

Table 3 The optimized electronegativity (χ_i^*) and hardness (η_i^*) data for AMBER force-field resolution

	1/*R*		M-N		Ohno		L-V	
	η_i^*	χ_i^*	η_i^*	χ_i^*	η_i^*	χ_i^*	η	χ_i^*
HC	30.09	10.00	18.52	10.00	23.34	10.00	29.40	10.00
H	32.38	9.09	15.98	9.95	28.38	8.31	31.15	9.69
HO	31.63	9.93	18.98	9.29	25.26	9.02	32.59	8.35
HS	32.41	9.07	23.44	9.16	19.59	10.20	26.83	8.73
H2	34.71	8.14	23.62	8.42	25.67	9.34	34.83	7.99
H3	32.66	8.31	20.50	7.47	19.05	11.06	33.14	8.61
C	18.25	13.21	10.99	11.16	13.82	12.30	15.62	13.81
CA	17.99	14.33	8.31	12.86	12.59	13.07	15.76	14.21
CB	16.58	13.89	9.76	13.08	12.27	13.18	14.34	14.09
CC	19.08	14.14	23.42	10.93	17.90	12.70	24.33	12.82
CK	2.92	18.96	25.19	9.23	33.11	8.59	28.62	10.55
CM	17.42	14.36	8.63	13.24	12.52	13.00	15.15	14.13
CN	31.92	8.17	16.94	10.35	16.46	11.94	19.97	12.40
CQ	30.72	11.33	25.21	9.96	28.64	10.52	27.32	11.90
CR	17.52	14.83	29.87	8.39	19.73	12.24	19.75	13.59
CT	19.19	14.40	10.10	13.56	14.41	13.16	17.03	14.12
CV	31.39	15.17	17.74	14.30	26.76	14.24	30.72	14.84
CW	30.98	14.78	21.54	13.46	34.77	13.83	34.92	14.45
C*	27.02	17.92	20.37	16.76	24.18	16.84	30.54	18.98
N	31.49	28.04	21.97	24.14	32.03	27.76	19.29	18.60
NA	23.37	21.52	33.99	31.66	26.12	23.68	24.36	23.68
NB	34.96	27.25	16.18	18.27	18.51	18.41	26.55	18.41
NC	34.70	28.26	34.94	28.04	28.07	22.74	25.95	22.74
NT	27.71	24.85	27.49	27.36	17.71	17.99	22.60	17.99
N2	32.38	29.27	26.12	26.51	27.02	23.94	33.21	23.94
N3	26.51	22.17	29.87	28.45	33.41	28.57	22.20	28.57
N*	27.30	26.21	25.84	27.13	28.15	25.54	23.70	25.54
O	27.54	23.52	34.81	26.78	29.01	22.98	36.57	22.98
OH	32.26	28.73	35.00	31.41	31.73	26.95	37.39	26.95
OS	22.27	22.26	29.90	28.00	28.80	25.08	29.47	25.08
O2	21.00	20.97	21.86	23.37	17.67	19.13	22.18	19.13
S	23.06	9.68	8.30	11.22	35.00	6.59	25.08	6.59
SH	17.73	12.21	9.72	11.44	11.65	11.61	16.80	11.61

data in our approach. In the case of electronegativity, the relative values should be compared. Our data are in qualitative agreement with those of Mortier at. al. [19, 20], Bultinck and coworkers [25], and Koča et al. [27]. Unfortunately, there was no quantitative agreement. This indicates that optimization is a very complex process and that many local minima exist on the energy hypersurface spanned in the variation parameter space. The optimization also depends on the training set and its size.

