JOURNAL OF PHYSICAL ORGANIC CHEMISTRY
J. Phys. Org. Chem. 2001; **14:** 355–369
DOI:10.1002/poc.378

Novel point charge models: reliable instruments for molecular electrostatics

Alexander A. Oliferenko, Vladimir A. Palyulin, Sergei A. Pisarev, Alexei V. Neiman and Nikolai S. Zefirov*

Department of Chemistry, Moscow State University, Moscow 119899, Russia

epoc

ABSTRACT: A novel method to compute atomic charge distribution is proposed. The starting data for the computation are the topology of a molecule and electronegativities of either constituent atoms or atomic orbitals. The topology of a molecule is represented by a molecular graph or, in a more advanced model, by an orbital graph that reflects an orbital constitution of atoms. The distinctive feature of the method is that the starting atomic (orbital) electronegativities equilibrate in the same fashion that electrical potentials do in the nodes of a closed electrical network. Thus, the well-developed formalism of the theory of electrical circuits was applied for computation of partial atomic charges. In addition, the problem of parametrization of halogens was solved using a specially developed procedures. Comparison of the obtained charges with those produced by means of various computational schemes [Partial Equalization of Orbital Electronegativities (PEOE), Mulliken population analysis, natural population analysis, Bader's AIM, method of generalized atomic polar tensors (GAPT) and an electrostatic potential-based method, CHELPG] and with some molecular spectral characteristics proves that both models, but especially the orbital graph charge model, provide fast, convenient and reliable methods to calculate theoretically justifiable atomic charges for a variety of chemical species. The developed charge models are well suited for application in many areas of molecular modeling and QSAR/QSPR studies.
Additional material for this paper is available from the epoc website at http://www.wiley.com/epoc

KEYWORDS: atomic charges; molecular electrostatics; electronegativity equilibration; topological models of chemical structures; predictions of molecular properties

INTRODUCTION

Few chemical concepts are as enduring and widely used as that of partially charged atoms in molecules. Electronic charge distribution of chemical systems is considered an extremely powerful tool to describe physico-chemical properties, reactivity and biological activity of molecules.[1–4] Thus, a knowledge of charge distribution is fundamental for molecular modeling of chemical systems.

The model of point charges on atoms is the most vital for numerous applications. In the framework of this model the electron density of a molecule is partitioned into atomic components: the so-called partial charge measured in fractions of the elementary charge is assigned to each atom in the molecule. Over many years there were numerous attempts to provide a quantitative background to the intuitively clear notion of the partial atomic charge as a characteristic associated with an atom in a molecule that reflects a local excess or deficit of the electronic charge.[5–12]

One should clearly understand that the notion of atomic charges is shrouded in uncertainty because the atomic charge is not a physical observable, and there does not exist a quantum-mechanical operator whose expectation value would be the atomic charge. Hence obtaining unique exact values of partial charges involves fundamental difficulties because there is no unique way to localize the global charge distribution in a molecule on its constituent atoms. The interpretation of the charges obtained by empirical, semiempirical or non-empirical calculations is completely arbitrary, and all of them coexist as useful tools for chemical reasoning.

Nonetheless, the point charge approximation is of great importance in the description of electrostatic properties of molecules, so the development of new advanced methods of atomic charge calculation has become a subject of permanent interest to chemists.

The most apparent way to derive atomic charges is to analyze the orbital populations obtained as a result of MO-LCAO calculations (either non-empirical or semiempirical). Mulliken population analysis,[5] which assigns the entire electron density to atom-centered orbitals, is recognized to be both the most widely used and the most criticized method.[13,14] Arbitrariness in dividing overlap population, a strong dependence on the basis set, and

**Correspondence to:* N. S. Zefirov, Department of Chemistry, Moscow State University, Moscow 119899, Russia.
E-mail: zefirov@org.chem.msu.su

sometimes unrealistic orbital populations (more than two or less than zero) stimulated many attempts[6–9] to elaborate an MO-LCAO computational scheme free of this kind of disadvantage.

Another family of methods is based on a theoretical analysis of experimentally or computationally obtainable physical observables. The integration of the electron density according to Bader's theory of atoms in molecules (AIM)[10] enables one to derive theoretically well-defined atomic charges. The electron density can be built up from an *ab initio* calculation or, in principle, from an accurate x-ray diffraction experiment. Fitting a positively charged probe to a molecular electrostatic potential obtained *ab initio* or semiempirically is another suitable approach to the evaluation of charge distribution.[11] Another way to obtain atomic charges is to analyze atomic polar tensors that may be derived from experimental IR bands[12a] or calculated theoretically.[12b]

Finally, there exists one more group of computational schemes that makes use of different classical approximations to simulate the electron distribution in a molecule.[15] Empirical parameter-based methods for the atomic charge evaluation operate fast and usually involve different manipulations with such atomic and molecular characteristics as molecular connectivity,[16] electronegativity,[17–26] polarizability,[24] dipole moments[25] or covalent radii.[20] The development of similar methods is invoked by the fast growth of such disciplines as medicinal chemistry, QSAR and computer-aided molecular design that need efficient tools for the fast and reliable charge distribution evaluation for a large number of molecules, including large-sized molecules, because the time cost necessary for an *ab initio* or even semiempirical calculation turns out to be unacceptable (computational time of a non-empirical run increases in proportion to near the fourth power of the number of basis functions involved[13]).

We will further concentrate our attention on the methods based on electronegativity (EN). Electronegativity, being traditionally perceived as a power attracting electrons to an atom,[27] is intimately related to partial atomic charges. It always seems attractive to derive this relationship as quantitatively as possible. Indeed, EN equalization has become a basic logic and algorithmic technique for deriving atomic charges,[15] and it is worth inspecting the development of this idea over the years.

For the first time the geometric mean principle of electronegativity equalization was suggested by Sanderson,[17] and the idea of full EN equalization on the formation of a molecule turned out to be very fruitful and promising.[15] However, in a practical sense the Sanderson method was limited only to small, mainly inorganic molecules, because it did not account for molecular connectivity. In other words, the atoms of the same type (say all carbons, oxygens, nitrogens, etc.) in a molecule attained the same charges, in spite of their structural difference. This shortcoming has been overcome with a number of approaches[21–24,26] that had fallen back on the notion of molecular graph[28] expressed, for instance, in the form of a connectivity matrix. [The connectivity (adjacency) matrix of a graph, usually designated as $A = \{a_{ij}\}$, is a square symmetrical $n \times n$ matrix whose element $a_{ij} = 1$ if the corresponding vertices of the graph ν_i and ν_j are connected by an edge, and $a_{ij} = 0$ if they are not connected; once the vertices are connected by a multiple edge, the corresponding matrix entry is equal to the edge multiplicity.]

We would like to consider here several approaches, which make use of the molecular graph model, that enable one to account explicitly for the molecular connectivity. First, the method suggested by Kirpichenok and Zefirov[23] allows Sanderson's electronegativity equalization principle to be augmented to include rather arbitrary molecular systems. Progress was made towards the charge calculation for branched non-cyclic systems including species with multiple bonds. This approach has found application in QSAR-related calculations.[29]

Another procedure of EN equalization was realized within the method of Partial Equalization of Orbital Electronegativities (PEOE) developed by Gasteiger and Marsili.[21] In the PEOE method, an iteration equation based on the notion of the orbital EN[30] was proposed, whereas the connectivity of constituent atoms was the only molecular structural information. We should emphasize that the PEOE method has become the most useful one to date for correlation with molecular properties and has found extremely wide application in a number of QSAR/QSPR and molecular design computer programs, e.g. PETRA[31] and SYBYL[32] and others.[33]

In this work we develop a new approach to the problem of molecular charge distribution expressed in the form of point atomic charges. The key point of the suggested method is an unconventional technique used for the EN equilibration. We consider a molecule as a closed system of electrically interacting atoms and formulate the atomic charge problem in terms of classical physics. In other words, we represent the molecule (strictly the molecular graph) as a closed electrical network and analyze it with the help of appropriate laws of electrical current. As will be demonstrated in subsequent sections, this model based on a rather provisional analogy allows one to obtain reasonable atomic charges for many types of molecules in a fast and efficient way. (The history of modern chemistry exhibits a number of bright examples of fruitful cooperation between classical and quantum-mechanical paradigms. Molecular mechanics,[36] combined QM/MM schemes[37] and classical potential functions for treating intermolecular interactions[38] are good examples.)

METHOD

Two very different aspects of the charge calculation problem are the goal of this paper. First, the task is to elaborate a novel, different technique of electronegativity equalization or, better equilibration (because our technique provides only a partial equalization, we prefer to name it *equilibration*). Second, a formal model for the chemical structure representation should be judiciously chosen. Such a model must be in conformity with the computational procedure involved and cover the diversity of chemical structures as far as possible.

Topological models of chemical structure

In this paper we consider two possibilities of representation of molecular structure. The simplest and most straightforward way is to use the well-known notion of molecular graph (MG).[28] Most of the methods discussed before are based on the MG model and we also will use it.

However, fine features of molecular charge distribution cannot be covered by the simple MG model, since it can take into account only bonding orbitals, but not free electrons, lone pairs, etc. Thus, the second model we have developed is based on the notion of the orbital graph (OG).

