

Struct Chem (2012) 23:1449–1458  
DOI 10.1007/s11224-012-0054-7

ORIGINAL RESEARCH

# Generalized charge sensitivity analysis

Anna Stachowicz · Jacek Korchowiec

Received: 1 July 2011 / Accepted: 20 December 2011 / Published online: 19 June 2012  
© The Author(s) 2012. This article is published with open access at Springerlink.com

**Abstract** Charge sensitivity analysis was originally introduced in the trivial-atom resolution. Here, we extend this resolution into force-field atoms. The AMBERff99 force-field resolution was employed. The effective electronegativities and hardnesses were derived for five different population analyses (Mulliken, Hirschfeld, AIM, NPA and Voronoi charges) by applying evolutionary algorithms.

**Keywords** Charge sensitivity analysis · Electronegativity equalization principle · Electronegativity · Hardness · AMBER force field

## Introduction

Charge sensitivity analysis (CSA) [1], the formalism employed in this study, originates from density functional theory (DFT) and is a part of conceptual DFT [2–4]. It may be considered as a generalization of electronegativity equalization (EE) method [5]. The EE-based methods [6–12] were mainly applied for deriving charge distribution inside non-interacting molecules. Such static model was introduced to molecular dynamics by Rappe and Goddard [8] as initial guess generator (charge distribution generator). The use of extended Lagrangian method extended the area of application of EE method to interacting molecules

[13, 14]. However, the model introduced unphysical thermal fluctuations and additional thermostat was required to cool down the “polarization” degree of freedom. In contrast to EE methods, CSA can describe polarization directly without resorting to Lagrangian method and in addition it allows to define a wide range of sensitivities for each chemically interesting partitioning of the system, therefore, one can monitor the progress of a given chemical process. It was demonstrated in static applications that CSA has a huge potential in the reactivity theory [15–17]. Therefore, the dynamic aspect of the CSA may provide many additional information.

The charge sensitivity,  $P_{AB}^R = (\partial p_A / \partial t_B)_{\mathcal{R}}$ , is a differential quantity and represents the response of the parameter  $p$  characterizing equilibrium state of fragment A to a displacement of parameter  $t$  of the equilibrium state of the another fragment B (see Chapter 1 of ref [1]). This response is measured under specific constraints imposed on the molecular remainder  $\mathcal{R} = (C, D, E, \dots, X, Y, Z)$ . The remainder can be further divided into the freely relaxing and frozen parts  $\mathcal{R} = (C|D|E| \dots |X|Y|Z)$ . Here, the broken lines are placed between relaxing fragments while the solid lines are placed between frozen fragments. Based on this convention, one can define rigid [ $\mathcal{R} = (C|D|E| \dots |X|Y|Z)$ ], relaxed [ $\mathcal{R} = (C|D|E| \dots |X|Y|Z)$ ] and any intermediate sensitivities. In such a way, the whole hierarchy of  $P_{AB}^R$  can be computed.

Routinely, CSA was applied as supplementary tool to semiempirical or ab initio calculations. Our main intention is to extend its area of applications. We plan to couple CSA with molecular dynamics (MD) simulations. Prior to this, the method should be parameterized for a given force field. This will be also a step towards polarizable force fields since standard force fields used in molecular modeling describe electrostatic interactions in terms of fixed, atom

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

centered, charges. In addition, CSA will introduce “dynamic” qualitative structure activity relationships (QSAR) or qualitative structure property relationships (QSPR) models into MD simulations.

The article is organized as follows. First, CSA methodology is given. Next, the optimization procedure is described. Afterwards, the results obtained are discussed. Finally, conclusions and future prospects are briefly discussed.

#### A short survey of charge sensitivity analysis

The CSA in atomic resolution is based on second-order Taylor expansion of the system energy ( $E_M$ ) with respect to atomic charges [ $\mathbf{q} = (q_1, q_2, \dots, q_N)$ ]:

$$\begin{aligned} dE_M &= \sum_{i=1}^N \left( \frac{\partial E_M}{\partial q_i} \right) dq_i + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \left( \frac{\partial^2 E_M}{\partial q_i \partial q_j} \right) dq_i dq_j \\ &= \sum_{\Omega} \sum_{i=1}^{n_{\Omega}} \left( \frac{\partial E_M}{\partial q_i} \right) dq_i \\ &\quad + \frac{1}{2} \sum_{\Omega} \sum_{i=1}^{n_{\Omega}} \sum_{\Xi} \sum_{j=1}^{n_{\Xi}} \left( \frac{\partial^2 E_M}{\partial q_i \partial q_j} \right) dq_i dq_j, \\ &\equiv \sum_{\Omega} \sum_{i=1}^{n_{\Omega}} \chi_i dq_i + \frac{1}{2} \sum_{\Omega} \sum_{i=1}^{n_{\Omega}} \sum_{\Xi} \sum_{j=1}^{n_{\Xi}} dq_i \eta_{ij} dq_j, \\ &\quad (\Omega, \Xi) \in \{A, B, \dots, Z\}. \end{aligned} \quad (1)$$

The sum over fragments  $\Omega$  and  $\Xi$  is introduced for further derivations. The number of atoms in the fragments  $\Omega$  and  $\Xi$  are denoted by  $n_{\Omega}$  and  $n_{\Xi}$ , respectively. The overall number of atoms in the system is equal to  $N$ . All differentiations are carried out for a fixed external potential due to nuclei  $\mathbf{v}$  and frozen atomic charges except those distinguished in derivatives. The atomic electronegativities  $\chi = (\chi_1, \chi_2, \dots, \chi_N) = (\chi_A, \chi_B, \dots, \chi_Z)$  and the elements of hardness matrix  $\eta = \{\eta_{ij}\} = \{\eta_{\Omega\Xi}\} [(i, j) \in \{1, 2, \dots, N\}; (\Omega, \Xi) \in \{A, B, \dots, Z\}]$  are the main CSA parameters which should be adjusted to reproduce atomic charges on the reference set of molecules. Both  $\chi$  and  $\eta$  are totally rigid quantities since by definition molecular system  $M$  is divided into  $N$  mutually closed atoms.