Table 2 collects the effective electronegativities and hardnesses for hybridized-atom resolution. The number of variation parameters is increased. Two types of hybridization (sp^2 and sp^3) of carbon, nitrogen, and oxygen atoms were considered. For Ohno and L-V hardness matrices, the atomic electronegativities are in disjoint domains, thus preserving the qualitative trends obtained for the trivial atoms. For $1/R$ and M-N hardness matrices, the domains of nitrogen overlap these of oxygen. The overlapping of the electronegativity and hardness domains is more obvious for AMBER force-field atoms, as shown in Table 3. This indicates that the effective electronegativity and hardness data are strongly dependent on the nearest environment.

To demonstrate the quality of the effective electronegativity and hardness data obtained, the fitness parameters (S^2), the linear correlation coefficients (R^2), and the best linear fits ($y=ax+b$) between the Mulliken and CSA charges are compared for all model hardness matrices and resolutions in Table 4. The first, second and third entries in Table 4 correspond to trivial-atom, hybridized-atom and force-field atom resolutions, respectively. It is clear from the table that at all resolutions the Ohno scheme is the best while M-N is the worst. The introduction of hybridization only slightly lowers the fitness parameter and improves the correlation. More pronounced is the influence of the nearest chemical neighborhood. The fitness parameter is halved after moving to AMBER force-field resolution. The data collected in Table 4 reflect the trends that should be expected after increasing the number of variation parameters. Namely, the fitness parameter decreases, the correlation coefficient increases and the slope of the fit line (a) increases and is close to the ideal correlation ($a=1$). In all cases, b is very small, a few orders of magnitude smaller than a. Therefore, we can safely assume that it is equal to zero. The correlation diagram between the Mulliken and CSA charges for the Ohno scheme is shown in Fig. 1. One can see that the correlation is indeed very good.

The data reported in Table 4 may give an impression that the efficiencies of computational schemes based on different resolutions and different hardness matrices are almost the same. To dispel this, histograms where different resolutions and schemes based on different hardness matrices are compared have been plotted in Fig. 2. It is evident from Fig. 2a that the deviation from the reference Mulliken charges is lessened when new atom types are introduced. With force-field resolution, the tail beyond 0.15 is reduced. The remaining two resolutions extend significantly beyond this value. In Fig. 2b, all computational schemes for force-field atoms are compared. It is now evident that Ohno's interpolation formula gives the smallest deviation from the reference values. The average absolute deviations (AAD) per atom type are highest for H (H attached to N), HC (H attached to C), and CT (aliphatic carbon), and are 0.02, 0.02

Table 4 Parameters characterizing simulations and the parameterizations obtained: fitness functions (S^2), correlation coefficients (R^2) and linear fits ($y=ax+b$) between CSA and Mulliken charges

	$\|S^2\|$	R^2	$y= ax+b$
	Trivial-atom resolution		
$1/R$	6.290	0.9765	$y=0.9771x+8\times10^{-5}$
M-N	9.726	0.9636	$y=0.9657x+0.0001$
Ohno	5.017	0.9812	$y=0.9819x+7\times10^{-5}$
L-V	5.271	0.9803	$y=0.9766x+9\times10^{-5}$
	Hybridized-atom resolution		
$1/R$	5.680	0.9788	$y=0.9752x+9\times10^{-5}$
M-N	9.313	0.9652	$y=0.9674x+0.0001$
Ohno	4.201	0.9843	$y=0.9826x+6\times10^{-5}$
L-V	4.430	0.9834	$y=0.9811x+7\times10^{-5}$
	AMBER force-field resolution		
$1/R$	3.977	0.9852	$y=0.9794x+8\times10^{-5}$
M-N	5.480	0.9795	$y=0.9861x+5\times10^{-5}$
Ohno	2.589	0.9903	$y=0.9859x+5\times10^{-5}$
L-V	3.267	0.9878	$y=0.9822x+7\times10^{-5}$