OG is a special graph constructed starting from the initial MG by substituting each MG vertex (point-like atom) by a complete graph containing as many vertices as there are valence shell atomic orbitals centered on the corresponding atom. All vertices of such a complete graph are pairwise connected by edges whose weights are equal to unity; this allows for the integrity of the electronic constitution of an atom. Furthermore, they are connected with corresponding vertices of other similar complete graphs by edges whose weights are equal to unity (for σ-bonds and non-conjugated π-bonds) in accordance with interatomic orbital overlaps. Some examples of molecular graphs and corresponding orbital graphs are shown in Fig. 1.

Being applied to simple organic molecules, the idea of the OG gives the graph pictures that are reminiscent of the old representation of molecular structures as tetrahedra connected by vertices, by edges and by faces in the case of ordinary [Fig. 1(a)], double [Fig. 1(b)] and triple bond [Fig. 1(c)], respectively. We especially emphasize that our model can include also doubly occupied [Fig. 1(d), (e) and (g)] and even vacant orbitals [Fig. 1(f)] in addition to bonding orbitals.[21,24] The idea behind this was to be able to apply the OG model, accounting for various orbital interactions (for example, d_π–p_π), for charge calculations in the case of coordinate and hypervalent compounds. The chemical bonding with the participation of d-orbitals is shown in Fig. 1(e).

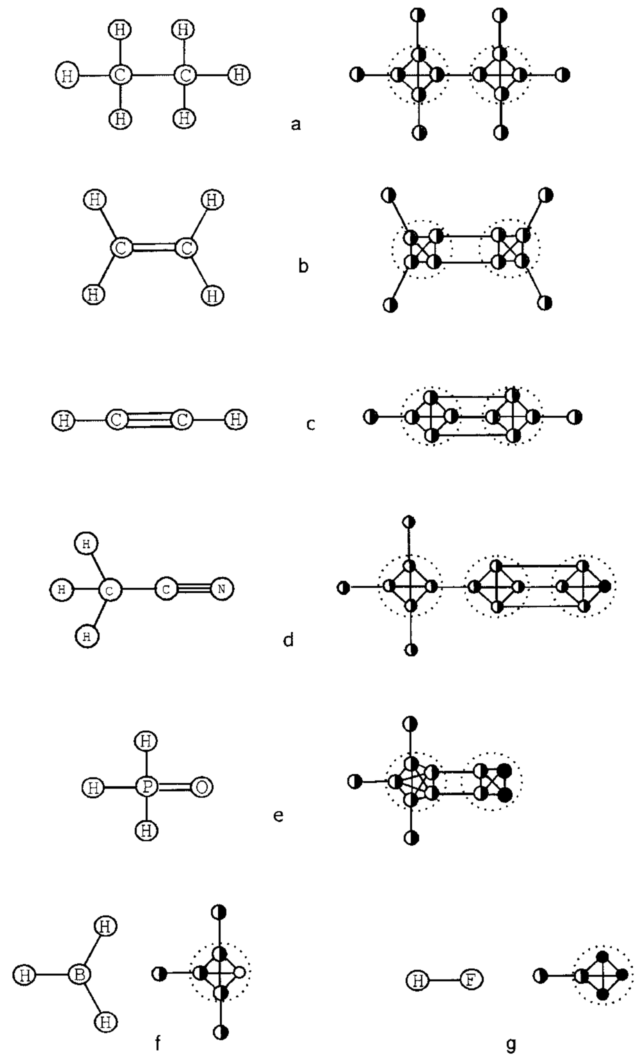


Figure 1. Molecular graphs for (a) ethane, (b) ethylene, (c) acetylene, (d) acetonitrile, (e) phosphine oxide, (f) borane and (g) hydrogen fluoride and corresponding orbital graphs (half-filled, filled and empty circles designate bonding, lone pair and vacant orbitals, respectively)

Model of electronegativity equilibration

In this section we describe a general model, which is common to both the straightforward MG charge model (MGC model) and the advanced OG charge model (OGC model), and permits one to obtain atomic (orbital) equilibrated electronegativities starting from molecular topology and standard EN values by means of a special procedure based on the theory of electrical circuits.

In general, the model of charge distribution in a molecule rests on the following assumptions: (i) the molecule is considered as a topological object, i.e. an object characterized only by its connectivity (no geometric characteristics are taken into account); (ii) the charge density is assumed to be distributed over the molecular framework in accordance with the electronegativities of the constituent atoms; (iii) the electronegativities of the constituent atoms tend to be equilibrated in the same manner as electrical potentials in the nodes of a closed electrical network. The last statement means that one may use the well-known

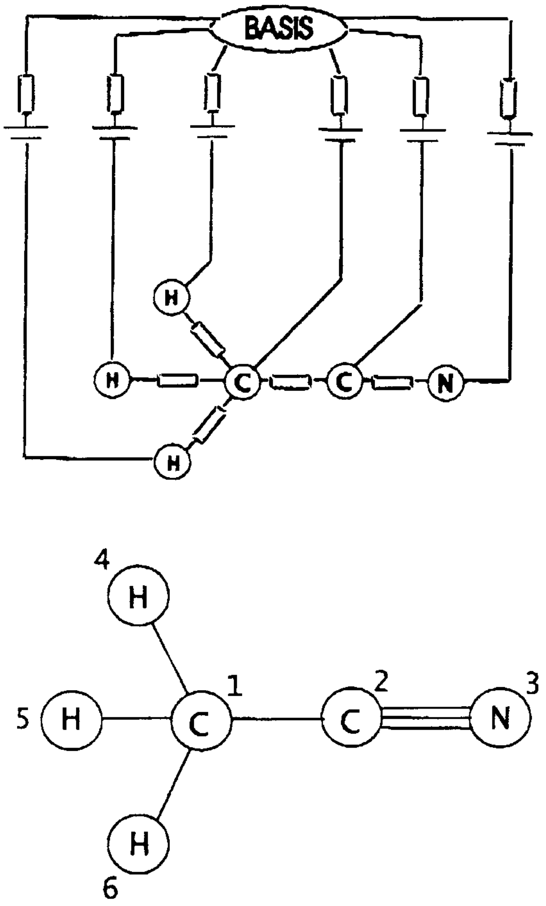


Figure 2. Molecular graph and corresponding electrical network of the acetonitrile molecule

mathematical formalism of the theory of electrical circuits[39] in order to simulate an EN equilibration. Thus, the molecule is considered as an electrical network composed of the nodes (junction points) that correspond to atoms (orbitals) and conducting sites that simulate chemical bonds.

General algorithm

Using the above concept, one may accept that the MG (OG) vertices correspond to nodes of the electrical network and the edges correspond to conducting branches (strictly, not all graph vertices can be deemed nodes, only those having a degree of >1; however, without loss of generality, we adopt this term for all vertices). In the next step, one should close the circuit and insert current sources into it. For this purpose we introduce an additional artificial node termed 'basis node' into the circuit and connect it with all other nodes by special branches having resistance and current sources. In terms of graph theory we introduce an additional vertex into the graph and connect it by edges with all vertices. Figure 2 demonstrates the transformation of the acetonitrile molecular graph into an electrical network.

Now consider the application of the nodal potentials method[39] to analyze the network obtained. Given the topology of a network and its electrical characteristics (currents and conductances), one could derive a set of linear simultaneous equations for all n nodes except the basis $(n+1)$th node to determine the nodal potentials φ_j in accordance with Kirchhoff's first law. The potential of the 'basis' node is taken as zero, and the currents are written according to Ohm's law as follows:

$$\sum_{i=1}^{n} Y_{ij}\cdot\varphi_j = \sum_{i} J_i \qquad (1)$$

where Y_{ij} is the negative of conductance of a branch connecting nodes i and j, Y_{ii} is the sum of conductances attached to node i and the sum in the right-hand side consists of algebraic sums of all currents coming to node i.

Rewriting this expression in a more suitable matrix form we have the following matrix equation:

$$\mathbf{Y\Phi} = \mathbf{J} \qquad (2)$$

where $\mathbf{\Phi}$ is the column matrix of the sought potentials, $\mathbf{Y}$ is the matrix of nodal conductance (being complex in the general case) and $\mathbf{J}$ is the column matrix of nodal currents. Provided that matrix $\mathbf{Y}$ is not degenerate, the single solution of these equations is $\mathbf{\Phi} = \mathbf{Y}^{-1}\,\mathbf{J}$.

Let us discuss in more detail the structure of matrix $\mathbf{Y}$. It is a square matrix whose diagonal element Y_{ii} is equal to a sum of (complex) conductances of all the branches attached to this node i, whereas its off-diagonal entry $Y_{ij} = Y_{ji}$ is equal to a sum of all conductances of all the branches connecting i and j.