In fragment resolution, the following set of equations:

$$\begin{aligned} \chi_1^A &= \chi_2^A = \dots = \chi_{n_A}^A = \chi_A \equiv (\partial E_M / \partial q_A) \\ \chi_1^B &= \chi_2^B = \dots = \chi_{n_B}^B = \chi_B \equiv (\partial E_M / \partial q_B) \\ &\vdots \\ \chi_1^Z &= \chi_2^Z = \dots = \chi_{n_Z}^Z = \chi_Z \equiv (\partial E_M / \partial q_Z) \end{aligned} \quad (2)$$

marks the intra-fragment equilibrium. In general, fragment electronegativities are different:  $\chi_A \neq \chi_B \neq \dots \neq \chi_Z$ . This equation should be completed by charge conservation equations:

$$\sum_{i \in \Omega} dq_i = q_{\Omega}, \quad \Omega \in \{A, B, \dots, Z\}, \quad (3)$$

The whole formalism can be summarized in a single matrix equation [17]:

$$\begin{pmatrix} 0 & 0 & 0 & \mathbf{1}_A & \mathbf{0}_B & \dots & \mathbf{0}_Z \\ \vdots & \ddots & \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & \dots & 0 & \mathbf{0}_A & \mathbf{0}_B & \dots & \mathbf{1}_Z \\ \mathbf{1}_A^\dagger & \dots & \mathbf{0}_A^\dagger & \eta_{AA} & \eta_{AB} & \dots & \eta_{AZ} \\ \mathbf{0}_B^\dagger & \dots & \mathbf{0}_B^\dagger & \eta_{BA} & \eta_{BB} & \dots & \eta_{BZ} \\ \vdots & & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_Z^\dagger & \dots & \mathbf{1}_Z^\dagger & \eta_{ZA} & \eta_{ZB} & \dots & \eta_{ZZ} \end{pmatrix} \begin{pmatrix} -\chi_A \\ \vdots \\ -\chi_Z \\ \mathbf{q}_A \\ \mathbf{q}_B \\ \vdots \\ \mathbf{q}_Z \end{pmatrix} = \begin{pmatrix} q_A \\ \vdots \\ q_Z \\ -\chi_A \\ -\chi_B \\ \vdots \\ -\chi_Z \end{pmatrix}, \quad (4)$$

where vectors  $\mathbf{1}_{\Omega}$  and  $\mathbf{0}_{\Omega}$  are filled by 0 and 1, respectively. In order to simplify this equation, the neutral atom limit,  $dq_{\Omega} = q_{\Omega}$  have been used. By inverting Eq 4, the charge distributions inside fragments are obtained:

$$\begin{pmatrix} -\chi_A \\ \vdots \\ -\chi_Z \\ \mathbf{q}_A \\ \mathbf{q}_B \\ \vdots \\ \mathbf{q}_Z \end{pmatrix} = \begin{pmatrix} -\eta_{AA} & \dots & -\eta_{AZ} & \mathbf{f}_{A,A} & \mathbf{f}_{A,B} & \dots & \mathbf{f}_{A,Z} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \dots & \vdots \\ -\eta_{ZA} & \dots & -\eta_{ZZ} & \mathbf{f}_{Z,A} & \mathbf{f}_{Z,B} & \dots & \mathbf{f}_{Z,Z} \\ \mathbf{f}_{A,A}^\dagger & \dots & \mathbf{f}_{A,Z}^\dagger & -\beta_{AA} & -\beta_{AB} & \dots & -\beta_{AZ} \\ \mathbf{f}_{B,A}^\dagger & \dots & \mathbf{f}_{B,Z}^\dagger & -\beta_{BA} & -\beta_{BB} & \dots & -\beta_{BZ} \\ \vdots & & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{f}_{Z,A}^\dagger & \dots & \mathbf{f}_{Z,Z}^\dagger & -\beta_{ZA} & -\beta_{ZB} & \dots & -\beta_{ZZ} \end{pmatrix} \begin{pmatrix} q_A \\ \vdots \\ q_Z \\ -\chi_A \\ -\chi_B \\ \vdots \\ -\chi_Z \end{pmatrix}. \quad (5)$$

The elements of the inverse matrix are [17]: diagonal ( $\mathbf{f}_{A,A} = \{\partial q_i^A / \partial q_A\}$ ) and off-diagonal ( $\mathbf{f}_{A,B} = \{\partial q_i^A / \partial q_B\}$ ) Fukui function (FF) vectors, diagonal ( $\boldsymbol{\beta}_{A,A} = \{\partial q_i^A / \partial v_j^A\}$ ) and off-diagonal ( $\boldsymbol{\beta}_{A,B} = \{\partial q_i^A / \partial v_j^B\}$ ) polarization matrices, and hardness matrix ( $\boldsymbol{\eta}^{\text{frg}} = \{\eta_{AB} = \partial \chi_A / \partial q_B\}$ ) in fragment (frg) resolution. Diagonal FF vectors are normalized to unity ( $\mathbf{f}_{A,A} \mathbf{1}_A^\dagger = 1$ ) while off-diagonal are normalized to zero ( $\mathbf{f}_{A,B} \mathbf{1}_B^\dagger = 0$ ). Both diagonal and off-diagonal polarization matrices are normalized to zero ( $\sum_{i \in \Omega} \boldsymbol{\beta}_{\Omega \Xi} = 0$ ,  $(\Omega, \Xi) \in \{A, B, \dots, Z\}$ ), since perturbation in the external potential at the position of atom  $j$  belonging to fragment  $B$  ( $v_j^B$ ) does not change the overall charge. When one is interested in global equilibrium ( $\chi_A = \chi_B = \dots = \chi_Z = \chi$ ), Eqs 4 and 5 have simpler form:

$$\begin{pmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & \eta_{11} & \eta_{12} & \cdots & \eta_{1N} \\ 1 & \eta_{21} & \eta_{22} & \cdots & \eta_{2N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \eta_{N1} & \eta_{N2} & \cdots & \eta_{NN} \end{pmatrix} \begin{pmatrix} -\chi \\ q_1 \\ q_2 \\ \vdots \\ q_N \end{pmatrix} = \begin{pmatrix} q \\ -\chi_1 \\ -\chi_2 \\ \vdots \\ -\chi_N \end{pmatrix}, \quad (6)$$

$$\begin{pmatrix} -\chi \\ q_1 \\ q_2 \\ \vdots \\ q_N \end{pmatrix} = \begin{pmatrix} -\eta & f_1 & f_2 & \cdots & f_N \\ f_1 & -\beta_{11} & -\beta_{12} & \cdots & -\beta_{1N} \\ f_2 & -\beta_{21} & -\beta_{22} & \cdots & -\beta_{2N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_N & -\beta_{N1} & -\beta_{N2} & \cdots & -\beta_{NN} \end{pmatrix} \begin{pmatrix} q \\ -\chi_1 \\ -\chi_2 \\ \vdots \\ -\chi_N \end{pmatrix}, \quad (7)$$