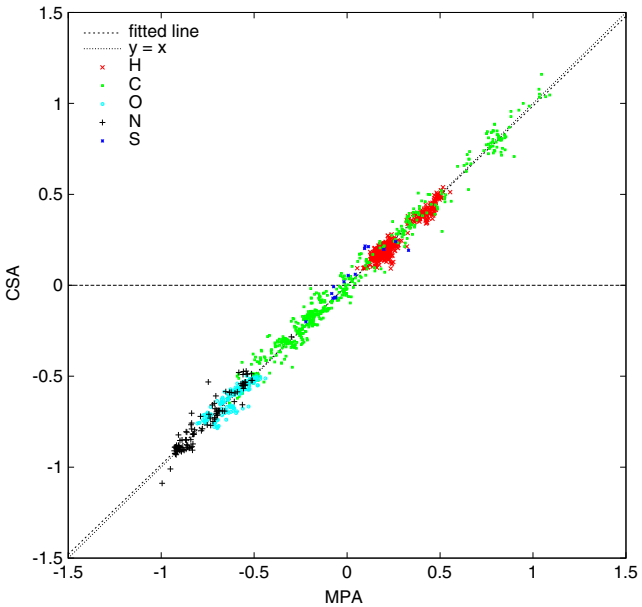

Fig. 1 Correlation diagram between Mulliken [HF/6-31G(d) level of theory] and CSA charges for the training set of molecules

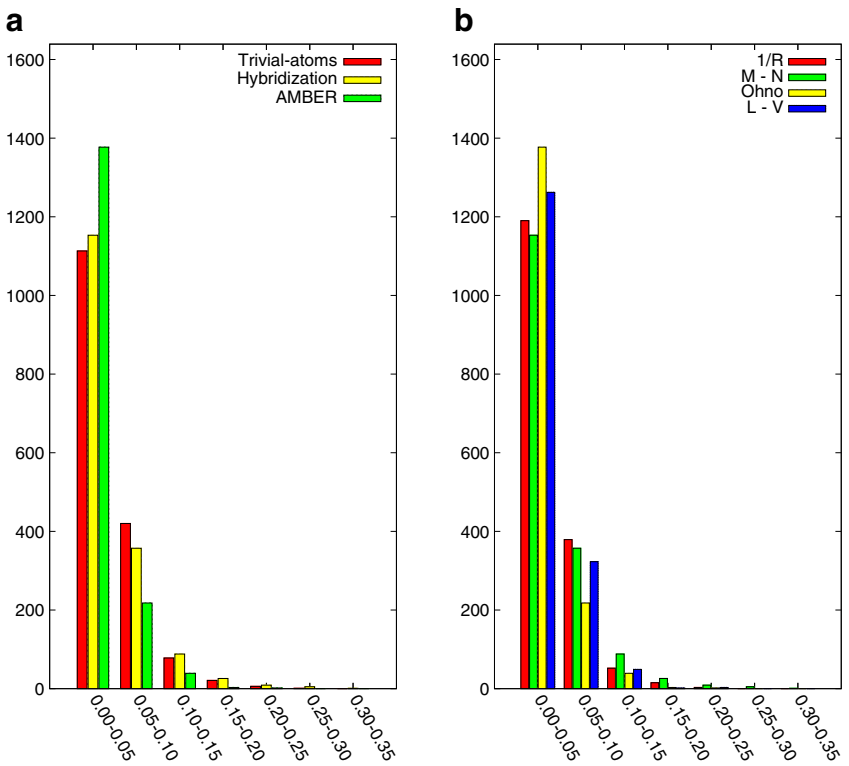

Fig. 2 **a**–**b** Histograms of the absolute differences between CSA and Mulliken charges [HF/6-31G(d) level of theory]: **a** comparison of trivial-atom, hybridized-atom and AMBER force-field resolutions; **b** charges derived from four different hardness matrices for AMBER force-field resolution are shown

and 0.01, respectively. For the remaining atoms, the average deviation is zero with the assumed accuracy. The corresponding root mean square deviations (RMSD) per atom type are 0.03, 0.03 and 0.02, respectively. The full list of AAD/atom and RMSD/atom data is included in Table 1 of the ESM. The AAD and RMSD per atom in each molecule are reported in Table 2 of the ESM. The highest deviations does not exceed 0.05 (AAD/atom) and 0.07 (RMSD/atom). A detailed examination of the training set demonstrates that H, HC and CT atoms are responsible for the deviations from the reference Mulliken charges.