Applying the above electrical engineering formalism to the case of graphs (this formalism was first applied to molecular graphs by Golender *et al.*[40] in order to attack the important problem of automorphism partitioning of molecular graph vertices), we have to consider the particular case of matrix $\mathbf{Y}$, in which all conductances are real numbers and equal in particular cases to the multiplicity of the corresponding graph edges. Let us label such a matrix as $\mathbf{S}$. This specific matrix can be constructed as $\mathbf{S} = -\mathbf{A} + \mathbf{D} + \mathbf{I}$, where $\mathbf{A}$ is the connectivity matrix of the graph;[28] $\mathbf{D} = \{d_{ij}\}$ is the degree matrix, i.e. a square diagonal matrix whose diagonal entry is equal to the degree of the corresponding vertex, and all off-diagonal entries are equal to zero; $\mathbf{I}$ is the identity matrix, i.e. a square matrix whose all diagonal entries are equal to 1, and all off-diagonal entries are zeros. All these matrices are of order n, where n is the number of graph vertices. In fact, matrix $\mathbf{S}$ is a very simplified real-valued variant of the matrix of nodal conductances $\mathbf{Y}$. The explicit derivation of matrix $\mathbf{S}$ for the acetonitrile molecular graph (Fig. 2) is shown below:

$$\overset{\mathbf{D}}{\begin{vmatrix} 4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{vmatrix}} - \overset{\mathbf{A}}{\begin{vmatrix} 0 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 3 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{vmatrix}} + \overset{\mathbf{I}}{\begin{vmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{vmatrix}} = \overset{\mathbf{S}}{\begin{vmatrix} 5 & -1 & 0 & -1 & -1 & -1 \\ -1 & 5 & -3 & 0 & 0 & 0 \\ 0 & -3 & 4 & 0 & 0 & 0 \\ -1 & 0 & 0 & 2 & 0 & 0 \\ -1 & 0 & 0 & 0 & 2 & 0 \\ -1 & 0 & 0 & 0 & 0 & 2 \end{vmatrix}} \quad (3)$$

Rewriting Eqn. (2), as applied to a molecule, but not to an electrical network, we take standard atomic (orbital) electronegativity values, χ_i^0, as the components of right-hand side column matrix $\mathbf{X}^0$, whereas the column matrix $\mathbf{X}$ is composed of unknown electronegativities of atoms (orbitals) in a molecule, χ_i:

$$\mathbf{SX} = \mathbf{X}^0 \quad (4)$$

Solving these equations in the same manner as for Eqn. (2), namely $\mathbf{X} = \mathbf{S}^{-1}\,\mathbf{X}^0$, we obtain new electronegativity values perturbed by the molecular environment. Further treatment of the problem depends on the graph model chosen.

Molecular graph charge model[41]

In the case of the MG model the partial atomic charges are derived as dimensionless values that are numerically equal to relative changes of the atomic ENs upon the formation of the molecule. These quantities are defined as follows:

$$\mathbf{Q} = \frac{\mathbf{X} - \mathbf{X^0}}{\chi_\mathrm{m}} = \frac{\Delta\mathbf{X}}{\chi_\mathrm{m}} \quad (5)$$

where $\mathbf{Q}$ is the vector of atomic charges, $\mathbf{X^0}$ and $\mathbf{X}$ are the vectors of standard atomic EN values and those in a valence state, respectively and χ_m is the geometric mean of atomic electronegativities. Due to the above definition, the charges obtained obey the following normalization conditions: (a) $\sum_i^n q_i = 0$;(b) $|q_i| < 1$.

Therefore, the computational scheme designed on the basis of the simple MG model is self-consistent and needs no iteration steps as compared with the PEOE method of Gasteiger and Marsili[21] or the method of Abraham *et al.*[24b]

Let us illustrate the above theory. Taking particularly the Pauling electronegativities (tabulated in Ref. 42) one can fill in the entries of the vector $\mathbf{X^0}$ in order to write Eqn. (4) in an explicit way; here the vector $\mathbf{X}$ contains unknowns:

$$\overset{\mathbf{S}}{\begin{vmatrix} 5 & -1 & 0 & -1 & -1 & -1 \\ -1 & 5 & -3 & 0 & 0 & 0 \\ 0 & -3 & 4 & 0 & 0 & 0 \\ -1 & 0 & 0 & 2 & 0 & 0 \\ -1 & 0 & 0 & 0 & 2 & 0 \\ -1 & 0 & 0 & 0 & 0 & 2 \end{vmatrix}} \times \overset{\mathbf{X}}{\begin{vmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{vmatrix}} = \overset{\mathbf{X^0}}{\begin{vmatrix} 2.55 \\ 2.55 \\ 3.04 \\ 2.20 \\ 2.20 \\ 2.20 \end{vmatrix}} \quad (6)$$

One should solve this matrix equation with respect to the column matrix $\mathbf{X}$. A routine procedure to solve such equations is to act on the column matrix $\mathbf{X^0}$ with a matrix inverse to $\mathbf{S}$, i.e. $\mathbf{S^{-1}}$:

$$\begin{array}{cccccc}
\mathbf{X} & & \mathbf{S^{-1}} & & \mathbf{X^0} & & \mathbf{X} \\
\begin{vmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{vmatrix} & = & \begin{vmatrix} 0.319 & 0.116 & 0.087 & 0.159 & 0.159 & 0.159 \\ 0.116 & 0.406 & 0.304 & 0.058 & 0.058 & 0.058 \\ 0.087 & 0.304 & 0.478 & 0.043 & 0.043 & 0.043 \\ 0.159 & 0.058 & 0.043 & 0.58 & 0.08 & 0.08 \\ 0.159 & 0.058 & 0.043 & 0.08 & 0.58 & 0.08 \\ 0.159 & 0.058 & 0.043 & 0.08 & 0.08 & 0.58 \end{vmatrix} & \times & \begin{vmatrix} 2.55 \\ 2.55 \\ 3.04 \\ 2.20 \\ 2.20 \\ 2.20 \end{vmatrix} & = & \begin{vmatrix} 2.425 \\ 2.638 \\ 2.739 \\ 2.313 \\ 2.313 \\ 2.313 \end{vmatrix}
\end{array} \qquad (7)$$

The geometric mean of atomic ENs involved was calculated as being $\chi_m = 2.439$. Hence in its final form the solution of Eqn. (5) is as follows:

$$\begin{array}{ccc}
\Delta\mathbf{X} & {\chi_m}^{-1} & \mathbf{Q} \\
\begin{vmatrix} -0.125 \\ 0.088 \\ -0.301 \\ 0.113 \\ 0.113 \\ 0.113 \end{vmatrix} & \times 1/2.439 = & \begin{vmatrix} -0.051 \\ 0.036 \\ -0.124 \\ 0.046 \\ 0.046 \\ 0.046 \end{vmatrix}
\end{array} \qquad (8)$$

As can be seen from Eqn. (8), the charges in the acetonitrile molecule are as follows: −0.051 at the methyl carbon atom, +0.036 at the nitrile carbon atom, −0.124 at the nitrogen and +0.088 at all methyl hydrogens.

Orbital graph charge model

As we have mentioned above, the first stage of the charge calculation, namely electronegativity equilibration, is common for both the MG and OG models. However, the final derivation of atomic charges within the OG model differs significantly, because here we use more advanced parameters, namely the orbital electronegativity and chemical hardness. From the standpoint of DFT[34,35] the electronegativity is considered as the first partial derivative of the total energy, E, with respect to the occupational number, n, with the external potential being constant. The hardness is defined as the second derivative of E. The finite-difference approximations for EN and chemical hardness are as follows:[43]

$$\chi = -\left[\frac{\partial E}{\partial n}\right]_V \approx \frac{I + A}{2} \qquad \eta = \left[\frac{\partial^2 E}{\partial n^2}\right]_V \approx I - A \qquad (9)$$

where I is the ionization potential and A is the electron affinity of an atom. As for a particular form of E, a number of functions were considered,[30b,44–46,] all of them are more or less warranted within Slater's expansion of the energy in terms of the powers of atomic charge.[47] As noted by Kohn *et al.*,[48] within a certain subsystem the energy E can be considered as a continuous and differentiable function of n. This point lends support to the validity of the notion of orbital EN introduced by Hinze and co-workers.[30] The energy expansion was limited to the quadratic term:

$$E(n_i) = a + bn_i + cn_i^2 \qquad (10)$$

Whence the charge dependence of the electronegativity of the ith orbital, χ_i, was expressed in terms of the parameters b and c which are, in fact, electronegativity ${\chi_i}^0$ and hardness ${\eta_i}^0$ of a corresponding unperturbed orbital, and n_i is its occupation number being the negative of the orbital charge, q_i:

$$\chi_i = \chi_i^0 + \eta_i^0 q_i \qquad (11)$$

Tabulated values of χ^0 and η^0 for many pure and hybridized states can be found elsewhere[30,46]. In such a way the orbital graph vertices are assigned appropriate EN values corresponding to σ, π, lone pair, unpaired electron or vacant orbitals. Substituting these values in Eqn. (4) gives the desirable equilibrated orbital ENs, χ_i, for all atoms in a molecule. Given those equilibrated orbital EN values and tabulated values of hardness, the orbital charges can be expressed from Eqn. (11) as follows:

$$q_i = \frac{\chi_i - \chi_i^0}{\eta_i^0} \qquad (12)$$

Obviously, the total charge of the jth atom should be written as a sum of the orbital charges:

$$Q = \sum\nolimits_i^A q_i \qquad (13)$$

Table 1. Summary on the charge calculation of acetonitrile with the orbital graph charge (OGC) model[a]

Atom	Orbital	EN[b] (χ^0) (V)	Equil. EN[c] (χ) (V)	Hardness[b] (η), (V e^{-1})	Cov. rad.[d] (Å)	Norm (η)[e] (V e^{-1})	Orbital charge[f] e	Atomic charge[g, e]
C	σ(C–C)	8.060	8.109	6.565	0.77	6.576	0.007	−0.064
	σ(C–H)	8.060	7.904	6.565		6.529	−0.024	
	σ(C–H)	8.060	7.904	6.565		6.529	−0.024	
	σ(C–H)	8.060	7.904	6.565		6.529	−0.024	
C	σ(C–N)	10.47	9.119	6.972	0.60	6.614	−0.204	0.265
	π(C–N)	5.730	8.086	5.473		6.452	0.365	
	π(C–N)	5.730	8.086	5.473		6.452	0.365	
	σ(C–C)	10.47	8.774	6.972		6.529	−0.260	
N	n	7.370	8.748	9.885	0.55	10.19	0.135	−0.365
	σ(N–C)	15.59	10.18	8.732		7.144	−0.757	
	π(N–C)	7.870	8.721	6.325		6.624	0.128	
	π(N–C)	7.870	8.721	6.325		6.624	0.128	
3 × H	σ(H–C)	7.18	7.542	6.422	0.37	6.605	0.055	0.055

[a] Normalizing factor D_m is equal to 0.81164 as follows from Eqn. (A8).
[b] Taken from Ref. 46.
[c] Calculated from Eqn. (4).
[d] Taken from Ref. 50.
[e] Calculated from Eqn. (16).
[f] Calculated from Eqns (19) and (A9).
[g] Calculated from Eqn. (13).