where  $\chi = \partial E_M / \partial q$  is the global electronegativity [18],  $\eta = \partial^2 E_M / \partial q^2 = \partial \chi / \partial q$  is the global hardness [19],  $\boldsymbol{\beta} = \{\beta_{ij} = \partial q_i / \partial v_j, i, j = 1, 2, \dots, N\}$  is the polarization or linear response matrix and  $\mathbf{f} = \{f_i = \partial q_i / \partial q, i = 1, 2, \dots, N\}$  is the FF [20] vector. The FF vector is normalized to 1 ( $\mathbf{f} \mathbf{1}^\dagger = 1$ ). All elements in rows or columns of  $\boldsymbol{\beta}$  sum up to zero ( $\sum_i \beta_{ij} = \sum_i \beta_{ji} = 0$ ). Equation 6 and 7 describe transformation from constrained equilibrium to global equilibrium. In the state of constrained equilibrium, all atoms in the system  $M$  are closed to each other:  $M = (1|2|\dots|N)$ . Namely, the charge-transfer (CT) among atoms is not allowed. By removing barriers on CT the global equilibrium is restored:  $M = (1|2|\dots|N)$ . Intermediate, hypothetical equilibria are described by Eq. 4 and 5.

Empirical parameters of CSA are vector  $\boldsymbol{\chi}$  and hardnesses matrix  $\boldsymbol{\eta}$ . The off-diagonal hardnesses, which have the meaning of two electron integrals, can be related to diagonal hardnesses by empirical interpolation formulas. We have demonstrated previously [21] that Ohno interpolation formula [22]:

$$\eta_{ij} = \frac{1}{\sqrt{a_{ij}^2 + R_{ij}^2}}, \quad (8)$$

where  $a_{ij} = 2 / (\eta_{ii} + \eta_{jj})$  and  $R_{ij}$  is interatomic separation, gave the sharpest distribution around the reference ab initio charges and for this reason is employed in the article. For isolated atoms,  $\chi_i$  and  $\eta_{ii}$  depend solely on the atomic number [2, 23]. In molecules, however, they are also influenced by atom's hybridization and nearest chemical environment [24, 25]. These quantities are named effective electronegativity ( $\chi_i$ ) and effective hardness ( $\eta_i \equiv \eta_{ii}$ ) throughout the article. They have nothing in common with global electronegativity ( $\chi$ ) and global hardness ( $\eta$ ). Matching rules should be applied to compute  $\chi(\eta)$  from  $\chi_i(\eta_i)$ . The resultant atomic hardnesses:  $\bar{\eta}_i = (\partial \mu / \partial N_i) = \sum_j \eta_{ij} f_j = \eta$ , which fulfills hardness equalization principle, are different from the effective atomic hardnesses. More about equalization principles in fragment and global resolutions can be found in Ref. [26].

The effective quantities are determined by reproducing gas phase charge distribution for a set of training molecules calculated with an ab initio method. In our previous article [21], we have determined CSA parameters based on original Wiener's AMBERff84 [27] atomic types. Two population analyses were used independently to calculate the reference ab initio charge distribution of training molecules, namely Mulliken Population Analysis (MPA) [28] and electrostatic potential fitting (ChelpG scheme) [29, 30] calculated at HF/6-31G\* level of theory. It was shown that MPA charges are much better reproduced by CSA than ChelpG charges, what indicated their better transferability from system to system. In this study, we extend force-field resolved CSA to other population analyses. Hirshfeld [31], Voronoi [32, 33], natural population analysis (NPA) [34] and Bader's atoms-in-molecules (AIM) [35] charges have been employed. We have also introduced the additional atomic types consistent with AMBERff99 [36, 37] parameterization. The goal of this study is to explore to what extent these charges can be reproduced by CSA.

## Methods

All ab initio charges were calculated at B3LYP/6-31G\* level of theory. Gaussian [38] program was used for calculating Hirshfeld and NPA charges. Voronoi and AIM charges were calculated with the use of Amsterdam density functional [39] package and DZP basis set. We have also performed parameterization with Mulliken charges computed at B3LYP/6-31G\* level of theory for comparison.

The training set consisted of a hundred small and medium sized organic molecules. The total number of

atoms in this set was equal 1617. Most of the molecules were of biological importance in order to cover the area of application of the AMBER force field. In particular, standard amino acids and DNA/RNA bases were included. The structure of all molecules was optimized at HF/6-31G\* level of theory. The information concerning these molecules can be found in Ref. [21].

Our training set contains 38 AMBERff99 atom types. It means that the number of CSA parameters is equal to 76. For this reason systematic search of any kind could not have been employed. Instead we have used evolutionary algorithms (EA). Nonetheless, simultaneous optimization of so many parameters is still difficult. That is the reason why a sequential procedure has been used. First, the parameterization for different elements: H, C, N, O and S was performed (trivial-atom resolution). Then  $sp^2$  and  $sp^3$  hybridization for carbon, nitrogen and oxygen was introduced (hybridized-atom resolution). Parameters obtained at trivial-atom resolution were perturbed no more than  $\pm 30\%$  of their starting values. Finally AMBERff99 atom types were introduced. Optimization was conducted as follows: each element was parameterized in turn; the values of parameters from hybridized-atom resolution were perturbed. Hardnesses were perturbed in the range of  $\pm 30\%$  of their initial (hybridized-atom) values. Since electronegativities exhibited much lower changeability in the previous runs, the range for those parameters was narrowed to  $\pm 15\%$  of their initial values. Fitness function was defined as:

$$S^2 = - \sum_A \sum_{\alpha \in A} \left( q_{A,\alpha}^{\text{B3LYP}} - q_{A,\alpha}^{\text{CSA}} \right)^2, \quad (9)$$

where  $q_A^{\text{B3LYP}}$  and  $q_A^{\text{CSA}}$  denote vectors collecting A-th molecule's atomic charges calculated with ab initio methods and CSA, respectively. The first sum in Eq 9 goes over all the training molecules.  $S^2$  tends to zero in the ideal case of  $q_A^{\text{CSA}}$  being identical to  $q_A^{\text{B3LYP}}$ . Negative sign of the function is introduced in order to make it increasing. The details of genetic calculations were the same as in our previous study. GAUL [40] library was used for performing EA calculations. It was coupled with the CSA package developed in our group.

