Full Research Paper

Electronegativity Equalization Method: Parameterization and Validation for Large Sets of Organic, Organohalogene and Organometal Molecule

Radka Svobodová Vařeková [1], Zuzana Jiroušková [1], Jakub Vaněk [1], Šimon Suchomel [1] and Jaroslav Koča [1],*

1 National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Kotlářská 2, 611 37 Brno, Czech Republic; Tel.: +420 549 494 947, Fax: +420 549 492 556,
E-mail: jkoca@chemi.muni.cz

* Author to whom correspondence should be addressed; E-mail: jkoca@chemi.muni.cz

Abstract: The Electronegativity Equalization Method (EEM) is a fast approach for charge calculation. A challenging part of the EEM is the parameterization, which is performed using *ab initio* charges obtained for a set of molecules. The goal of our work was to perform the EEM parameterization for selected sets of organic, organohalogen and organometal molecules. We have performed the most robust parameterization published so far. The EEM parameterization was based on 12 training sets selected from a database of predicted 3D structures (NCI DIS) and from a database of crystallographic structures (CSD). Each set contained from 2000 to 6000 molecules. We have shown that the number of molecules in the training set is very important for quality of the parameters. We have improved EEM parameters (STO-3G MPA charges) for elements that were already parameterized, specifically: C, O, N, H, S, F and Cl. The new parameters provide more accurate charges than those published previously. We have also developed new parameters for elements that were not parameterized yet, specifically for Br, I, Fe and Zn. We have also performed crossover validation of all obtained parameters using all training sets that included relevant elements and confirmed that calculated parameters provide accurate charges.

Keywords: Charge distribution, Electronegativity Equalization Method, Parameterization, Organohalogenes, Organometals.

1. Introduction

Electronegativity Equalization Method (EEM) [1,2,3] is a fast approach for charge calculation. The basic idea is based on the density functional theory (DFT) [4,5]. First, Parr et al. applied the DFT and formulated a new definition and explanation of electronegativity [6,7]. Later on, Mortier et al. applied Parr's definition of electronegativity and Sanderson's Electronegativity Equalization Principle (EEP) [8,9,10] and created the EEM.

This method is able to calculate atomic charges markedly faster than common *ab initio* approaches. The *ab initio* charge calculations exhibit time complexity of $O(B^4)$, where B is greater or equal to the number of valence electrons. The EEM approach shows a time complexity of $\theta(N^3)$, where N is the number of atoms. Nevertheless, accuracy of the EEM corresponds to the *ab initio* methods.

A challenging part of the EEM is the parameterization that is performed using *ab initio* charges obtained for a set of molecules. The parameterization is very time-consuming with time complexity of $O(S.B^4)$, where S is a number of molecules in the set. The most common parameterization of the EEM is a parameterization for the HF method with the STO-3G basis set, where the charges are calculated by Mulliken population analysis (MPA) [11,12]. Principally, it is also possible to parameterize the EEM for other basis sets (i.e., 6-31G*) and methods for charge calculation (i.e., CHELPG, MK, NPA, ESP, Hirshfeld method) [13,14]. First attempts to calculate EEM parameters were published in eighties [1,2]. These publications contained only parameters for C, H, N and O, which were developed using training sets of about one hundred molecules. Further parameterizations were performed during the nineties and contained parameters for new elements (S, Si, P, F, Cl) and more complex bases [15,16,17]. The EEM parameterization still remains attractive to chemists' attention [18,19,20].

The goal of this work is to perform the EEM parameterization based on large sets of organic, organohalogen and organometal molecules (containing Zn and Fe) selected from databases NCI DIS [21] and CSD [22], and to validate the quality of calculated parameters on reference sets of molecules selected from these databases. The parameterization was performed for STO-3G MPA charges.