However, problems arise if one tries to compute atomic charges in such a way. In the general case the electroneutrality condition for the molecule can be violated: $\Sigma_j^N Q_j \neq 0$, i.e. the procedure has ceased to be self-consistent. One of the reasons is that the quadratic form of the energy function does not account for any variations in atomic size upon ionization[45]. In order to remedy this problem we considered a more accurate cubic form of the energy function:

$$E(n_i) = a + bn_i + cn_i^2 + dn_i^3 \tag{14}$$

On differentiation, this equation gives the quadratic dependence of orbital EN on the atomic charge:

$$\chi_i = \chi_i^0 + \eta_i^0 q_i + dq_i^2 \tag{15}$$

It is apparent that a successive differentiation of Eqn. (14) will result in the following linear charge dependence of the hardness:

$$\eta_i = \eta_i^0 + dq_i \tag{16}$$

Where the correction term dq_i should be defined in a reasonable way. An important point is that η_i should be a function of d. As the hardness varies inversely with the atom size,[45] it would be justified to establish d_j as the reciprocal of the covalent radius, r_j, of the jth atom under consideration:

$$d_j = \frac{D_{\mathrm{M}}}{r_j} \tag{17}$$

where D_{M} is the normalization factor common for all atoms in the molecule. The introduction of D_{M} will provide the electroneutrality condition for the molecule. From Eqns (11) and (17), a set of $n \times k$ simultaneous equations can be composed for determining the coefficients d and the atomic charges *per se*:

$$\begin{cases} d_1 q_{11}^2 + \eta_1^0 q_{11} + \chi_{11}^0 - \chi_{11} = 0 \\ \vdots \\ d_n q_{nk}^2 + \eta_{nk}^0 q_{nk} + \chi_{nk}^0 - \chi_{nk} = 0 \\ \sum_j^N Q_j = 0, \pm 1, \ldots \end{cases} \tag{18}$$

A more detailed description of the procedure of computing the coefficients d is given in the Appendix. In addition, standing the normalization condition in Eqn. (18) for ±1, ±2, etc., allows one to calculate the charge distribution for ionic structures. Finally, the orbital

Table 2. Halogen atomic graphs and calculated hybrid orbitals

Atomic graph	F	Cl	Br	I
∠XAE(°)	99.63	103.5	100.8	101.5
sp Hybrid	sp$^{16.44}$	sp$^{7.58}$	sp$^{12.65}$	sp$^{11.05}$
s Character, %	5.7	11.7	7.33	8.3
Orbital EN (V)	13.24	10.52	9.03	8.66
Hardness (V e^{-1})	8.63	5.61	4.73	4.46
∠EAE(°)	117.2	114.6	116.5	116.1
sp Hybrid	sp$^{2.18}$	sp$^{2.40}$	sp$^{2.24}$	sp$^{2.27}$
s Character, %	31.4	29.4	30.9	30.5
Orbital EN (V)	18.16	12.27	11.36	10.35
Hardness (V e^{-1})	8.18	5.46	4.51	3.96

charge can be computed according to the following equation:

$$q_i = \frac{\chi_i - \chi_i^0}{\eta_i} \qquad (19)$$

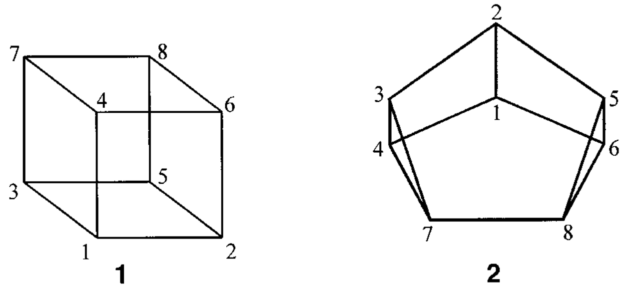
where η_i is the normalized hardness from Eqn. (16).

The above theory can be illustrated by the picture of acetonitrile orbital graph [Fig. 1(e)] and by the details of the acetonitrile charge calculation listed in Table 1.

PARAMETRIZATION

Parameters for the MGC model

In order to obtain the best set of parameters, a number of atomic EN scales were tested, including those developed by Pauling,[27,42] Sanderson,[17] Allred and Rochow[42] and Luo and Benson.[49] The Pauling scale turned out to be the best with regard to correlations with molecular spectral properties such as ESCA chemical shifts and ^{35}Cl NQR coupling constants. Thus, for the single-parameter MGC model we finally selected Pauling's ENs[42] without any changes or additions.



1 **2**

Figure 3. Molecular graphs for cubane (**1**) and cuneane (**2**) (hydrogen atoms are depleted)

Parameters for the OGC model

The parameters needed for the OG model are orbital electronegativity and hardness and the covalent radii of the atoms involved. As a main source of the parameters we used the list of orbital EN and hardness values computed by Bergmann and Hinze.[46] Covalent radii were taken from the work of Gillespie and Hargittai.[50]

Despite the abundance of valence states energies given in the computed collection,[46] we had to augment the above listing in order to obtain parameters for some specific valence states. In particular, halogen valence states were restricted only by their pure s- and p-

Table 3. Calculated atomic charges for hydrocarbons

		Charge calculation scheme							
Compound	Atom	MGC	OGC	PEOE	MPA	NPA	AIM	GAPT	CHELPG
Methane	C	−0.103	−0.180	−0.0772	−0.473	−0.88	0.244	0.073	−0.333
	H	0.026	0.045	0.0193	0.118	0.22	−0.061	−0.018	0.083
Ethane	C	−0.092	−0.144	−0.0678	−0.335	−0.647	0.237	0.147	−0.010
	H	0.031	0.048	0.0226	0.112	0.216	−0.079	−0.049	0.003
Ethylene	C	−0.076	−0.120	−0.107	−0.254	−0.419	0.082	−0.041	−0.280
	H	0.038	0.060	0.0535	0.127	0.210	−0.041	0.021	0.140
Acetylene	C	−0.049	−0.086	−0.1225	−0.233	−0.244	−0.121	−0.212	−0.259
	H	0.049	0.086	0.1225	0.233	0.244	0.121	0.212	0.259
Propane	C1	−0.089	−0.141	−0.0652	−0.333	−0.646	0.225	0.126	−0.211
	H	0.032	0.048	0.0228	0.111	0.222	−0.079	−0.050	0.036
	C2	−0.083	−0.111	−0.0584	−0.216	−0.432	0.220	0.178	0.272
	H	0.035	0.051	0.0259	0.109	0.216	−0.092	−0.073	−0.051
Cyclopropane	C	−0.074	−0.089	−0.107	−0.261	−0.454	0.104	0.054	−0.228
	H	0.037	0.045	0.0535	0.131	0.227	−0.052	−0.027	0.114
Bicyclobutane	C1	−0.041	−0.018	−0.0673	−0.147	−0.251	−0.061	−0.123	−0.032
	H	0.038	0.038	0.0626	0.146	0.242	−0.009	0.008	0.098
	C2	−0.072	−0.106	−0.1353	−0.241	−0.427	0.186	0.190	−0.336
	H	0.037	0.043	0.0775	0.126	0.225	−0.055	−0.004	0.149
Tetrahedrane	C	−0.035	−0.036	−0.1068	−0.164	−0.257	−0.049	−0.105	−0.128
	H	0.035	0.036	0.1068	0.164	0.257	0.049	0.105	0.128
Correlation coefficient	MGC	—	0.983	0.880	0.985	0.984	0.698	0.220	0.667
	OGC	0.983	—	0.837	0.980	0.977	0.720	0.236	0.687
	PEOE	0.880	0.837	—	0.890	0.829	0.337	0.134	0.799
Intercept	MGC	—	−0.003	−0.007	0.003	−0.002	0.011	0.013	0.01
	OGC	−0.003	—	−0.011	0.002	−0.003	0.01	0.012	0.011
	PEOE	−0.007	−0.011	—	0.0229	−0.0518	0.0294	0.0238	0.003
Slope	MGC	—	1.475	0.613	3.755	6.575	−1.549	−0.437	2.233
	OGC	1.475	—	0.876	2.489	4.351	−1.064	−0.311	1.435
	PEOE	0.613	0.876	—	2.367	3.864	−0.520	0.186	1.863

orbitals,[46] although it was shown at the experimental[51] and theoretical[52] level that bonding orbitals of univalent halogens contain a certain fraction of s character and can vary, according to the above references, from 5 to 33%. This problem will be specially considered below.

Analogously, it is well known[42,53] that the chemical bonding in strained cyclic and cage molecules differs, sometimes dramatically, from that in unstrained species. Therefore, one should derive orbital parameters for three- and four-membered rings accounting for the specific features of their orbital hybridization.