## Results

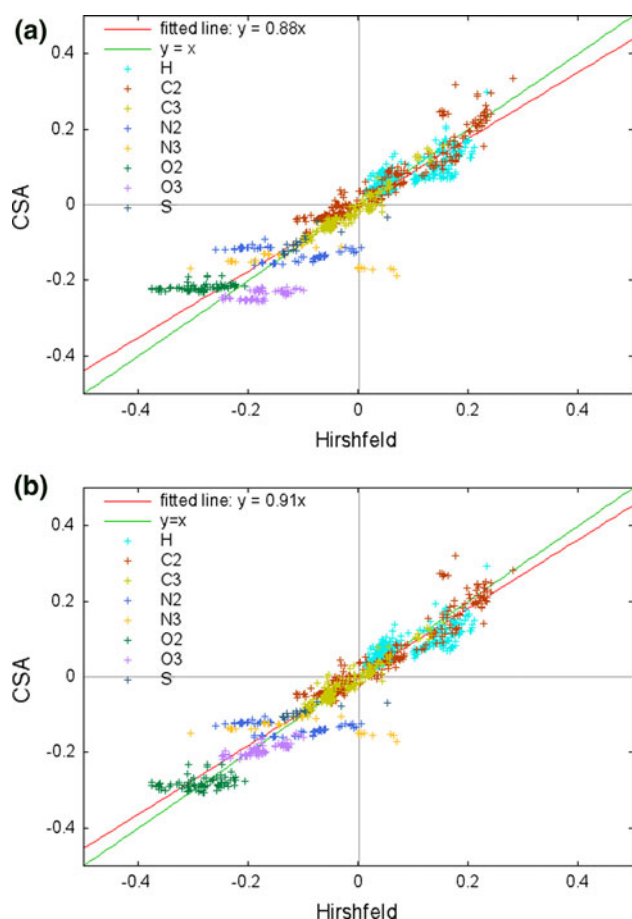
In Table 1, the values of the fitness function, correlation coefficients ( $R^2$ ) and best linear fits ( $y = ax + b$ ) between CSA and respective ab initio charges are collected. Examples of correlation plots between CSA and ab initio charges are presented in Figs. 1, 2, 3. The obtained values of y-intercepts in linear fits were negligible so  $y = ax$  type fits were adopted. It can be seen that out of the five population analyses employed three, namely AIM, MPA and NPA, are

**Table 1** Parameters characterizing simulation process and quality of obtained parameterization, i.e., fitness function ( $S^2$ ), correlation coefficient ( $R^2$ ) and linear fit ( $y = ax$ ) for MPA, AIM, NPA, Hirshfeld and Voronoi population analyses

Reference population analysis	$ S^2 $	$R^2$	$y = ax$
Trivial-atom resolution			
MPA	3.030	0.9826	$y = 0.98x$
AIM	6.391	0.9870	$y = 0.98x$
NPA	5.142	0.9812	$y = 0.98x$
Hirshfeld	2.528	0.8787	$y = 0.88x$
Voronoi	2.394	0.9070	$y = 0.91x$
Hybridized-atom resolution			
MPA	3.025	0.9826	$y = 0.98x$
AIM	3.460	0.9930	$y = 0.99x$
NPA	4.997	0.9817	$y = 0.98x$
Hirshfeld	1.929	0.9075	$y = 0.91x$
Voronoi	1.691	0.9344	$y = 0.94x$
AMBER force-field resolution			
MPA	1.982	0.9886	$y = 0.96x$
AIM	2.675	0.9945	$y = 0.99x$
NPA	2.905	0.9894	$y = 0.99x$
Hirshfeld	0.879	0.9578	$y = 0.96x$
Voronoi	1.033	0.9599	$y = 0.96x$

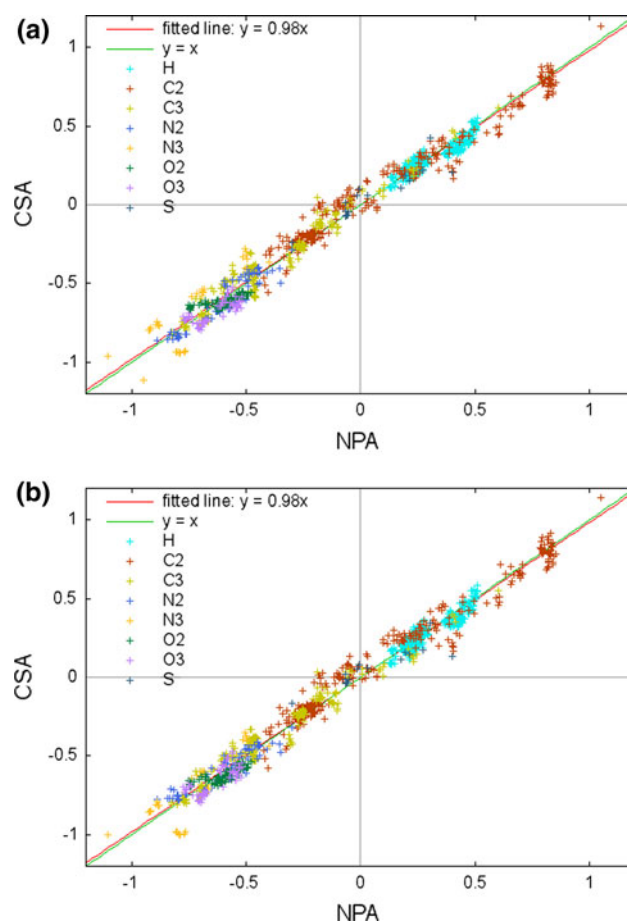
already very well reproduced at the trivial-atoms resolution. For these analyses correlation coefficients are very close to unity (0.987, 0.983, and 0.981, respectively) and fitness function exhibits low absolute values (6.39, 3.03 and 5.14, respectively). The slopes of linear fits are close to unity and are equal to 0.98 in all three cases. The reproduction of Hirshfeld and Voronoi charges is quite good. For these population analyses correlation coefficients (0.879 and 0.907, respectively) and linear fit slopes (0.88 and 0.91, respectively) are lower than for AIM, MPA and NPA charges. Despite that, the absolute value of the fitness function is considerably smaller (2.53 and 2.39, respectively) that is in contradiction with previous observations. The discrepancy between values of fitness function and correlation coefficients can be attributed to charge distributions. Depending on the population analysis employed different ranges for the charges are obtained. AIM charges cover the range between  $-1.5$  and  $2.5e$ . NPA and MPA charges do not exceed  $\pm 1.5$  and  $\pm 1.0e$ , respectively. For Hirshfeld and Voronoi population analyses the absolute values of the charges are small ( $\leq 0.5e$ ). Less spread charges are responsible for decrease in the fitness functions.