2. Theoretical basis

2.1. EEM

Using DFT, the effective (charge-dependent) electronegativity of the atom i in a molecule can be calculated by eq. (1) [1,2 3]:

$$\chi_i = A_i + B_i.q_i + \kappa.\sum_{j=1(j\neq i)}^{N} \frac{q_j}{R_{i,j}} \tag{1}$$

where N is the number of atoms in the molecule, q_i and q_j are the charges distributed on the atoms i and j, respectively, $R_{i,j}$ is the distance between atoms i and j, and κ is the adjusting factor. The coefficients A_i and B_i are defined by eqs. (2):

$$A_i = \chi_i^* = \chi_i^0 + \Delta\chi_i$$
$$B_i = 2\eta_i^* = 2\left(\eta_i^0 + \Delta\eta_i\right) \tag{2}$$

where χ_i^0 is the electronegativity of an isolated neutral atom *i*, η_i^0 is the hardness, and $\Delta\chi_i^0$ and $\Delta\eta_i$ describe the molecular environment. The coefficients A_i, B_i and κ are empirical parameters, which must be obtained via EEM parameterization. Such a parameterization is a topic of this work.

According to Sanderson's Electronegativity Equalization Principle [8, 9, 10], the effective electronegativity of each atom in the molecule is equal to the molecular electronegativity $\overline{\chi}$:

$$\chi_1 = \chi_2 = \ldots = \chi_N = \overline{\chi} \tag{3}$$

The total charge *Q* of the molecule is equal to the sum of all the atomic charges:

$$\sum_{i=1}^{N} q_i = Q \tag{4}$$

The atomic charges are described using the equation system (5), which contains $N+1$ equations with $N+1$ unknowns: $q_1, q_2, \ldots, q_N$ and $\overline{\chi}$. This system was derived from equations (1), (3) and (4) [1]:

$$\begin{pmatrix} B_1 & \frac{\kappa}{R_{1,2}} & \cdots & \frac{\kappa}{R_{1,N}} & -1 \\ \frac{\kappa}{R_{2,1}} & B_2 & \cdots & \frac{\kappa}{R_{2,N}} & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\kappa}{R_{N,1}} & \frac{\kappa}{R_{N,2}} & \cdots & B_N & -1 \\ 1 & 1 & \cdots & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_N \\ \overline{\chi} \end{pmatrix} = \begin{pmatrix} -A_1 \\ -A_2 \\ \vdots \\ -A_N \\ Q \end{pmatrix} \tag{5}$$

The matrix of the equation system (5) is called EEM matrix.

2.2. EEM Parameterization

Empirical parameters A_i, B_i and κ (described by eqs. (1) and (2)) can be calculated in the following way [17]:

From eq. (1) and (3), eq. (6) can be derived:

$$\overline{\chi} = A_i + B_i q_i + \kappa \sum_{j=1(j\neq i)}^{N} \frac{q_j}{R_{i,j}} \tag{6}$$

Eq. (6) can be rewritten as:

$$A_i + B_i q_i = \overline{\chi} - \kappa \sum_{j=1(j\neq i)}^{N} \frac{q_j}{R_{i,j}} \tag{7}$$

Meaning that eq. (7) is in the form:

$$A_i + B_i x_i = y_i \tag{8}$$

where:

$$x_i = q_i, \qquad y_i = \overline{\chi} - \kappa \sum_{j=1(j\neq i)}^{N} \frac{q_j}{R_{i,j}}$$

Then, empirical parameters can be obtained using eq. (8) in the following way:

1. Selection of a set of molecules used for the EEM parameterization.
2. *Ab initio* calculation of atomic charges q_i for all atoms within all selected molecules.
3. Calculation of the molecular electronegativity $\overline{\chi}$ as a harmonic average of atomic electronegativities χ_i^0 (for isolated atoms *i*):

$$\overline{\chi} = N\left(\sum_{i=1}^{N} \frac{1}{\chi_i^0}\right)^{-1} \tag{9}$$

4. Selection of κ values for which the parameterization will be performed.

5. For each of the above selected κ:
   - Calculation of x_i and y_i values for all atoms in all molecules using eq. (8).
   - Separation of x_i and y_i couples into subsets according to the chemical symbol and hybridization of the atom i (for example C in sp^3, C in sp^2 etc.).
   - Calculation of parameters A_i and B_i for each of these subsets using the least square minimization.
6. Finding the optimal κ value.