The data presented in Table 4 and Figs. 1 and 2 correspond to the training set of molecules. To validate the obtained parameters, we applied CSA to a validation set. None of the molecules from the validation set were in the training set. The validation set included completely new classes of molecules: mono- and disaccharides, lactams, keto acids, thio acids, thioesters, carbamic acid and its derivatives, and others. The structures of molecules from the validation set are shown in the ESM. The correlation obtained and the deviations from the reference charge distributions are shown in Fig. 3. All calculations were performed with AMBER resolution and Ohno's interpolation formula (hardness matrix). The correlation coefficient was found to be 0.9762. The distribution obtained is very close to the reference values. Thus, the results obtained for the training and validation sets are close to each other.

The AAD and RMSD per force-field atom and per atom in each molecule of the validation set are summarized in Tables 3 and 4 of the ESM. Except for one molecule from the *p*-tertbutylcalix [4] family, the AAD and RMSD are in the same ranges as those reported for the training set. Actually, this molecule has the highest number of HC and CT atoms. The reason for the poor reproduction of Mulliken charges for these atom types is their low specificity. Namely, the CT type corresponds to any sp^3 carbon, and HC to any hydrogen attached to carbon. Further partitioning of HC and CT atom types, as done in CHARMM [42] and recent AMBER [43] parameterizations, should improve the agreement between CSA and Mulliken charges. Confirmation of this observation can be found in Table 3 of the ESM, where AAD and RMSD values per force-field-atom type are listed for the Ohno hardness matrix. Just as observed for the training set, high deviations occur for HC and CT atoms. In addition, a slight increase in the error is observed for OH and HO force-field atoms (hydroxyl group). It seems that distinguishing different aliphatic carbon atoms improves the descriptions of OH and HO atoms.

In Table 5 we have summarized the results of the simulation performed to derive CHELPG charges. The data obtained indicate that the reproduction of charges fitted to electrostatic potential is not satisfactory. The main reason for this is that the level of theory used to obtain CHELPG charges was insufficient. Such charges are very sensitive to the *N*-electron wavefunction and possess huge