It can be seen that hybridization improves the system's description. Reproduction of all population analyses is better. The absolute values of  $S^2$  are lowered. In the same time, correlation coefficients and linear fit slopes increase. The improvement is more pronounced for these population



**Fig. 1** Correlation diagram between Hirshfeld [B3LYP/6-31G(d) level of theory] and CSA-derived charges for trivial-atom (a) and hybridized-atom (b) resolutions

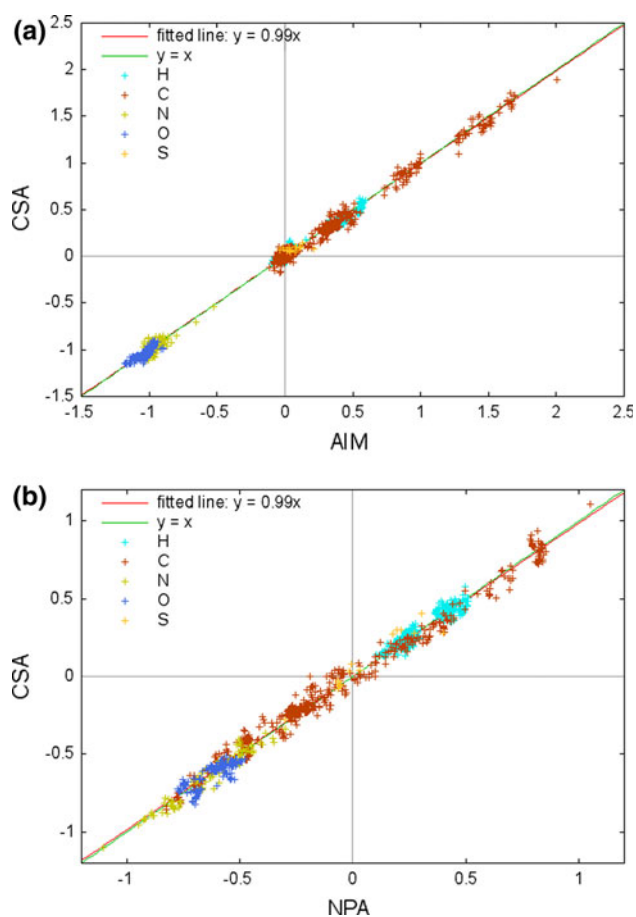
analyses which were worse reproduced in the trivial-atom resolution. For Hirshfeld and Voronoi charges  $R^2$  rises by approximately 0.03. In the case of MPA, NPA and AIM charges the improvement is less pronounced.  $R^2$  rises by  $<0.01$ . Such results are not surprising. Firstly, the increase in the number of optimization parameters always improves the correlation. Secondly, with the introduction of additional atomic types the system is described more accurately. An illustration of this effect is shown in Fig. 1 where correlation plots between CSA and Hirshfeld charges for trivial-atom (a) and hybridized-atom (b) resolutions are presented. It can be seen that Hirshfeld charges of oxygen atoms form two islands corresponding to  $sp^2$  and  $sp^3$  hybridization. This is not captured by CSA in trivial-atom resolution and all charges obtained for oxygen are of similar magnitude. This effect also explains lower values of linear fit slopes obtained for Hirshfeld and Voronoi charges in trivial-atom resolution. Usually lower values of  $y$ -intercepts would suggest that CSA charges are underestimated compared to ab initio reference charges. In this case, however, this is caused by the property of the least square



**Fig. 2** Correlation diagram between NPA [B3LYP/6-31G(d) level of theory] and CSA-derived charges for trivial-atom (a) and hybridized-atom (b) resolutions

method used in the fit and the fact that there were more  $sp^2$  than  $sp^3$  oxygen atoms in the training set. Distinction of  $sp^2$  and  $sp^3$  oxygens allows to fix this problem. In fact, the improvement of correlation between oxygen charges is the most significant factor responsible for the improved reproduction of both Hirshfeld and Voronoi analyses. In the case of MPA and NPA analyses oxygen charges do not form separate islands corresponding to different hybridizations (Fig. 2). Instead, they form overlapping domains and this is the reason why they are well reproduced by CSA in the trivial-atom resolution. Further betterment of correlation between CSA and ab initio charges is achieved by introducing force-field atom types. Here, the largest improvement is observed for carbon and nitrogen atoms, which is the consequence of the very sophisticated distinction between chemical environments introduced in AMBER force field for these elements.

In the first entry of Table 2 optimal effective electronegativities for trivial-atom resolution are collected. It can be seen that different relative electronegativity scales are obtained for different reference charge distributions. For



**Fig. 3** Examples of correlation plots obtained for force-field resolution. AIM (a) and NPA (b) reference charges

**Table 2** The optimized electronegativity ( $\chi_i$  in eV/e) and hardness ( $\eta_i$  in eV/e<sup>2</sup>) data for trival-atom resolution

	MPA	AIM	NPA	Hirshfeld	Voronoi
$\eta_i$					
H	19.98	25.38	18.67	25.02	21.70
C	12.88	14.70	12.51	16.22	17.09
N	19.88	26.83	20.46	30.33	38.50
O	22.45	33.29	25.58	26.79	33.08
S	9.79	24.04	14.68	29.30	16.70
$\chi_i$					
H	10.00	10.00	10.00	10.00	10.00
C	12.12	10.12	12.76	10.80	11.12
N	16.95	25.10	17.74	13.19	15.52
O	18.37	35.69	21.85	15.00	16.63
S	11.12	9.00	10.96	12.99	11.33

MPA and NPA charges, effective electronegativities can be ordered as follows:  $\chi_H < \chi_S < \chi_C < \chi_N < \chi_O$ . Such sequence is in accordance with Pauling electronegativity

scale for isolated atoms. For AIM charges sulfur is less electronegative than hydrogen. In the case of Hirshfeld and Voronoi charges electronegativity of sulfur is higher than that of carbon. One should remember that electronegativity of a given atom in molecule is modified by electrostatic contribution due to the environment and isolated, non-interacting atom limit is not the best reference. The deviations between values of optimal electronegativities obtained for different reference population analyses are the smallest for sulfur and carbon. For oxygen and nitrogen differences between them are more pronounced, especially for AIM charges where very high values of  $\chi_N$  and  $\chi_O$  are observed. Such high values of effective electronegativities of nitrogen and sulfur can be rationalized by analyzing the reference AIM charge distribution. First, in this case partial charges of N and O atoms exhibit relatively large absolute values compared to other ab initio population analyses under consideration (see Fig. 3)—except for two nitrogen atoms they are all over 0.8e. Second, charges obtained for oxygen and nitrogen form an island, separated by over 0.4e from the lowest charge obtained for any other element (or over 0.6e with the exclusion of the aforementioned nitrogen atoms).