3. Methods

In this work, two databases were used. The first one was the NCI DIS 3D database [21], created as a part of DTP NCI (Developmental Therapeutics Program of National Cancer Institute). This database contains organic molecules tested against cancer, specifically their topologies and also geometries, predicted by the program CHEM-X [23] and stored in SDF format [24]. The second database used was CSD (Cambridge Structural Database) [22], administered by CCDC (Cambridge Crystallographic Data Centre). Geometries of molecules are stored also in SDF format. However, in this case information is obtained experimentally using the X-ray and/or neutron diffraction. Both these databases are sufficiently large, containing more than two hundred thousand molecules.

Table 1: Sets of molecules that were used as training and testing sets for the EEM parameterization.

Database	Denotation of the set	Number of molecules	Atoms included	Position of the set in the database
NCI DIS (predicted data)	n_{beg}	2000	C, O, N, H, S	beginning (ID between 1 and 3162)
	n_{mid}	2000	C, O, N, H, S	middle (ID between 300 000 and 314 026)
	n_{end}	2000	C, O, N, H, S	end (ID between 705 000 and 712 703)
	n_{all}	6000	C, O, N, H, S	n_{beg}, n_{mid} and n_{end}
	n_{hal}	4000	C, O, N, H, S, Br, Cl, F, I	beginning (ID between 106498 and 114688)
CSD (crystallo-graphic data)	c_{beg}	2000	C, O, N, H, S	beginning (ID starting by A and B)
	c_{mid}	2000	C, O, N, H, S	middle (ID starting by J, K and L)
	c_{end}	2000	C, O, N, H, S	end (ID starting by W and Y)
	c_{all}	6000	C, O, N, H, S	c_{beg}, c_{mid} and c_{end}
	c_{hal}	4000	C, O, N, H, S, F, Cl, Br, I	beginning (ID starting by A, B and C)
	c_{met}	2000	C, O, N, H, S, Fe, Zn	beginning (ID starting by A and B)
	$c_{h,m}$	6000	C, O, N, H, S, F, Cl, Br, I, Fe, Zn	c_{hal} and c_{met}

ID is a unique identification of a molecule in a database. In the NSC DIS database, ID is a number between 1 and about 720 000. Database CSD uses alphabetically sorted string IDs that contains six upper case characters.

From these two databases, several training sets of molecules were selected (see Table 1). Our goal was to generate training sets, which cover most of bonding situations and also conformational variability in real molecules. For that reason, we have chosen large sets containing randomly selected molecules. As molecules are unsorted in NCI DIS and CSD databases, the simplest random selection is to take a continuous part of the database. We selected three training sets, containing elements C, H, O, N and S from each database. To obtain the most versatile training sets, we selected first training set from the beginning, second from the middle and the third from the end of the databases. We have also used unions of these sets. For organohalogenes and organometals, we did not need so many training sets as the process of parameterization was already debugged on above mentioned six training sets and their unions. Therefore we used only one training set of organohalogenes from each database. Just one training set (from CSD database) was used for organometals, because the NCI DIS database does not contain enough organometal molecules. These organometal and organohalogene molecules were selected from the beginning of the databases.

For the parameterization, *ab initio* charges were calculated using the HF method with the STO-3G basis set for all molecules in all sets. The charge calculation was performed by Gaussian 98 program [25]. After that, the EEM parameterization was performed using calculated *ab initio* charges for all training sets.

Table 2: Quality of parameters that were obtained by EEM parameterization using all training sets.