Another problem is the lack of electronegativity values for d-orbitals of the non-transition elements belonging to the third and further rows of the Periodic Table. These orbitals, being favorably disposed towards accepting the electron pair, play an important role in the structure of coordinate and hypervalent species. Halogen higher valence compounds are also formed due to the promotion of their decoupled electrons to higher d-levels. However, the parameters for d-orbitals are difficult to obtain[42] and currently few of them are available. Santry and Segal[54] calculated the d-orbital ENs for the elements from Na to Cl; their data were taken into our parameter set. The overall set of parameters used in our calculations is listed in Table S1 of the supporting material.

Halogens

As was mentioned above, s characteristics of halogens were often taken 'from scratch' just to satisfy the fit with experimental data.[18,21,22] Recently we have suggested a new method[55] for the evaluation of halogen s character based on the analysis of the 'geometry' of valence shell electron pairs. The latter were localized due to the topology analysis[10] of the electron density Laplacian $L(\mathbf{r}) = -\nabla^2\rho(\mathbf{r})$ obtained from *ab initio* calculations. As a result, an atomic graph[10] for each halogen atom can be constructed. It could be considered as a coordination polyhedron and thus analyzed in terms of valence bond theory.[53] Halogen methyl derivatives were used as model structures. Given the hybridization characteristics of atoms, one can compute the values of particular orbital EN and hardness due to linear interpolation, since the values of s and p states are known.[46] These data for all halogens are listed in Table 2. According to Gillespie and Hargittai[50] the central atom, the bonding electron pair and the lone pair are designated as A, X and E, respectively.

Small rings

It is well known[53] that bonding orbitals in strained hydrocarbons demonstrate a different s character: endocyclic C—C bonding orbitals are less enriched by the 2s state compared with exocyclic C—H orbitals. The generally recognized reason is that the bent character of C—C bonds, e.g. in cyclopropane, causes a small interorbital angle, about 104–105°, compared with the tetrahedral one that corresponds to a low s-content of the orbital. In contrast, the interorbital angle HCH of the unstrained C—H bond is 114–115°, which brings it closer to sp^2 hybridization.

Randic[56] has devoted a series of papers to the computation of the s character of the carbon bonding orbitals in strained rings. Applying the criterion of maximum overlap[53] for the construction of the hybrids, he succeeded in explaining the geometry of a number of highly strained cage derivatives of cyclopropane and cyclobutane. They were the same hybrids that had been utilized by Guillen and Gasteiger[21d] for the additional parametrization of their original method.[21a] Making use of the Clementi orbitals[57] instead of Slater's, Klasinc *et al.*[58] achieved more accurate values that have come into textbooks.[59] These hybrids were used in our OGC model for interpolation of electronegativity and hardness values with respect to the known pure *s* and *p* states.[46] The parameters obtained from Randic's hybrid orbitals are as follows:

Bond	Hybrid	s character %	χ^0(V)	η^0(V e^{-1})
C—C in C_3H_6	$sp^{3.86}$	20.58	7.37	5.72
C—H in C_3H_6	$sp^{2.40}$	29.41	8.26	5.76
C—C in C_4H_8	$sp^{3.47}$	22.37	7.56	5.73
C—H in C_4H_8	$sp^{2.62}$	27.62	8.08	5.75

There are two indicative examples of cage hydrocarbons being structural isomers, cubane and cuneane (Fig. 3). In terms of graph theory, both of them are the regular graphs of degree three [a regular graph is one all of the vertices of which have the same degree (the number of neighbouring vertices)], thus none of the simple topological schemes can reveal three classes of charge equivalence of carbon atoms in cuneane without recourse to special parametrization.

This is the case for both the PEOE initial version[21a] and the first version of our MGC model.[41] As for the PEOE, this shortcoming was remedied by the particular analysis of the molecular graph cyclic structure[29b,60] and the use of special EN parameters based on Randic's hybrid orbitals (see above). At the same time, our OGC model allows for a natural way of using hybrid orbitals of any type. In addition, we used more comprehensive data on the hybridization in strained cycles.[58] Therefore, it would be interesting to compare the results of charge calculation produced with our OGC model, Gasteiger's PEOE model implemented in the PETRA program,[31] and those obtained using an *ab initio* method. With this aim in mind, we performed charge calculations for structures **1** and **2** by means of our OGC model, the PETRA program

Table 4. Calculated atomic charges for fluorosubstituted methanes

Compound	Atom	Charge calculation scheme							
		MGC	OGC	PEOE	MPA	NPA	AIM	GAPT	CHELPG
Methane	C	0.025	0.368	0.0795	0.087	−0.095	0.869	0.719	0.171
	F	−0.268	−0.616	−0.2527	−0.407	−0.429	−0.743	−0.550	−0.277
	H	0.081	0.083	0.0259	0.107	0.175	−0.042	−0.056	0.035
CH_2F_2	C	0.125	0.822	0.2299	0.561	0.562	1.506	1.304	0.482
	F	−0.186	−0.533	−0.213	−0.380	−0.429	−0.744	−0.576	−0.263
	H	0.124	0.122	0.098	0.100	0.149	−0.009	−0.076	0.022
CHF_3	C	0.203	1.234	0.3804	0.944	1.120	2.193	1.790	0.644
	F	−0.120	−0.465	−0.1738	−0.350	−0.421	−0.744	−0.576	−0.230
	H	0.156	0.160	0.1411	0.106	0.142	0.047	−0.061	0.047
CF_4	C	0.262	1.621	0.5617	1.313	1.620	2.951	2.202	0.814
	F	−0.065	−0.405	−0.1404	−0.328	−0.405	−0.737	−0.551	−0.203
Correlation coefficient	MGC	—	0.877	0.933	0.867	0.855	0.803	0.780	0.850
	OGC	0.877	—	0.984	0.990	0.969	0.990	0.984	0.998
	PEOE	0.933	0.984	—	0.982	0.970	0.955	0.939	0.972
Intercept	MGC	—	0.104	0.024	0.07	0.076	0.225	0.181	0.055
	OGC	0.104	—	−0.008	0.008	0.015	0.031	0.026	1.416
	PEOE	0.024	−0.008	—	0.138	0.0088	0.089	0.075	0.0167
Slope	MGC	—	3.876	1.416	2.908	3.431	6.153	4.694	1.906
	OGC	3.876	—	0.337	0.752	0.88	1.716	1.339	0.506
	PEOE	1.416	0.337	—	2.174	2.568	4.832	3.727	1.437

(via the Internet) and the *ab initio* MP2/6–31G* method[13] (all quantum chemical calculations were performed with the use of the PC-GAMESS program adapted by Dr Granovsky, Department of Chemistry, Moscow State University). The results obtained are as follows. Obviously, all cubane carbon atoms should have identical charges. For cuneane all three methods partition the atoms into three classes of charge equivalence. The PEOE method assigns the following charges: −0.105 to atoms 1 and 2, −0.082 to atoms 3, 4, 5 and 6 and −0.060 to atoms 7 and 8. In the case of the OGC model the following pattern of charge distribution is obtained: −0.046 for atoms 1 and 2, −0.026 for atoms 3, 4, 5 and 6 and −0.067 for atoms 7 and 8. Lastly, the *ab initio* MP2/6–31G* methods assigns −0.190 for atoms 1 and 2, −0.157 for atoms 3, 4, 5 and 6 and −0.201 for atoms 7 and 8. As can be seen, in spite of some difference in the absolute charge values, our OGC model demonstrates a concerted pattern of charge distribution with the non-empirical calculation, whereas the PEOE method does not.

RESULTS AND DISCUSSION

Any discourse on a new method for atomic charge calculation should begin with the evaluation of the consistency of the obtainable values with physical considerations and chemical experience. Therefore, one should primarily answer the following questions:

1. To what extent do the results produced by to the method conform with those obtained by means of other methods?
2. How is the method sensitive to variations of physical and chemical properties that reflect the specific features of the electronic charge distribution in molecules?
3. What is the applicability range of the method? In other words, which classes of compounds can be correctly treated by the method?
4. How efficient and easy to implement is the method?

Comparison with other charge calculation schemes

The atomic charges were calculated for three sets of neutral organic molecules (hydrocarbons, fluoromethanes, and oxygen-containing species) using both the MGC and OGC models. Then the results were compared with those obtained from various theoretical approaches. For the comparison we used the PEOE[21] method, Mulliken population analysis (MPA),[5] natural population analysis (NPA),[9] Bader's AIM method,[10] GAPT[12b] and CHELPG.[11b] The last set of data obtained at the Hartree–Fock level of theory with the 6–31G** basis set[13] on the 6–31G* geometry were taken from the comprehensive analysis by Wiberg and Rablen.[2] The data on PEOE charges were obtained with the help of the PETRA program[31] (via the Internet). The values for hydrocarbons, fluoromethanes and oxygen-containing compounds are listed in Tables 3, 4 and 5, respectively.