The hardness data reported in Table 2 reveals that relative hardness scales also differ for different reference charge distributions. For MPA charges it agrees with the isolated atom limit, namely,  $\eta_O > \eta_N > \eta_C > \eta_S$ . The value of  $\eta_H$  is in between  $\eta_O$  and  $\eta_N$ . For Hirshfeld charges the first three elements share the trend with other population analyses but hardness then drops for O and rises again for S. Similarly to the case of electronegativities, the influence of molecular environment can be invoked to explain these differences. Comparison of the hardness values obtained for different reference charges shows that the biggest differences are observed for nitrogen, oxygen and sulfur atoms. Hardnesses of hydrogen and carbon are less variable between population analyses. Standard deviations ( $\sigma$ ) between parameters obtained for these elements are lower than 3 eV/e<sup>2</sup> compared to 5–8 eV/e<sup>2</sup> obtained for N, O and S. The hardnesses are responsible for resistance on charge flows. The harder is the atom, the stronger is its resistance on CT. Charge distribution for Voronoi and Hirshfeld analyses is less spread than for remaining analyses. Therefore, the hardness parameters for these population analyses are usually higher than for remaining population analyses.

Table 3 collects the effective electronegativities and hardnesses for the hybridized-atom resolution. The number of variation parameters for O, N and C atoms is doubled since two types of hybridization ( $sp^2$  and  $sp^3$ ) were considered. For some of population analyses the hardnesses and electronegativities of the same elements are in disjoint domains. For other populations the atomic domains overlap. The overlapping of electronegativity and hardness domains is more

**Table 3** The optimized electronegativity ( $\chi_i$  in eV/e) and hardness ( $\eta_i$  in eV/e<sup>2</sup>) data for hybridized-atom resolution

	MPA	AIM	NPA	Hirshfeld	Voronoi
$\eta_i$					
H	19.98	24.44	22.42	29.05	20.35
C $sp^2$ (C2)	12.93	14.69	12.50	17.58	17.01
C $sp^3$ (C3)	13.58	15.92	13.20	16.06	17.61
N $sp^2$ (N2)	20.61	25.07	18.80	30.85	50.58
N $sp^3$ (N3)	18.58	24.16	22.95	26.92	51.28
O $sp^2$ (O2)	23.14	42.98	21.76	29.26	36.58
O $sp^3$ (O3)	25.12	30.28	20.23	25.20	35.54
S	9.72	30.41	10.57	25.35	17.12
$\chi_i$					
H	10.00	10.00	10.00	10.00	10.00
C $sp^2$ (C2)	12.09	10.03	13.32	11.27	11.22
C $sp^3$ (C3)	12.35	9.68	13.48	10.92	11.02
N $sp^2$ (N2)	17.38	23.65	17.49	13.74	17.43
N $sp^3$ (N3)	16.31	22.36	20.84	12.79	17.48
O $sp^2$ (O2)	18.40	47.64	21.01	17.46	18.99
O $sp^3$ (O3)	19.78	30.79	19.85	14.13	15.62
S	11.15	7.98	11.73	12.90	11.28

obvious for AMBER force-field atoms shown in Table 4, especially for hardnesses. This indicates that the effective electronegativity and hardness data are strongly dependent on the nearest environment. It can be seen that for hybridized and force-field atom types, likewise the trivial-atom resolution, the obtained values of electronegativities and hardnesses for oxygen and nitrogen atoms are most variable among population analyses. In the hardness domain the most variable element is nitrogen. For this element the standard deviation  $\sigma$  is equal to 13 and 14 eV/e<sup>2</sup> for hybridized and force-field atoms, respectively. In the case of electronegativities oxygen exhibits the biggest differences ( $\sigma = 13$  eV/e for hybridized atoms and  $\sigma_{av} = 11$  eV/e for force-field atoms). Apart from that, it can be seen that for all three resolutions employed hardnesses are much more variable than electronegativities both between population analyses and between atom types corresponding to the same element. For example in the case of hydrogen in the force-field resolution  $\sigma_{av} = 7.5$  eV/e<sup>2</sup> for hardnesses and 1 eV/e for electronegativities. It is not surprising since electronegativity differences determine the direction of charge flow between atoms in molecular system, whereas atomic hardnesses are responsible for the amount of charge transferred between atoms. Hardness parameters are therefore more sensible to both the reference charge distribution employed and chemical environment of a given atomic type.

It is hard to indicate how good the obtained parameterizations are. Therefore, in Fig. 4 we have plotted a histogram illustrating distribution of CSA-derived charges around the reference values  $|q_i^{CSA} - q_i^X|$  ( $X = \text{MPA, NPA, AIM, Hirshfeld and Voronoi charges}$ ) for all molecules from the training set. It can be seen from the figure that Hirshfeld and Voronoi

charges have the sharpest distributions and there are no deviations beyond 0.15e. The same observation is valid for MPA, AIM, and NPA charges. However, the number of atoms in the first region (0,0.05) is smaller than for Hirshfeld and Voronoi analyses. Now, the other two regions, (0.05–0.10) and (0.10–0.15) are more populated than Hirshfeld and Voronoi cases.

Data presented in Tables 1, 2, 3, 4 and Fig. 4 correspond to the training set of molecules. To validate the obtained parameters we have applied CSA to a validation set. None of the molecules from the validation set was in the training set. The validation set included completely new classes of molecules, namely, mono- and disaccharides, lactams, keto acids, thio acids, thioesters, carbamic acid and its derivatives and others. The structure of molecules from the validation set can be found in Ref [17]. We have performed calculation for force-field resolution. All investigated population analyses were taken into account. The results are illustrated in Fig. 5. The obtained distribution is close to the reference values. The Voronoi charge distribution is the sharpest. The differences in CSA-derived charges for Voronoi, Hirshfeld, and AIM population analyses do not go beyond 0.15e as it was observed for the training set of molecules. Slightly worse agreement is observed for the other population schemes.

## Conclusion and future prospects

The extension of CSA to AMBERff99 force-field resolution was performed. The effective electronegativity and hardness data were found using evolutionary algorithms.