R_{mol}^{avg}		Training set												
		Lit.	n_{beg}	n_{mid}	n_{end}	n_{all}	n_{hal}	c_{beg}	c_{mid}	c_{end}	c_{all}	c_{hal}	c_{met}	$c_{h,m}$
T	n_{beg}	0.966	0.955	**0.962**	0.930	0.959	0.961	0.950	0.924	0.938	0.945	0.944	0.928	0.938
e	n_{mid}	0.957	0.939	**0.951**	0.910	0.947	**0.951**	0.941	0.902	0.932	0.936	0.930	0.918	0.929
s	n_{end}	0.960	0.944	**0.958**	0.922	0.944	0.956	0.956	0.894	0.942	0.945	0.932	0.929	0.942
t	n_{all}	0.961	0.946	**0.957**	0.921	0.953	0.956	0.949	0.907	0.937	0.942	0.935	0.925	0.936
e	n_{hal}	-	-	-	-	-	**0.928**	-	-	-	-	0.919	-	0.887
d	c_{beg}	0.945	0.918	0.928	0.870	0.928	0.930	**0.946**	0.917	0.934	0.941	0.936	0.916	0.937
	c_{mid}	0.934	0.912	0.922	0.867	0.921	0.920	**0.932**	0.902	0.921	0.928	0.922	0.898	0.921
s	c_{end}	0.936	0.913	0.925	0.870	0.922	0.922	**0.936**	0.902	0.923	0.930	0.927	0.903	0.927
e	c_{all}	0.939	0.914	0.925	0.869	0.924	0.924	**0.938**	0.907	0.926	0.933	0.928	0.906	0.928
t	c_{hal}	-	-	-	-	-	0.903	-	-	-	-	**0.910**	-	0.885
	c_{met}	-	-	-	-	-	-	-	-	-	-	-	**0.887**	0.879
	$c_{h,m}$	-	-	-	-	-	-	-	-	-	-	-	-	**0.885**

R_{mol}^{avg} describes quality of parameters obtained via EEM parameterization using the training set. This value is between 0 and 1. The closer it is to 1, the more accurate charges are provided employing the EEM method using the parameters. The R_{mol}^{avg} is an average of R_{mol} values for all molecules in the training set. R_{mol} is the R-squared value of the linear regression line, which was inserted into a set of points [q_i(*ab initio*), q_i(EEM)], where q_i(*ab initio*) and q_i(EEM) are *ab initio* and EEM charges (calculated using the parameters) of atom *i*, respectively. Lit. means parameters obtained from literature [17]. For our parameters, the best R_{mol}^{avg} for each tested set is bolded.

Parameters, calculated for all training sets presented in Table 1, and also parameters obtained from the literature were validated for all training sets that contained suitable atoms. Validation of parameters for a selected training set was done in such a way that *ab initio* charges and the EEM charges calculated for each molecule from the training set using the developed parameters were compared via the least square method. In other words, the linear regression line was fitted to a set of points [q_i(*ab initio*), q_i(EEM)], where q_i(*ab initio*) and q_i(EEM) are *ab initio* and EEM charges of the atom *i*, respectively. Correlation between *ab initio* and EEM charges was described by the R-squared value [26] of this line. This R-squared value is between 0 and 1. The closer it is to 1, the better the correlation is. The R-squared value (R_{mol}) was calculated for each molecule in the training set. An average value of (R_{mol}^{avg}) was calculated from all R_{mol} values in each set to express the quality of parameters for the set.

Table 3: Information about numbers of molecules and atoms in newly created training sets c_{beg2}, c_{hal2}, c_{met2} and $c_{h,m2}$. For more details see the text.