As can be seen, the charge distributions obtained are in

Table 5. Calculated atomic charges for oxygen-containing organic molecules

		Charge calculation scheme							
Compound	Atom	MGC	OGC	PEOE	MPA	NPA	AIM	GAPT	CHELPG
Formaldehyde	C	0.036	0.334	0.1055	0.245	0.325	1.365	0.800	0.496
	O	−0.209	−0.534	−0.3041	−0.432	−0.577	−1.296	−0.675	−0.478
	H	0.086	0.100	0.0993	0.093	0.126	−0.034	−0.063	−0.009
Acetaldehyde	C	0.038	0.350	0.1156	0.379	0.513	1.292	0.853	0.620
	O	−0.217	−0.531	−0.3008	−0.475	−0.606	−1.328	−0.740	−0.539
	H	0.091	0.101	0.1028	0.091	0.130	−0.051	−0.089	−0.048
	C	−0.050	−0.079	−0.0084	−0.435	−0.759	0.183	−0.033	−0.256
	H	0.046	0.053	0.0303	0.160	0.245	−0.020	0.003	0.088
Acetone	C	0.039	0.365	0.1256	0.501	0.685	1.219	0.864	0.761
	O	−0.220	−0.529	−0.2975	−0.516	−0.627	−1.349	−0.783	−0.589
	C	−0.051	−0.077	−0.0057	−0.421	−0.748	0.181	−0.038	−0.373
	H	0.047	0.053	0.0305	0.159	0.248	−0.021	−0.001	0.104
Formamide	C	0.064	0.525	0.1837	0.562	0.670	1.977	1.304	0.790
	O	−0.189	−0.451	−0.2828	−0.562	−0.702	−1.392	−0.904	−0.619
	H	0.100	0.116	0.1216	0.097	0.129	−0.035	−0.056	−0.023
	N	−0.152	−0.512	−0.2761	−0.732	−0.944	−1.476	−0.854	−0.987
	H	0.088	0.161	0.1269	0.322	0.427	0.471	0.252	0.435
Correlation coefficient	MGC	—	0.898	0.968	0.848	0.816	0.770	0.739	0.769
	OGC	0.898	—	0.966	0.932	0.889	0.965	0.954	0.940
	PEOE	0.968	0.966	—	0.882	0.8325	0.893	0.862	0.859
Intercept	MGC	—	−0.017	0.014	0.02	0.015	0.16	0.098	0.052
	OGC	−0.017	—	−0.009	−0.021	−0.04	0.077	0.49	0.0087
	PEOE	0.014	−0.009	—	−0.0057	−0.02	0.115	0.07	0.026
Slope	MGC	—	0.308	1.46	2.9	3.8	6.82	4.029	3.33
	OGC	0.308	—	0.502	1.097	1.418	2.93	1.786	1.396
	PEOE	1.46	0.502	—	2.0	2.56	5.22	3.106	2.46

harmony with a chemist's intuition. The data in Tables 3–5 also show good agreement among the obtained charges and those resulting from the above-mentioned *ab initio* methods. The best results were obtained for the OG charge model. In almost all cases the PEOE[21] method demonstrated worse correlations. This is especially the case for the alkane dataset (Table 3). In the case of fluoromethanes and oxygen-containing compounds (Tables 4 and 5), the correlations of PEOE charges are slightly better than MG charges and worse than the OG charges.

For hydrocarbons the OG model charges correlate well with the MPA and NPA charges. The poor correlation with AIM and GAPT charges is most likely related to the inverse polarity of the C^{+}—H^{-} bond, contrary to electronegativity considerations. The excellent correlations obtained for fluorosubstituted methanes (Table 4) suggest a concerted simulation of the inductive effect for both *ab initio* methods and the method developed. Oxygen-containing species listed in Table 5 can be considered as fragments occurring in biomolecules, and therefore a consistent pattern of the charge distribution produced by all methods would be desirable. A similar situation takes place, as judged from the data in Table 5.

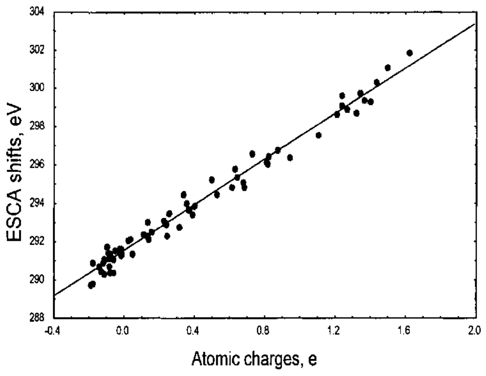


Figure 4. Correlation plot of ESCA experimental shifts against atomic charges calculated with the OGC model

Of special note is the good quality of the correlation between our OG model charges and electrostatic potential (ESP) derived CHELPG charges[11b] because the ESP charges turn out to be crucial in the development of the force fields intended for application to the modeling of biomolecules.[61] Yet another advantage of the OG charge model is a good correlation with Bader's AIM charges,[10] since the latter were recognized to be the preferential ones for a number of reasons.[2]

Correlation with molecular properties

One of the often-used means for estimating atomic charges is the correlation with core electron binding energies.[21a] In the framework of the classical potential

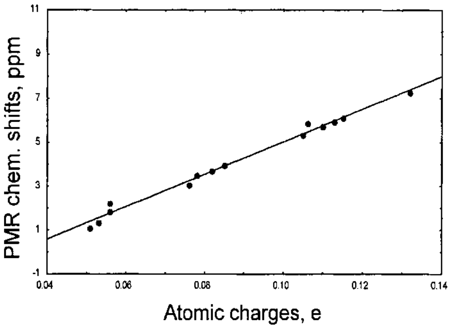


Figure 5. Correlation plot of ^{1}H NMR chemical shifts of chloro-substituted alkanes (Ref. 21b) against the OGC atomic charges

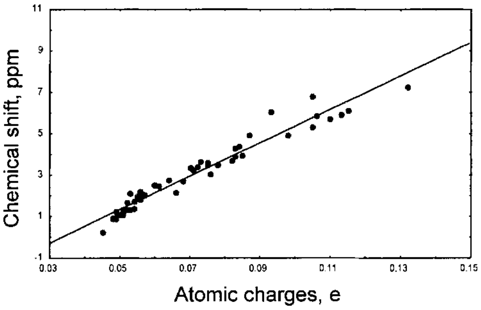


Figure 7. Correlation plot of ^{1}H NMR chemical shifts against the OGC atomic charges, full dataset (Ref. 21b)

model[62] it is believed that there exists an almost linear charge dependence of core electron binding energies that are determined as ESCA shifts. The validity of a number of charge calculation schemes[19,21] was tested on ESCA shifts, especially C-1s shifts. The simplified potential model substantiated by Gasteiger and Marsili[21a] implies a direct correlation of ESCA shifts with the atomic charges. Following the common approach we compiled a database including 70 experimental ESCA C-1s shifts (57 molecules),[63] and put them into correspondence with the atomic charges computed with both developed models. Also, for the sake of comparison, a similar correlation was carried out with the PEOE atomic charges computed via the Internet.[31] The results obtained showed the good ability of our charges to reproduce C-1s electron binding energies. Charges produced from the orbital graph model demonstrate a better correlation ($r = 0.991$, $s = 0.44$ eV, $F = 3562$, intercept = 291.58, slope = 5.905) than those from the simpler molecular graph model ($r = 0.977$, $s = 0.68$ eV, $F = 1433$). The correlation parameters for PEOE charges are almost the same as those for MG charges ($r = 0.979$, $s = 0.65$ eV, $F = 1512$, intercept = 291.82, slope = 15.667) The OG model charges are plotted against C-1s core binding energies in Fig. 4 (more details are available Table S2 in the supporting material). The only more or less significant deviations within the OGC model can be found for oxygen- and especially carbonyl-containing molecules. This may be caused by an insufficient consistency of the orbital EN scale. In the case of the MGC model the largest deviations are observed for multiple-bonded molecules because Pauling's EN scale assigns the same values for the atom in any valence state. Another example of this type is molecules of high symmetry (e.g. CX_4, $X = H$, Hal), which is related to the peculiarities of the EN equilibration procedure being applied to molecular graphs. The OG charge model is free from similar shortcomings.

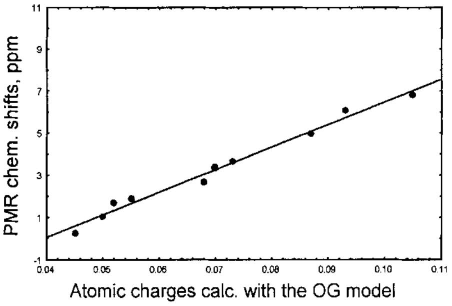


Figure 6. Correlation plot of ^{1}H NMR chemical shifts of bromo-substituted alkanes (Ref. 21b) against the OGC atomic charges

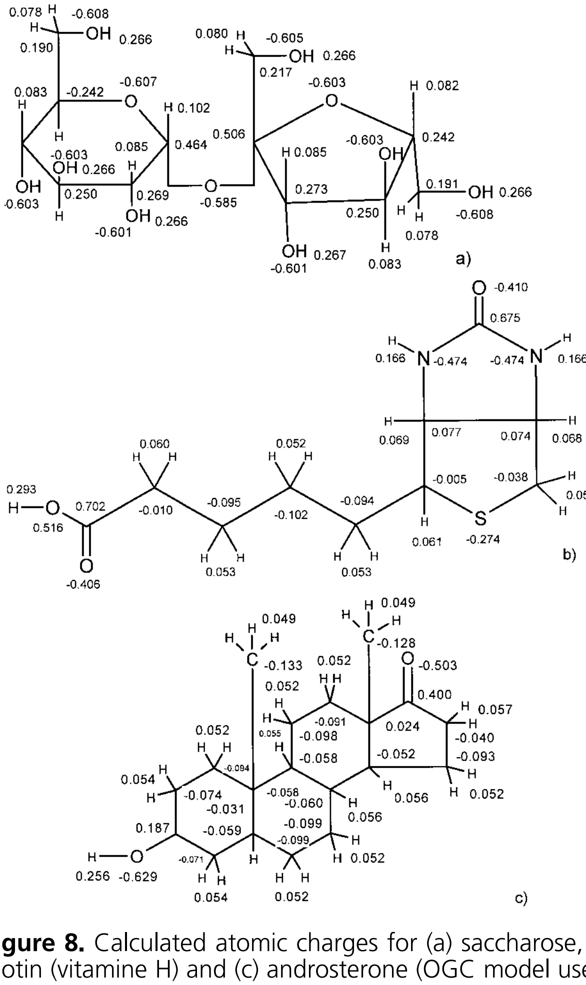


Figure 8. Calculated atomic charges for (a) saccharose, (b) biotin (vitamine H) and (c) androsterone (OGC model used)