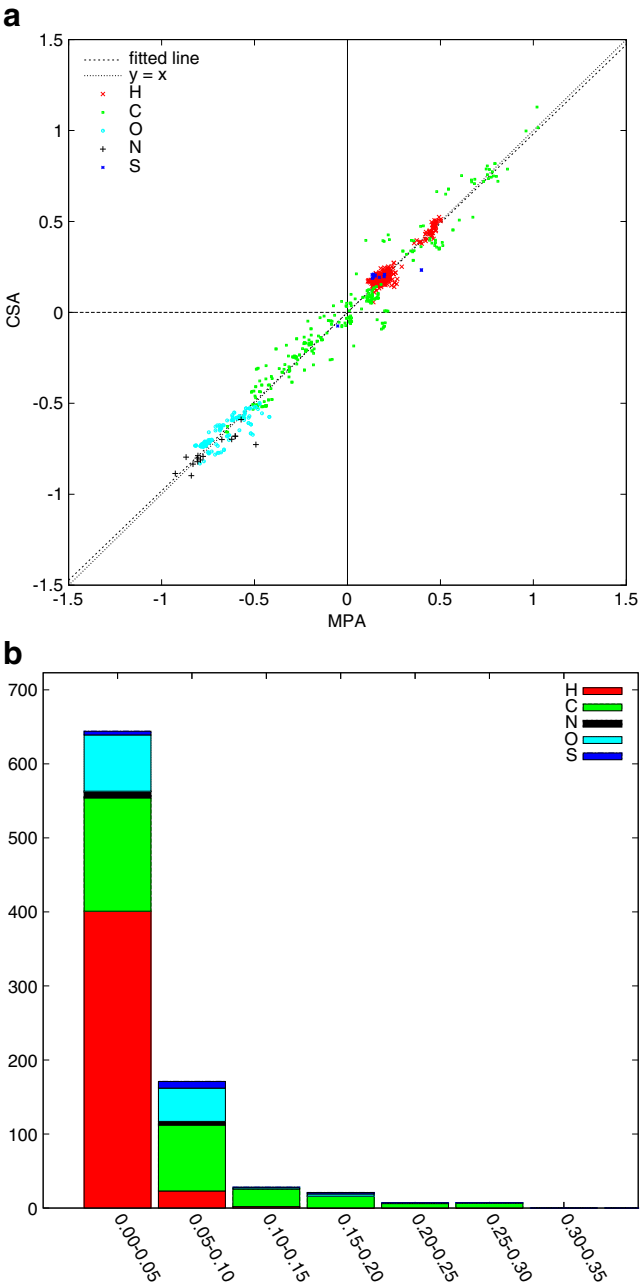

Fig. 3 **a**–**b** Correlation between Mulliken [HF/6-31G(d) level of theory] and CSA charges (**a**), and histograms of the absolute differences between the CSA and Mulliken charges (**b**) for the validation set of molecules

statistical inaccuracies. Extension beyond the HF scheme is probably required to obtain "transferable" CHELPG charges.

Conclusions and future prospects

This article has presented the extension of Charge Sensitivity Analysis to AMBER force-field resolution. The effective electronegativity and hardness data were found using evolutionary algorithms. The parameters for trivial-atom and hybridized-atom resolutions are also reported. The Mulliken population analysis and charges derived from fitting to the electrostatic potential (CHELPG) were considered. Four different hardness matrices were taken into account. The current parameterization included hydrogen, carbon, nitrogen, oxygen, and sulfur atoms. For hybridized-atom resolution, sp^2 and sp^3 states of carbon, nitrogen, and oxygen atoms were considered. AMBER force-field resolution distinguished 33 different chemical environments of H, C, N, O, and S atoms. It was demonstrated that the effective electronegativities and hardnesses depend strongly on the nearest neighborhood. The reproduction of Mulliken charges by CSA was very good, with only very small average deviations from the reference distribution. The best agreement was observed for Ohno's interpolation formula. CSA based on Ohno's formula gave the narrowest and sharpest distribution around the reference Mulliken charges. The same observations were valid for the validation set. Taking into account that the validation set contained new classes of molecules, none of which were included in the training set, the obtained electronegativities and hardnesses would appear to be reasonable.

CHELPG charges are not as transferable as Mulliken charges. They show huge statistical inaccuracies that reflect the quality of the *N*-electron wavefunction. A post-Hartree–Fock treatment and a large basis set are probaby required to obtain "transferable" charges.

Next, we plan to connect force-field CSA with molecular dynamics calculations. The first step towards this involves

Table 5 Parameters characterizing simulation and the parameterization obtained: fitness function (S^2), correlation coefficient (R^2) and linear fit ($y=ax+b$) between the CSA and CHELPG charges

	$\lvert S^2\rvert$	R^2	$y=ax+b$
	Trivial-atom resolution		
1/*R*	51.989	0.8028	$y=0.8055x+0.0007$
M-N	63.619	0.7587	$y=0.7601x+0.0009$
Ohno	55.170	0.7908	$y=0.7962x+0.0007$
L-V	52.937	0.7992	$y=0.8012x+0.0007$
	Hybridized-atom resolution		
1/*R*	46.834	0.8224	$y=0.8168x+0.0007$
M-N	62.566	0.7627	$y=0.763x+0.0009$
Ohno	51.912	0.8031	$y=0.8085x+0.0007$
L-V	47.359	0.8204	$y=0.8178x+0.0007$
	AMBER force-field resolution		
1/*R*	38.472	0.8541	$y=0.8578x+0.0005$
M-N	51.248	0.7896	$y=0.8004x+0.0007$
Ohno	42.368	0.8393	$y=0.8424x+0.0006$
L-V	37.829	0.8565	$y=0.8598x+0.0005$