Element	Bond order	Number of molecules and atoms in training set							
		c_{beg2}		c_{hal2}		c_{met2}		$c_{h,m2}$	
		molecules	atoms	molecules	atoms	molecules	atoms	molecules	atoms
H	1	530	11187	810	13214	1112	25894	3082	60873
C	1	498	4113	729	5128	1070	10918	2847	24359
N	1	325	605	378	689	641	1353	1598	3195
O	1	400	1162	536	1258	830	2636	2185	6030
S	1	58	116	87	160	168	416	358	756
C	2	518	5871	843	10058	1078	12756	3086	37612
N	2	172	350	289	561	374	825	1062	2163
O	2	401	907	546	991	786	1943	2123	4449
Cl	1	-	-	455	1158	-	-	929	2319
Br	1	-	-	211	324	-	-	477	735
F	1	-	-	188	805	-	-	411	1745
I	1	-	-	57	95	-	-	134	202
Zn	1	-	-	-	-	103	178	155	268
Fe	1	-	-	-	-	186	317	203	335
Total		544	24311	870	34441	1154	57236	3258	145041

4. Results

Table 4: EEM parameters A, B and κ (see eqs. (1) and (2)) obtained via parameterization using training sets c_{beg2}, c_{hal2}, c_{met2} and $c_{h,m2}$.

		EEM parameters created using training sets							
		c_{beg2}		c_{hal2}		c_{met2}		$c_{h,m2}$	
		κ 0.44		κ 0.66		κ 0.42		κ 0.55	
Element	Bond order	A	B	A	B	A	B	A	B
H	1	2.396	0.959	2.404	1.461	2.386	0.937	2.394	1.212
C	1	2.459	0.611	2.503	0.899	2.452	0.593	2.476	0.772
N	1	2.597	0.790	2.653	1.017	2.550	0.663	2.597	0.835
O	1	2.625	0.858	2.713	1.211	2.624	0.847	2.676	1.077
S	1	2.407	0.491	2.465	0.705	2.424	0.400	2.440	0.665
C	2	2.464	0.565	2.516	0.850	2.462	0.527	2.495	0.704
N	2	2.554	0.611	2.633	0.869	2.547	0.639	2.600	0.790
O	2	2.580	0.691	2.757	1.348	2.567	0.622	2.622	0.850
Cl	1	-	-	2.791	2.365	-	-	2.759	2.092
Br	1	-	-	2.496	1.345	-	-	2.494	1.315
F	1	-	-	2.789	1.494	-	-	3.032	2.985
I	1	-	-	2.421	2.309	-	-	2.454	1.387
Zn	1	-	-	-	-	2.378	0.259	2.422	0.301
Fe	1	-	-	-	-	2.557	0.061	2.575	0.087

For each training set of molecules in Table 1, the parameters were found. As it was described in the Methods section, calculated parameters were validated for all training sets that contained suitable atoms and also compared with the parameters from literature [17]. As the literature does not show the κ value (see eq. (1)), we had to find the κ value via our methodology. The best fit for κ was found to equal 1.25. The results of this parameter quality validation expressed by R_{mol}^{avg} are summarized in Table 2. This table shows that the quality of parameters varies for different training sets. Moreover, the quality of parameters from literature is generally slightly better than the quality of our parameters. Therefore, our effort was to further improve our methodology and parameters. The main idea of this improvement was based on results, obtained for training sets n_{all} and its subsets n_{beg}, n_{mid} and n_{end} and also for training set c_{all} with subsets c_{beg}, c_{mid} and c_{end}. It is seen from Table 2 that R_{mol}^{avg} (n_{all}) is better than the average value from R_{mol}^{avg} (n_{beg}), R_{mol}^{avg} (n_{mid}) and R_{mol}^{avg} (n_{end}), but the best results are obtained for the set n_{mid}. Analogically, in the training set c_{all}, the subset c_{beg} provides the best parameters. Randomly sorted molecules that create the training sets imply the good accuracy of parameters from subsets n_{mid} and c_{beg}. Therefore, the quality of parameters can be increased by selection of an appropriate subset of the input training set. We have tested two methods of appropriate subset selection:

    1. Select only molecules, which have R_{mol} greater than a defined limit (for example 0.8).

2. Sort molecules from the training set *T* randomly and create a sequence of them (1, 2, ..., |*T*|), where |*T*| is a cardinality of *T*. Calculate parameters for all subsets ST_i, where ST_i is obtained from *T* by removing the subset DST_i. The subset DST_i is composed of elements $T_{(i-1).K+1}$, $T_{(i-1).K+2}$, …, $T_{i.K}$; where *K* can be, for example, 100. Now create the selection in the following way: From the input training set, sorted into the above described sequence, delete every subset DST_i, for which $R_{mol}^{avg}(T) < R_{mol}^{avg}(ST_i)$.