Another way to validate the charges obtained is to correlate them with ^{1}H NMR chemical shifts.[21b] One can agree that NMR chemical shifts do not show any direct and unambiguous correlation with atomic charges, because the chemical shifts are influenced not only by the electrons at the resonating nucleus, but also by remote effects through several bonds as well as magnetic anisotropy of the environment (especially in conjugated systems). Nonetheless, we have revealed the above-mentioned charge dependence of ^{1}H NMR shifts. The dataset we used in this study was previously compiled by Gasteiger and Marsili.[21b] In particular, for the subsets of 15 chloro- and 11 bromoalkanes excellent correlations of ^{1}H NMR chemical shifts with calculated hydrogen atomic charges were obtained. Within the OGC model the statistical characteristics are $r = 0.994$, $s = 0.22$ ppm for chloro- and $r = 0.994$, $s = 0.26$ ppm for bromoalkanes. Similar results were demonstrated with the use of the MGC model. The correlation diagrams where the OG model charges are plotted against the ^{1}H NMR chemical shifts for chloro- and bromoalkanes are presented in Figs 5 and 6, respectively. In addition, we treated a more extended dataset including 55 proton chemical shifts. Fairly good correlations were produced with the MG model ($r = 0.967$, $s = 0.45$ ppm, $F = 758$) and, particularly, with the OGC model ($r = 0.975$, $s = 0.398$ ppm, $F = 1001$). The best plot related to the OGC model is presented in Fig. 7; tabulated data can be found in Table S3 of the supporting material. In spite of the excellent particular correlations (chloro- and bromoalkanes) one should note that in the case of the total dataset the PEOE charges demonstrated better results.[21a]

Generally, such good results can be attributed to adequacy of the developed models. This is especially the case with the OGC model, because it is more flexible with respect to both molecular representation and parametrization.

Applicability range

The MGC model is limited to neutral covalent molecules; better results are obtained for low-polarity organic molecules (hydrocarbons). By contrast, the OGC model can be applied to atomic σ-charge calculation for many types of chemicals. Organic, inorganic, coordinate compounds, charge-transfer complexes, ions and radicals will be treated in a uniform way. The only precondition is the availability of necessary parameters. We must specially emphasize that a method that permits one to treat conjugated and aromatic systems is in progress and will published in due course.

To present more illustrations, a set of fairly complex molecules is shown in Fig. 8 (all of them are of high biological importance) with the charge distribution obtained by means of our OGC model. One can clearly see that the charges obtained are realistic and look like those produced by more complicated and time-consuming methods.

Algorithm complexity and implementation

The algorithm that underlies the method suggested differs favorably in its ease of implementation and low computational complexity. The most time-consuming operation is the matrix inversion; its computational complexity is $O(N^3)$, where N is the number of graph vertices. We must emphasize that no iteration steps are needed and self-consistency is achieved in a single pass of the algorithm.

Both graph charge models (MGC and OGC) are implemented as Windows 95/98/NT programs written in C++ programming language.

CONCLUSIONS

The main result of this study is the development of a novel approach (realized on two graph models, MGC and OGC models) to the calculation of partial atomic charges based on an original principle of electronegativity equilibration. The core feature of the approach is the electronegativity equilibration that is assumed to occur in the molecule by analogy with the equilibration of electrical potentials in the nodes of an electrical network. The first, simpler model based on the molecular graph (MG) formalism makes use of only one parameter, Pauling's electronegativity. The second model uses the notion of the orbital graph (OG) that appears to be more adequate in the modeling of the molecular charge distribution. Electronegativity and hardness of the corresponding atomic orbitals are used as the main parameters of the method.

Second, we have elaborated ways to evaluate all parameters needed for the correct treatment of heteroatoms (univalent halogens, in particular), small rings, and other situations that usually bring specific difficulties in other empirical approaches.[18–23]

A set of correlations of obtained charges with spectral data and *ab initio* charges convincingly demonstrate the quality of the developed model and the meaningfulness of the obtained values. The obvious advantages of the proposed method are the wide applicability range due to the flexibility of the model and parametrization, high speed of computation and suitability for calculations of large data sets. These features are of great importance for application in methods of molecular modeling using electrostatic considerations such as computer-aided molecular design, QSAR studies, etc.

Finally, the methods suggested can be improved. The obvious progress is connected with a proper adequate treatment of conjugated species and an explicit account for molecular geometry, and both areas are now being actively pursued.

Acknowledgements

This work was supported by the Russian Foundation for Basic Research, Project No 98-03-32955a. The authors are undebted to Ms Polina Krylenko for her valuable assistance in the preparation of the manuscript.