checking how the force-field CSA describes mutually polarized subsystems and the polarization induced by an external electric field. Other possible applications are QSAR and QSPR models. The extension of this work to other elements and population analyses is also planned.

Acknowledgments This work was partly supported by the Ministry of Sciences and Higher Education (project no. 1206/ GDR/2007/03). Calculations were carried out with equipment purchased thanks to the financial support of the European Regional Development Fund in the framework of the Polish Innovation Economy Operational Program (contract no. POIG.02.01.00-12-023/08).

References

1. Mulliken RS (1955) Electronic polulation analisys on LCAOMO molecular wave functions I. J Chem Phys 23:1833–1841
2. Löwdin PO (1950) On the non-orthogonality problem connected with the use of atomic wave functions in the theory of molecules and crystals. J Chem Phys 18:365–376
3. Roby KR (1974) Definition of the charge on an atom in a molecule and of occupation numbers for electron density shared between atoms. Mol Phys 27:81–104
4. Bader RFW (1990) Atoms in molecules. Oxford University Press, Oxford
5. Hirshfeld FL (1977) Bonded-atom fragments for describing molecular charge densities. Theor Chem Acc 44:129–138
6. Reed AB, Weinstock RB, Weinhold F (1985) Natural population analysis. J Chem Phys 83:735–746
7. Besler BH, Merz KM Jr, Kollman PA (1990) Atomic charges derived from semiempirical methods. J Comput Chem 11:431–439
8. Chirlian LE, Francl MM (1987) Atomic charges derived from electrostatic potentials: a detiled study. J Comput Chem 8:894–905
9. Breneman CM, Wiberg KB (1990) Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in foramide conformational analisys. J Comput Chem 11:361–373
10. Li J, Zhu T, Cramer CJ, Truhlar DG (1998) New class IV charge model for extracting accurate partial charges from wave functions. J Phys Chem A 102:1820–1831
11. Cerofolini GF, Meda L, Re N (2000) Core electrons as probes of the net atomic charge. A theoretical and experimental investigations. Res Adv Phys Chem 1:77
12. Townes CH, Dailey BP (1949) Determination of electronic structure of molecules from nuclear quadrupole effects. J Chem Phys 17:782–796
13. Burnes WA, Phillips JA, Canagaratna M, Goodfriend H, Leopold KR (1999) Partially formed bonds In HCN−SO_3 and CH_3CN−SO_3: a comparison between donor−acceptor complexes of SO_3 and BF_3. J Phys Chem A 103:7445–7453
14. Szewczyk B, Sokalski WA, Leszczynski J (2002) Optimal methods for calculation of the amount of intermolecular electron transfer. J Chem Phys 117:6952–6958
15. Sanderson RT (1951) An interpretation of bond lengths and a classification of bonds. Science 114:670–672
16. Pauling L (1939) The nature of the chemical bond, 3rd edn. Ithaca, Cornell
17. Parr RG, Yang W (1989) Density functional theory of atoms and molecules. Oxford University Press, Oxford
18. Mulliken RS (1934) A new electroaffinity scale; together with data on valence states and on valence ionization potentials and electron affinities. Chem Phys 2:782–793
19. Mortier WJ, van Genechten K, Gasteiger J (1985) Electronegativity equalization: application and parametrization. J Am Chem Soc 107:829–835
20. Mortier WJ, Gosh SK, Shankar S (1986) Electronegativity equalization methodfor the calculations of atomic charges in molecules. J Am Chem Soc 108:4315–4320
21. Rappé AK, Goddard W (1991) Charge equilibration for molecular dynamics simulations. J Phys Chem 95:3358–3363