By comparison, the second approach was found to be more successful. It is interesting that sets selected via the first method provide worse quality of parameters than the input training sets themselves (results not shown here).

Using method 2, we have performed selections based on sets c_{beg}, c_{hal}, c_{met} and $c_{h,m}$ and created sets c_{beg2}, c_{hal2}, c_{met2} and $c_{h,m2}$ (see Table 3).

Table 5: Comparison of the quality of parameters obtained using original sets and their selected subsets.

R_{mol}^{avg}		Training set								
		Lit.	c_{beg}	c_{beg2}	c_{hal}	c_{hal2}	c_{met}	c_{met2}	$c_{h,m}$	$c_{h,m2}$
T e s t e d s e t	n_{beg}	0.966	0.950	***0.968***	0.944	0.958	0.928	0.959	0.938	0.950
	n_{mid}	0.957	0.941	***0.962***	0.930	0.952	0.918	0.951	0.929	0.943
	n_{end}	0.960	0.956	***0.970***	0.932	0.953	0.929	0.956	0.942	0.949
	n_{all}	0.961	0.949	***0.967***	0.935	0.954	0.925	0.955	0.936	0.947
	n_{hal}	-	-	-	*0.919*	**0.940**	-	-	*0.887*	*0.927*
	c_{beg}	0.945	*0.946*	***0.960***	0.936	*0.954*	0.916	*0.954*	0.937	*0.947*
	c_{mid}	0.934	0.932	***0.948***	0.922	*0.943*	0.898	*0.941*	0.921	*0.934*
	c_{end}	0.936	*0.936*	***0.951***	0.927	*0.945*	0.903	*0.944*	0.927	*0.937*
	c_{all}	0.939	0.938	***0.953***	0.928	*0.947*	0.906	*0.946*	0.928	*0.939*
	c_{hal}	-	-	-	*0.910*	**0.934**	-	-	*0.885*	*0.921*
	c_{met}	-	-	-	-	-	*0.887*	**0.927**	*0.879*	*0.917*
	$c_{h,m}$	-	-	-	-	-	-	-	*0.885*	**0.919**

For more details about R_{mol}^{avg} see Table 2. R_{mol}^{avg} values of our parameters that are better than the literature parameters (denoted as Lit., taken from reference [17]) are in italics. The best R_{mol}^{avg} value (our parameters) for each tested set is bolded.

We have chosen the CSD database as this database contains high quality experimental data. The set c_{beg} was selected as it exhibits R_{mol}^{avg} higher than c_{mid}, c_{end} and c_{all}. The parameters were calculated for selected subsets c_{beg2}, c_{hal2}, c_{met2} and $c_{h,m2}$ (see Table 4). Then the parameters were validated for all training sets containing corresponding atoms (see Table 5 and graphs in supplementary materials).

It is seen that the selected subsets c_{beg2}, c_{hal2}, c_{met2} and $c_{h,m2}$ provide markedly better parameters than the input sets c_{beg}, c_{hal}, c_{met} and $c_{h,m}$ themselves. In all cases we have found parameters that are better than the literature parameters.

The parameters c_{beg2} are of better quality than parameters obtained from literature [17] for both used databases. The parameters c_{hal2}, c_{met2} and $c_{h,m2}$ are of a worse quality than published parameters for the NCI DIS database, but are better for the experimental database CSD. Moreover, these parameters contain new data for halogens or Fe and Zn.