REFERENCES

1. Bachrach SM. In *Reviews in Computational Chemistry*, Lipkowitz KB, Boyd DB (eds). VCH: New York, 1993; vol. **5**, 171–227.
2. Wiberg K, Rablen P. *J. Comput. Chem.* 1993; **14**: 1504–1518.
3. (a) Cramer C, Truhlar D. *Theor. Chim. Acta* 1997; **98**: 206–211; (b) Li J, Zhu T, Cramer CJ, Truhlar DG. *J. Phys. Chem. A* 1998; **102**: 1820–1831.
4. Naray-Szabo G, Ferenczy G. *Chem. Rev.* 1995; **95**: 829–847.
5. Mulliken RS. *J. Chem. Phys.* 1955; **23**: 1833–1840.
6. Lowdin PO. *J. Chem. Phys.* 1953; **21**: 374.
7. Edgecombe KE, Boyd RJ. *J. Chem. Soc.* 1987; 1307–1315.
8. Fliszar S. *Lecture Notes in Chemistry*. Springer: Berlin, 1994; vol. **63**.
9. Reed A, Weinstock R, Weinhold F. *J. Chem. Phys.* 1985; **83**: 735–746.
10. (a) Bader RFW. *Atoms in Molecules: a Quantum Theory*. Clarendon Press: Oxford, 1990; (b) Bader RFW, Johnson S, Tang TH, Popelier P. *J. Phys. Chem. A* 1996; **100**: 15398–15415.
11. (a) Chirlian LE, Francl MM. *J. Comput. Chem.* 1987; **8**: 894–905; (b) Breneman CM, Wiberg KB. *J. Comput. Chem.* 1990; **11**: 361–373.
12. (a) King WT, Mast GB, Blanchette PP. *J. Chem. Phys.* 1972; **56**: 4440–4446; (b) Cioslowski J. *J. Am. Chem. Soc.* 1989; **111**: 8333–8336..
13. Hehre WJ, Radom L, Schleyer v. RP, Pople JA. *Ab initio Molecular Orbital Theory*. Wiley: New York, 1986.
14. Politzer P, Mulliken RS. *J. Chem. Phys.* 1971; **55**: 5135–5136.
15. Mortier WJ, Van Genechten K, Gasteiger J. *J. Am. Chem. Soc.* 1985; **107**: 829–835.
16. (a) Gimarc BM, Ott JJ. *J. Am. Chem. Soc.* 1983; **108**: 4298–4303; (b) Bonchev D, Kier L. *J. Math. Chem.* 1992; **9**: 75–85; (c) Galvez J, Garcia R, Salabert MT, Soler R. *J. Chem. Inf. Comput. Sci.* 1994; **34**: 520–525.
17. Sanderson RT. *J. Am. Chem. Soc.* 1983; **105**: 2259–2261.
18. Huheey JE. *J. Phys. Chem.* 1965; **69**: 3284–3291.
19. Jolly WL, Perry WB. *Inorg. Chem.* 1974; **13**: 2686–2692
20. Baumer L, Sala G, Sello G. *Tetrahedron Comput. Method* 1989; **2**: 37–46.
21. (a) Gasteiger J, Marsili M. *Tetrahedron* 1980; **36**: 3219–3228; (b) Gasteiger J, Marsili M. *Org. Magn. Reson.* 1981; **15**: 353–360; (c) Hutchings MG, Gasteiger J. *Tetrahedron Lett.* 1983; **24**: 2541–2544; (d) Guillen MD, Gasteiger J. *Tetrahedron* 1983; **39**: 1331–1335; (e) Marsili M, Gasteiger J. *Croat. Chem. Acta* 1980; **53**: 601–614.
22. Mullay J. *J. Am. Chem. Soc.* 1986; **108**: 1770–1775.
23. (a) Kirpichenok MA, Zefirov NS. *Russ. J. Org. Chem.* 1987; **23**: 673–690, 690–701; (b) Zefirov NS, Kirpichenok MA, Izmailov FF, Trofimov MI. *Dokl. Akad. Nauk SSSR* 1987; **296**: 883–887.
24. (a) Smith R, Ree T, Magee J, Eyring H. *J. Am. Chem. Soc.* 1951; **73**: 2263–2268; (b) Abraham R, Griffiths L, Loftus P. *J. Comput. Chem.* 1982; **3**: 407–416; (c) Fetisov VI, Kolesova VA, Kotov AV, Lisyanskii II, Chimishkyan AL, Martynov IV. *Dokl. Akad. Nauk* 1998; **358**: 210–214.
25. Del Re G. *J. Chem. Soc.* 1958; 4031–4040.
26. Mortier WJ, Ghosh SK, Shankar S. *J. Am. Chem. Soc.* 1986; **108**: 4315–4320.
27. Pauling L. *The Nature of the Chemical Bond.* Cornell University Press: Ithaca NY, 1960.
28. Balaban AT. *J. Mol. Struct. (THEOCHEM)* 1985; **21**: 117–142.
29. (a) Katrizky AR, Sild S, Karelson M. *J. Chem. Inf. Comput. Sci.* 1998; **38**: 1171–1176; (b) Trofimov MI, Smolenskii EA. *Russ. Chem. Bull.* 2000; 401–406.
30. (a) Hinze J, Jaffe H. *J. Am. Chem. Soc.* 1962; **84**: 540–546; (b) Hinze J, Whitehead M, Jaffe H. *J. Am. Chem. Soc.* 1963; **85**: 148–154.
31. *PETRA*. http://www2.ccc.uni-erlangen.de/software/petra/index.html
32. SYBYL 6.6.1, Tripos: St Louis, MO, 1999.
33. Palyulin VA, Radchenko EV, Zefirov NS. *J. Chem. Inf. Comput. Sci.* 2000; **40**: 659–667.
34. Parr RG, Yang W. *Density Functional Theory of Atoms and Molecules*. Oxford University Press: New York and Clarendon Press: Oxford, 1989.
35. (a) Nalewajski RF. *J. Am. Chem. Soc.* 1984; **106**: 944–945; (b) De Proft F, Langenaeker W, Geerlings P. *J. Mol. Struct. (THEOCHEM)* 1995; **339**: 45–55.
36. Burkert U, Allinger NL. *Molecular Mechanics*. American Chemical Society: Washington, DC, 1982.
37. (a) Théry V, Rinaldi D, Rivail J-L, Maigret B, Ferenczy JJ. *J. Comput. Chem.* 1994; **15**: 269–282.
38. Hobza P, Zahradnik R. *Intermolecular Complexes*. Elsevier: Amsterdam, 1988.
39. (a) Seshu S, Read MB. *Linear Graphs and Electrical Networks*. Addison-Wesley: Reading, MA, 1961; (b) *The Manual on Electrical Engineeing*. Energoatomizdat: Moscow, 1985.
40. Golender VE, Drboglav VV, Rosenblit AB. *J. Chem. Inf. Comput. Sci.* 1981; **21**: 196–204.
41. Oliferenko AA, Palyulin VA, Zefirov NS. *Dokl. Chem.* 1999; **368**: 209–212.
42. Huheey JE. *Inorganic Chemistry. Principles of Structure and Reactivity*. Harper and Row: New York, 1983.
43. Yang W, Lee G, Ghosh SK. *J. Phys. Chem.* 1985; **89**: 5412–5414.
44. (a) Iczkowski RP, Margrave IL. *J. Am. Chem. Soc.* 1961; **83**: 3547–3551; (b) Jorgenssen CK, Horner SM, Harfield WE, Tyree SY. *Int. J. Quantum Chem.* 1967; **1**: 191–195.
45. Komorowski L. *Chem. Phys.* 1987; **114**: 55–71.
46. (a) Bergman D, Hinze J. *Struct. Bonding (Behin)* 1987; **66**: 145–190; (b) Bergman D, Hinze J. *Angew. Chem. Int. Ed. Engl.* 1996; **35**: 150–163.
47. Slater J. *Self-Consistent Field for Molecules and Solids*. McGraw-Hill: New York, 1974.
48. Kohn W, Becke AD, Parr RG. *J. Chem. Phys.* 1996; **100**: 12974–12980.
49. Luo Y-R, Benson SW. *Acc. Chem. Res.* 1992; **25** 375–381.
50. Gillespie R, Hargittai I. *The VSEPR Model of Molecular Geometry*. Allyn and Bacon: Boston, 1991.
51. (a) Wilmshurst JK. *J. Chem. Phys.* 1959; **30**: 561–565; (b) Whitehead MA, Jaffe HH. *Theor. Chim. Acta* 1963; **1**: 209–221.
52. (a) Mulliken RS. *J. Chem. Phys.* 1952; **51**: 295–310; (b) Rue R, Ruedenberg K. *J. Phys. Chem.* 1964; **68**: 1676–1706; (c) Szentpàly LV. *J. Phys. Chem. A* 1998; **102**: 10912–10915.
53. Coulson C. *Valence*. Oxford University Press: London, 1961.
54. Santry DP, Segal GA. *J. Chem. Phys.* 1967; **47**: 158–174.
55. Oliferenko AA, Pisarev SA, Palyulin VA, Zefirov NS. *Bull. Moscow State Univ., Ser. 2, Chem.* 2000; **41**: 324–326.
56. (a) Randic M, Maksic Z. *Theor. Chim. Acta* 1965; **3**: 59–68; (b) Maksić Z, Klasinc L, Randic M. *Theor. Chim. Acta* 1966; **4**: 273–286.
57. (a) Clementi E, Roetti C. *At. Data Nucl. Data Tables* 1974; **14**: 177–478.
58. Klasinc L, Maksic Z, Randic M. *J. Chem. Soc. A* 1966; 755–757.
59. Minkin VI, Simkin BYa, Minyaev RN. *The Theory of Molecular Structure*. Vysshaya Shkola: Moscow, 1979.
60. Gasteiger J, Jochum C. *J Chem. Inf. Comput. Sci.* 1979; **19**: 43–48.
61. Cieplak P, Cornell WD, Bayly C, Kollman PA. *J. Comput. Chem.* 1995; **16**: 1357–1377.
62. Siegbahn K, Nordling C, Johansson G, Hedman J, Heden PF, Hamrin K, Gelius U, Bergmark T, Werne LO, Manne R, Baer Y. *ESCA Applied to Free Molecules*. North-Holland: Amsterdam, 1969.
63. Jolly WL, Bomben KD, Eyermann CY. *At. Data Nucl. Data Tables* 1984; **31**: 433–493.

APPENDIX

A set of equations ((A1, A2)) was obtained and to solve these equations, we should take some approximations.

$$\begin{cases} \frac{D_{\mathrm{M}}}{r_j} q_i^2 + \eta_i^{\mathrm{o}} q_i - \Delta\chi_i = 0 \qquad i = 1, \ldots, N & \text{(A1)} \\ \sum_i q_i = 0 & \text{(A2)} \end{cases}$$

Expressions (A3) are the roots of each of Eqns (A1):

$$q_{i(1,2)} = \frac{-\eta_i^{\mathrm{o}} r_j \pm \sqrt{(\eta_i^{\mathrm{o}} r_j)^2 + 4D_{\mathrm{M}} r_j \Delta\chi_i}}{2D_{\mathrm{M}}} \tag{A3}$$

provided that

$$|D_{\mathrm{M}}| < \frac{(\eta_i^{\mathrm{o}})^2 r_j}{4|\Delta\chi_i|}$$

to ensure positivity of discriminant. These expressions can be rewritten as follows:

$$q_{i(1,2)} = \frac{\eta_i^{\mathrm{o}} r_j \left[-1 \pm \sqrt{1 + \frac{4\Delta\chi_i}{(\eta_i^{\mathrm{o}})^2 r_j} D_{\mathrm{M}}} \right]}{2D_{\mathrm{M}}} \tag{A4}$$

Assuming that

$$|D_{\mathrm{M}}| \ll \frac{(\eta_i^{\mathrm{o}})^2 r_j}{4|\Delta\chi_i|}$$

we use Taylor's expansion (A5):

$$\sqrt{1+\alpha} \approx 1 + \frac{\alpha}{2} - \frac{\alpha^2}{8} \tag{A5}$$

for expressions (A4) and thus obtain

$$q_{i(1)} = \frac{\Delta\chi_i}{\eta_i^{\mathrm{o}}} - \frac{(\Delta\chi_i)^2}{(\eta_i^{\mathrm{o}})^3 r_j} D_{\mathrm{M}} \tag{A6}$$

$$q_{i(2)} = -\frac{\eta_i^{\mathrm{o}} r_j}{D_{\mathrm{M}}} - \frac{\Delta\chi_i}{\eta_i^{\mathrm{o}}} + \frac{(\Delta\chi_i)^2}{(\eta_i^{\mathrm{o}})^3 r_j} D_{\mathrm{M}} \tag{A7}$$

Root (A7) should be discarded as it reaches infinity when D_{M} approaches zero (constant hardness approximation). Substituting expressions (A6) in Eqn. (A2), we obtain the value of D_{M} [Eqn. (A8)]:

$$D_{\mathrm{M}} = \frac{\sum_{i=1}^{N} \sum_{j=1}^{A} \frac{\Delta\chi_{ij}}{\eta_{ij}^{\mathrm{o}}}}{\sum_{i=1}^{N} \sum_{j=1}^{A} \frac{(\Delta\chi_{ij})^2}{(\eta_{ij}^{\mathrm{o}})^3 r_j}} \tag{A8}$$

Finally, using the obtained value of D_{M} we can rewrite q_i values as

$$q_i = \frac{\Delta\chi_i}{\eta_i^{\mathrm{o}}} - \frac{(\Delta\chi_i)^2 D_{\mathrm{M}}}{(\eta_i^{\mathrm{o}})^3 r_j} \tag{A9}$$