**Table 4** The optimized electronegativity ( $\chi_i$  in eV/e) and hardness ( $\eta_i$  in eV/e<sup>2</sup>) data for AMBER99 force-field resolution

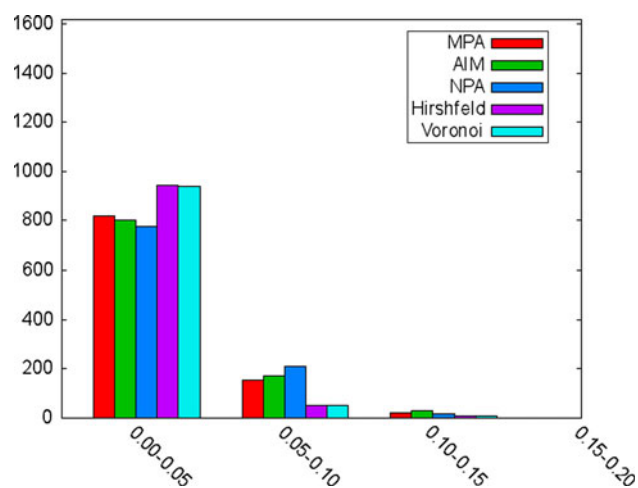
$\eta_i$	MPA	AIM	NPA	Hirshfeld	Voronoi	MPA $\chi_i$	AIM	NPA	Hirshfeld	Voronoi
HC	23.54	28.19	24.23	30.81	23.54	10.00	10.00	10.00	10.00	10.00
H	21.78	33.54	20.57	46.14	31.78	9.95	6.53	10.04	6.04	7.64
HO	18.87	25.84	25.16	35.57	20.79	10.76	8.29	9.75	7.82	9.87
HS	17.02	42.21	17.77	38.05	35.12	10.18	12.28	10.10	9.41	8.88
HP	30.32	28.03	24.81	42.14	28.33	8.03	9.76	9.90	10.71	9.83
H2	17.27	22.40	18.47	45.75	23.30	11.00	9.39	11.26	10.36	10.18
H1	27.03	30.16	24.25	39.83	23.11	9.66	9.38	10.10	9.80	10.08
H5	28.16	27.66	28.80	32.22	24.72	7.90	5.68	7.95	9.15	9.59
H4	19.93	28.01	16.38	26.01	27.52	9.83	9.25	11.34	10.10	9.36
HA	29.10	36.92	22.20	49.59	31.65	9.48	8.95	10.19	9.13	9.39
CT	13.39	16.55	13.84	19.21	17.16	12.44	7.14	14.01	10.99	11.11
C	13.44	16.51	14.02	28.27	25.12	11.61	8.92	12.87	9.02	9.96
CA	11.96	17.12	13.95	15.04	17.42	11.99	8.71	14.00	11.12	11.44
CB	14.62	23.94	12.54	17.73	17.50	10.95	8.06	14.65	10.93	11.74
CC	10.92	15.91	19.02	26.96	23.83	12.16	7.83	14.05	10.56	11.58
CK	14.06	14.61	17.54	19.12	22.31	11.67	9.39	12.56	10.50	11.10
CM	12.18	26.47	14.75	15.79	17.06	12.01	8.93	13.87	11.13	11.19
CD	10.76	12.66	14.34	15.69	24.32	11.79	10.42	14.20	11.08	12.06
CN	14.74	15.64	18.07	20.59	26.12	10.90	8.18	13.98	10.65	11.05
CQ	17.90	19.86	17.15	27.00	24.91	11.70	4.41	12.77	10.23	10.87
CR	13.41	15.46	14.19	21.83	19.07	11.97	9.65	13.36	10.29	11.13
CV	12.64	15.34	16.89	17.64	23.11	12.20	9.94	14.36	11.43	11.36
CW	10.93	18.27	22.69	18.68	21.57	12.14	9.47	14.88	11.23	11.49
CG	13.73	19.07	12.58	16.96	18.93	12.32	9.66	14.16	11.56	11.94
NT	16.18	29.02	19.68	26.45	49.35	15.55	26.51	19.50	12.71	17.54
N3	18.86	18.84	32.30	36.25	64.45	16.41	16.24	26.32	8.98	15.34
N	17.40	26.10	17.35	25.03	42.22	15.75	23.17	16.68	11.88	15.24
NA	16.09	21.95	20.99	51.66	69.11	15.28	20.57	18.50	13.36	17.39
NB	17.90	19.09	19.67	30.20	47.75	16.01	19.14	18.90	14.86	19.41
NC	16.78	22.91	15.87	28.84	44.26	15.62	23.21	17.37	14.76	19.33
N2	17.05	21.15	26.00	45.33	46.48	15.71	20.20	22.32	15.45	16.72
NG	16.86	26.75	24.45	37.49	43.34	15.96	24.38	20.18	11.85	14.17
OH	17.11	24.33	21.54	28.75	48.54	16.32	25.08	21.08	14.40	18.24
OS	21.37	46.57	27.54	43.04	51.36	18.13	43.80	24.33	16.51	18.11
O	14.93	46.84	28.52	35.99	45.63	15.26	50.31	23.95	19.83	21.33
O2	12.32	46.97	30.65	19.95	26.67	14.92	52.41	26.39	14.57	17.83
S	13.60	36.03	17.91	29.32	24.66	10.72	7.42	9.66	12.76	11.74
SH	11.86	35.69	11.41	16.49	16.44	11.57	8.62	11.86	11.97	11.11

Five independent sets of parameters reproducing different population analyses, namely, MPA, AIM, Hirshfeld, Voronoi, and NPA, were reported. Apart from force-field resolution, intermediate hybridized-atom resolution and the least resolved trivial-atom resolution were considered. The 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. The AMBER force-field resolution

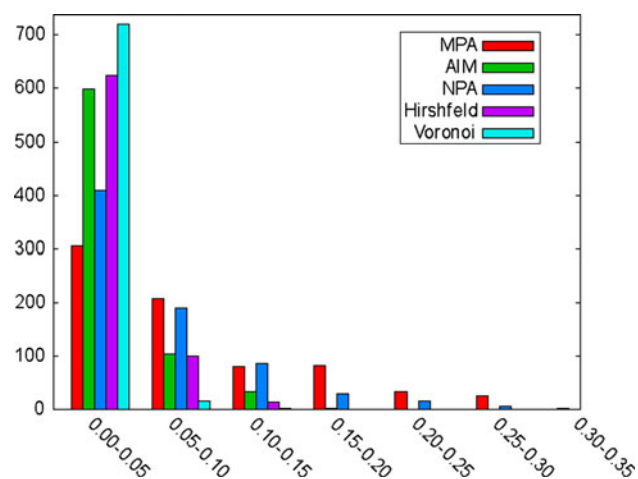
distinguished 38 different chemical environments of H, C, N, O, and S atoms.