Generally, we can conclude, that it is possible to calculate parameters using both the predicted and experimental databases. However, parameters that are based on experimental structures exhibit better charge calculation results. It can be caused by the fact that the theoretical structures from NCI DIS database may include some less realistic geometries compared to the experimental structures from CSD database. These parameters are more useful as they are portable and can be used for an arbitrary molecule that contains atoms for which the parameters were developed. Our results also show that it is useful to work with large training sets and select the best subset that provides the highest quality parameters. It is also reasonable to test several training sets.

We did a large validation of our parameters. For demonstration, tables with detailed results of EEM charge calculation method with our parameters for several different organohalogene and organometal molecules are attached in supplementary material. Also coordinates and charges on single atoms are available there.

5. Conclusions

In this work, we have improved the published EEM parameters to calculate the STO-3G MPA charges for C, O, N, H, S, F and Cl. The new parameters provide more accurate charges than those published previously [17]. We have developed parameters for elements not yet parameterized, specifically for Br, I, Fe and Zn.

The EEM parameterization we have performed has been based on 12 training sets, which are also the largest published training sets used for the EEM parameterization ranging from 2000 to 6000 molecules. We have shown that the number of molecules in the training set is very important for the quality of the parameters.

We have performed crossover validation of all obtained parameters using all training sets that include relevant elements. To the best of our knowledge, we have performed the most accurate testing of EEM parameters quality published so far.

This is the first work to compare EEM parameters calculated using two principally different training sets, one being a database of theoretically predicted 3D structures (NCI DIS) and the second being a database of crystallographic structures (CSD). Our results show that it is possible to use both databases, but parameters from the CSD database training sets give more accurate charges. Moreover, the parameters obtained from the NCI DIS database training sets are not very suitable to calculate charges for molecules from the CSD database.

These improved and newly developed parameters can be used for charge calculation using the program EEM SOLVER [27], which we have developed and which is freely available via the internet on http://ncbr.chemi.muni.cz/~n19n/eem_abeem.

Acknowledgements

References and Notes

1. Mortier, W.J.; Van Genechten, K.; Gasteiger, J. Electronegativity Equalization: Application and Parametrization. *J. Am. Chem. Soc.* **1985**, *107*, 829-835.
2. Mortier, W.J.; Ghosh, S.K.; Shankar, S. Electronegativity Equalization Method for the Calculation of Atomic Charges in Molecules. *J. Am. Chem. Soc.* **1986**, *108*, 4315-4320.
3. Van Genechten, K.A.; Mortier, W.J.; Greelings, P. Intrinsic Framework Electronegativity: A Novel Concept in Solid State Chemistry. *J. Chem. Phys.* **1987**, *86*, 5063-5071.
4. Parr, R.G.; Yang, W. *Density Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1989.
5. Bartolotti, L.J.; Flurchick, K. An Introduction to Density Functional Theory. *Rev. Comp. Chem.* **1996**, *7*, 187-216.
6. Parr, R.G.; Donnelly, R.A.; Levy, R.A.; Palke, W.E.J Electronegativity: The Density Functional Viewpoint. *Chem. Phys.* **1978**, *68*, 3801-3807.
7. Donnelly, R.A.; Parr, R.G. Elementary Properties of an Energy Functional of the First-order Reduced Density Matrix. *J. Chem. Phys.* **1978**, *69*, 4431-4439.
8. Sanderson, R.T. *Chemical Bond and Bond Energies*; Academic Press: New York, 1976.
9. Sanderson, R.T. *Polar Covalence*; Academic Press: New York, 1983.
10. Parr, R.G.; Pearson, R.G. Absolute Hardness: Companion Parameter to Absolute Electronegativity. *J. Am. Chem. Soc.* **1983**, *105*, 7512-7522.
11. Grant, G.H.; Richards, W.G. *Computational Chemistry*; Oxford University Press: USA, 1995.
12. Bachrach, S.M. Population Analysis and Electron Densities from Quantum Mechanics. *Rev. Comp. Chem.* **1994**, *5*, 171-227.
13. Bultinck, P.; Langenaeker, W.; Lahorte, P.; DeProft, F.; Geerlings, P.; Waroquier, M.; Tollenaere, J.P. The Electronegativity Equalization Method I: Parametrization and Validation for Atomic Charge Calculations. *J. Phys. Chem. A* **2002**, *106*, 7887-7894.
14. Bultinck, P.; Langenaeker, W.; Lahorte, P.; DeProft, F.; Geerlings, P.; Van Alsenoy, C.; Tollenaere, J.P. The Electronegativity Equalization Method II: Applicability of Different Atomic Charge Schemes. *J. Phys. Chem. A* **2002**, *106*, 7895-7901.
15. Yang, Z.-Z.; Shen, E.-Z.; Wang, L.-H. A scheme for Calculating Atomic Charge Distribution in Large Molecules Based on Density Functional Theory and Electronegativity Equalization. *J. Mol. Struct. (Theochem)* **1994**, *312*, 167-173.
16. Yang, Z.-Z.; Shen, E.-Z. Molecular Electornegativity in Density Functional Theory (I). *Sci. China B* **1995**, *38*, 521-528.