22. Rappé AK, Casewit C, Colwell K, Goddard W, Skiff W (1992) UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. J Am Chem Soc 114:10024–10035
23. Chelli R, Procassi P (2002) A transferable polarizable electrostatic force field for molecular mechanics based on the chemical potential equalization principle. J Chem Phys 117:9175–9189
24. Patel S, Brooks CL III (2004) CHARMM fluctuating charge force field for proteins. I. Parameterization and application to bulk organic liquid simulations. J Comput Chem 25:1–15
25. Bultinc P, Vanholme R, Popelier PLA, de Proft F, Geerlings P (2004) High-speed calculation of AIM charges through the electronegativity equalization method. J Chem Phys A 108:10359–10366
26. Chaves J, Barroso JM, Bultinck P, Carbo-Dorca R (2006) Toward an alternative hardness kernel matrix structure in the electronegativity equalization method (EEM). J Chem Inf Model 46:1657–1665
27. Svobodová Vařekova R, Jiroušková Z, Vanek J, Suchomel Š, Koča J (2007) Electronegativity equalization method: parameterization and validation for large sets of organic, organohalogene and organometal molecule. Int J Mol Sci 8:572–582
28. Jiroušková Z, Svobodová Vařekova R, Vanek J, Koča J (2009) Electronegativity equalization method: parameterization and validation for organic molecules using the Merz–Kollman–Singh charge distribution scheme. J Comput Chem 30: 11741178
29. Ouyang Y, Ye F, Liang Y (2009) A modified electronegativity equalization method for fast and accurate calculation of atomic charges in large biological molecules. Phys Chem Chem Phys 11:6082–6089
30. Korchowiec J, Kowalski P (1993) Charge polarization of molecular systems: charge sensitivity and MNDO study. Chem Phys Lett 208:135–138
31. Nalewajski RF, Korchowiec J (1998) Charge sensitivity approach to electronic structure and chemical reactivity. World Scientific Publishing, Singapore
32. Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S Jr, Weiner P (1984) A new force field for molecular mechanical simulation of nucleic acids and proteins. J Am Chem Soc 106:765–784
33. Weiner SJ, Kollman PA, Nguyen DT, Case DA (1986) An all atom forcefield for simulations of proteins and nucleic acid. J Comput Chem 7:230–252
34. Korchowiec J, Gerwens H, Jug K (1994) Relaxed fukui function indices and their application to chemical reactivity problems. Chem Phys Lett 222:58–64
35. Parr RG, Pearson RG (1983) Absolute hardness: companion parameter to absolute electronegativity. J Am Chem Soc 105:7512–7516
36. Parr RG, Yang W (1984) Density functional approach to the frontier-electron theory of chemical reactivity. J Am Chem Soc 106:4049–4050
37. Mataga M, Nishimoto K (1957) Electronic structure and spectra of nitrogen heterocycles. Z Phys Chem 13:140–152
38. Ohno K (1964) Some remarks on the Pariser–Parr–Pople method. Theor Chim Acta (Berlin) 2:219–227

39. Louwen JN, Vogh ETC (1998) Semi-empirical atomic charges for use in computational chemistry of molecular sieves. J Mol Catal A 134:63–77
40. Adcock S (2001–2009) GAUL open source programming library. http://gaul.sourceforge.net/
41. Schmidt MW, Baldridge KK, Boatz JA, Elbert ST, Gordon MS, Jensen JH, Koseki S, Matsunaga N, Nguyen KA, Su S, Windus TL, Dupuis M, Montgomery JA (1993) General atomic and molecular electronic structure system. J Comput Chem 14:1347–1363
42. MacKerell AD Jr, Feig M, Brooks CL III (2004) Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J Comput Chem 25:1400–1415
43. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1994) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc 117:5179–5197