Our investigations clearly demonstrate that effective hardness and electronegativities depend on the nearest chemical neighborhood. The Voronoi and Hirshfeld charges were more sensitive on chemical environment than the remaining population analyses. The sharpest distribution around the reference charges was observed for Voronoi and Hirshfeld charges. The most spread was distribution





**Fig. 4** Histograms of the absolute differences between MPA, AIM, NPA, Hirshfeld, Voronoi [B3LYP/6-31G(d) level of theory] and CSA-derived charges for the set of training molecules



**Fig. 5** Histograms of the absolute differences between MPA, AIM, NPA, Hirshfeld, Voronoi [B3LYP/6-31G(d) level of theory] and CSA-derived charges for the set of validating molecules

obtained by NPA charges. This observation was also confirmed by molecules from validating set. CSA-derived charges for MPA parameterization were more spread around the reference MPA charges as compared to the training set. This discrepancy can be explained by the fact that the validation set contained new classes of molecules, none of which were included to the training set. The better performance of the other population analyses can be attributed to smaller changeability of charges within the training and validating sets of molecules.

In the nearest future, we plan to connect force-field CSA with molecular dynamics calculations. We hope that we can adopt our formalism to derive a polarizable force field. Standard force fields used in molecular modeling describe electrostatic interactions in terms of fixed, atom centered,

charges. Real molecules are substantially polarized when placed in a high-dielectric medium. The polarization strongly affects the geometry and energetics of solute molecules. The force fields include polarization only in an averaged way by increasing the atomic charges in order to describe the bulk properties of liquid solvents. In available polarizable models [13, 14, 41–44] either the total field is determined self-consistently via an iterative energy minimization procedure or extended Lagrangian method is applied for polarization degree of freedom. In the latter case the second thermostat is required. There is no need to apply Lagrangian method in CSA formalism. In addition, self-consistency can also be omitted. We plan to derive polarizable force fields based on this formalism and parameterization obtained in this article.

**Acknowledgments** The calculations were carried out with the 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). The authors acknowledge grant from National Science Center (Project No UMO-2011/01/B/ST4/00636).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

## References

- Nalewajski RF, Korchowiec J (1998) Charge sensitivity approach to electronic structure and chemical reactivity. World Scientific Publishing, Singapore
- Parr RG, Yang W (1989) Density functional theory of atoms and molecules. University Press, Oxford
- Geerlings P, De Proft F, Langenaeker W (2003) Chem Rev 103:1793
- Chattaraj PK, Giri S, Duley S (2011) Chem Rev 111:PR43
- Sanderson RT (1951) Science 114:670
- Mortier WJ, Van Genechten K, Gasteiger J (1985) J Am Chem Soc 107:829
- Mortier WJ, Gosh SK, Shankar S (1986) J Am Chem Soc 108:4315
- Rappé AK, Goddard W (1991) J Phys Chem 95:3358
- Bultinc P, Vanholme R, Popelier PLA, de Proft F, Geerlings P (2004) J Chem Phys A 108:10359
- Patel S, Brooks CL III (2004) J Comput Chem 25:1
- Chelli R, Procacci P (2002) J Chem Phys 117:9175
- Chaves J, Barroso JM, Bultinc P, Carbo-Dorca R (2006) J Chem Inf Model 46:1657
- Rick SW, Stuart SJ, Bader JS, Berne BJ (1995) J Mol Liq 66/66:31
- Rick SW, Berne BJ (1996) J Am Chem Soc 118:672
- Nalewajski RF, Korchowiec J, Tokarz R, Broclawik E, Witko M (1992) J Mol Catal 77:165
- Korchowiec J, Kowalski P (1993) Chem Phys Lett 208:135
- Korchowiec J, Gerwens H, Jug K (1994) Chem Phys Lett 222:58
- Parr RG, Donnelly RA, Levy M, Palke WE (1978) J Chem Phys 69:4431

19. Parr RG, Pearson RG (1983) *J Am Chem Soc* 105:7512
20. Parr RG, Yang W (1984) *J Am Chem Soc* 106:4049
21. Stachowicz A, Styrz A, Korchowiec J (2011) *J Mol Model*. doi: [10.1007/s00894-011-1006-7](https://doi.org/10.1007/s00894-011-1006-7)
22. Ohno K (1964) *Theor Chim Acta (Berlin)* 2:219
23. Mulliken RS (1934) *Chem Phys* 2:782
24. Svobodová Vařeková R, Jiroušková Z, Vanek J, Suchomel Š, Koča J (2007) *Int J Mol Sci* 8:572
25. Jiroušková Z, Svobodová Vařeková R, Vanek J, Koča J (2009) *J Comput Chem* 30:1174
26. Korchowiec J (2005) *J Chem Sci* 117:491
27. Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S Jr, Weiner P (1984) *J Am Chem Soc* 106:765
28. Mulliken RS (1955) *J Chem Phys* 23:1833
29. Chirlian LE, Francl MM (1987) *J Comput Chem* 8:894
30. Breneman CM, Wiberg KB (1990) *J Comput Chem* 11:361
31. Hirshfeld FL (1977) *Theor Chim Acta* 44:129
32. te Velde G (1990) Numerical integration and other methodological aspects of bandstructure calculations. In: *Chemistry. Vrije Universiteit, Amsterdam*
33. Bickelhaupt FM, Eikema Hommes NJR, Fonseca Guerra C, Baerends EJ (1996) *Organometallics* 15:2923
34. Reed AE, Weinstock RB, Weinhold F (1985) *J Chem Phys* 83:735
35. Bader RFW (1990) *Atoms-in-molecules*. University Press, Oxford
36. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) *J Am Chem Soc* 117:5179
37. Junmei W, Cieplak P, Kollman PA (2000) *J Comput Chem* 21:1049
38. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery Jr JA, Peralta JE, Ogliaro F, Bearpark M, Heyd JJ, Brothers E, Kudin KN, Staroverov VN, Kobayashi R, Normand J, Raghavachari K, Rendell A, Burant JC, Iyengar SS, Tomasi J, Cossi M, Rega N, Millam NJ, Klene M, Knox JE, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Martin RL, Morokuma K, Zakrzewski VG, Voth GA, Salvador P, Dannenberg JJ, Dapprich S, Daniels AD, Farkas Ö, Foresman JB, Ortiz JV, Cioslowski J, Fox DJ (2009) *Gaussian 09, Revision A.1*. Gaussian, Inc., Wallingford
39. te Velde G, Bickelhaupt FM, van Gisbergen SJA, Fonseca Guerra C, Baerends EJ, Snijders JG, Ziegler T (2001) *J Comput Chem* 22:931
40. Adcock S. <http://gaul.sourceforge.net>
41. Caldwell J, Dang LX, Kollman PA (1990) *J Am Chem Soc* 112:9144
42. Sprik M, Klein ML (1988) *J Chem Phys* 89:7556
43. Dick BG Jr, Overhauser AW (1958) *Phys Rev* 112:90
44. Pratt LR (1980) *Mol Phys* 40:347