17. Yang, Z.-Z.; Shen E.-Z. Molecular Electronegativity in Density Functional Theory (II). *Sci. China B* **1996**, *39*, 20-28.

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

19. Zhang, Q.; Cagin, T.; van Duin, A.; Goddard III, W.A. Adhesion and Nonwetting-Wetting Transition in the $Al/\alpha - Al_2O_3$ Interface. *Phys. Rev. B* **2004**, *69*, 1-11.

20. Bollmann, L.; Hillhouse, H.W.; Delgass W.N. Calibration of an Electronegativity Equalization Method Based on Dft Results to Evaluate the Partial Charges, Global Softness, and Local Softness of Zeolite Structures. *Comp. Mol. Sci. Eng. Forum* **2005**, *21*, 597d.

21. Milne, G.W.A.; Nicklaus, M.C.; Driscoll, J.S.; Wang, S.; Zaharevitz. D. National Cancer Institute Drug Information System 3D Database. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1219-1224.

22. Allen, F.H.; Kennard, O. Cambridge Structural Database. *Chem. Des. Autom. News* **1993**, *8*, 31-37.

23. Eaton, P.E., Millikan, R.; Engel, P. *The CHEMX Reference Manual*; Chemical Design Ltd: Oxford, 1990.

24. Dalby, A.; Nourse, J.G.; Hounshell, W.D.; Gushurst, A.K.I.; Grier, D.L.; Leland, B.A.; Laufer, J. Description of Several Chemical Structure File Formats Used by Computer Programs Developed at Molecular Design Limited. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 244-255.

25. Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.E.; Robb, M.A.; Cheeseman, J.R.; Zakrzewski, V.G.; Montgomery Jr., J.A.; Stratmann, R.E.; Burant, J.C.; Dapprich, S.; Millam, J.M.; Daniels, A.D.; Kudin, K.N.; Strain, M.C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G.A.; Ayala, P.Y.; Cui, Q.; Morokuma, K.; Salvador, P.; Dannenberg, J.J.; Malick, D.K.; Rabuck, A.D.; Raghavachari, K.; Foresman, J.B.; Cioslowski, J.; Ortiz, J.V.; Baboul, A.G.; Stefanov, B.B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R.L.; Fox, D.J.; Keith, T.; Al-Laham, M.A.; Peng, C.Y.; Nanayakkara, A.; Challacombe, M.; Gill, P.M.W.; Johnson, B.; Chen, W.; Wong, M.W.; Andres, J.L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E.S.; Pople, J.A. *Gaussian 98*; Gaussian Inc.: Pittsburgh, 2001.

26. Nemhauser, G.L.; Rinnooy, A.H.G. *Optimization*; North-Holland: Amsterdam, 1989.

27. Svobodová Vařeková, R.; Koča, J. Optimized and Parallelized Implementation of the Electronegativity Equalization Method and the Atom-Bond Electronegativity Equalization Method. *J. Comput. Chem.* **2005**, *27*, 396-405.