sciendo

Atomic partial charge model in chemistry: chemical accuracy of theoretical approaches for diatomic molecules

Matej Uhliar

Institute of Physical Chemistry and Chemical Physics, Slovak University of Technology in Bratislava, Radlinského 9, SK-812 37 Bratislava, Slovakia
matej.uhliar@stuba.sk

Abstract: Atomic partial charges cannot be physically measured but they play a significant role in many chemical theories and theoretical models. Therefore, they are, evaluated from experimentally acquired properties or calculated by quantum chemistry computational methods. This study is focused on determining chemical accuracy of various theoretical methods of computing atomic partial charges based on quantum chemistry. Values of gas-phase atomic partial charges were acquired by Mulliken (MUL) population analysis, natural bond analysis (NBO), Merz-Singh-Kollman (MSK) scheme, and atomic polar tensor (APT) charges computed considering Density Functional Theory and *ab initio* Møller-Plesset up to the second order levels. Correlations between the calculated values were determined by principal component analysis (PCA) and further confirmed by linear regression. The best agreement between experimentally evaluated atomic partial charges and theoretical values was obtained with the MSK scheme.

Keywords: electric dipole moment; atomic charges; Mulliken charges; principal component analysis

Introduction

Physical properties of compounds and their chemical reactivity are affected by their molecular as well as electronic structure. In case of aromatic organic molecules, their molecular structure can be described using various aromaticity descriptors, especially popular are HOMA-based Harmonic Oscillator Model of Electron Delocalization (HOMED) (Raczyńska et al., 2010) and Harmonic Oscillator Model of Aromaticity for Heterocyclic Electron Delocalization (HOMHED) (Frizzo et al., 2012) which are based on theoretical or experimental X-ray, bond lengths, respectively. These indices seem to be suitable for aromaticity description of linear oligomers and their ring-fused analogues (Cagardová et al, 2020; Cagardová et al., 2022) as well as for large two-dimensional molecules, e.g., sunflower molecules (Cagardová et al., 2019; Cagardová et al., 2022). Bond length of relevant substituent groups also correlates very well with antioxidant activity (Klein et al., 2006), chemical reactivity (Klein et al., 2007), or biological activity (Štellerová et al., 2022). Description of the electronic structure is based on the knowledge of electron density distribution in the molecule or its changes during optical transitions, which opens a possibility to investigate several properties, i.e., atomic partial charges, total charge density, spin density for radicals, energies of selected frontier molecular orbitals, and energies of electronic transitions. Although, partial electric charges of atoms cannot be measured directly, they have a significant role in theoretical chemistry, i.e., atomic partial charges serve to estimate the reactivity of molecules (Klein et al., 2006; Mehmood et al., 2017), can help to determine structural differences between molecules and their conformers (Rostkowski et al., 2007), and be implemented for the prediction of pKa (Gross et al., 2002; Svobodová Vařeková et al., 2013; Ugur et al., 2014). Additionally, there are reports on good correlation of selected partial charges with toxicity of para phenols (Michalík et al., 2018a) and 3-hydroxynaphthalene-2-carboxanilides (Michalík et al., 2018b).

The physical concept of partial atomic charge is not uniquely defined. Classical electrostatics defines the partial electric charge of a diatomic polar molecule using the electric dipole moment (p), which is a measure of the separation of positive and negative electrical charges within a system. Two-point charges located on the atoms, one with positive ($+q$) and the other one with negative ($-q$) values, are separated by distance d. Absolute value of these charges is identical. Since the electric dipole moment and the interatomic distance are experimentally obtainable quantities, absolute value of the partial charge on the atom is determined as

$$|q| = p/d \tag{1}$$

In quantum chemistry, partial atomic charge can be evaluated from the electron wave function or electron density. From the historical point of view, the oldest method for atomic partial charge

determination is the Mulliken population analysis based on the molecular orbital scheme (Mulliken, 1955) based on the Hartree-Fock (HF) approach. Obtained Mulliken (MUL) partial charges are, in principle, sensitive to the basis set choice. Hypothetically, a complete basis set for a molecule can be obtained by locating a large number of atomic orbitals on a single atom, i.e., all electrons in the system would be assigned to that single atom. And thus, the method has no complete basis set limit as the exact value of MUL partial charge depends on the way the limit is approached. Such dependency results in incorrectly defined calculated charges. Additionally, the basis set convergence of charges does not exist and different basis set families may yield significantly different results (Thakkar, 2021). These problems can be overcome by newer methods for net atomic charge computing, e.g., electrostatic potential Merz-Singh-Kollman analysis (MSK) (Breneman et al., 1990) or natural bond orbital (NBO) analysis (Reed et al., 1985). Both these methods can be performed at the HF level, but also in calculations where correlation energy is considered, i.e., post HF approach or DFT.

In the MSK scheme suggested by U.C. Singh and P.A. Kollman (Singh et al., 1984), partial atomic charges are fit to reproduce molecular electrostatic potential (MEP) at a number of points around the molecule. Thus, MEP is calculated at gridpoints located on several layers around the molecule. Layers are constructed as an overlay of van der Waals spheres around each atom discarding all points located inside the van der Waals volume. The best results are achieved by sampling points not too close to the van der Waals surface and the van der Waals radii are therefore modified using scaling factors.

NBO analysis is based on the Lewis's structure picture, where the input atomic orbital basis set is transformed via natural atomic orbitals (NAOs) and natural hybrid orbitals (NHOs) into natural bond orbitals (NBOs). The obtained NBOs correspond to the widely used Lewis's picture, in which two-centre bonds and lone pairs are localized.

Among the population analysis schemes available in literature, atomic partial charges evaluated from generalized atomic polar tensor (GAPT) can be completely determined from experimental data, infrared intensities and frequencies, equilibrium dipole moments, and molecular geometries (Biarge et al., 1961; Person et al., 1974).

Although many theoretical works deal with the comparison of partial atomic charges calculated using various treatments of atomic molecules, systematic analysis statistically comparing chemical accuracy of standardly used population analysis has not been performed yet. Therefore, a comparative study of theoretical ab initio and DFT calculations of diatomic polar molecules is presented in this paper. Partial aims of this study are to calculate optimal geometries and atomic charges for 17 representative molecules. Chemical accuracy of obtained results will be compared with experimental values deduced from experimental gas-phase dipole moments and distances. Finally, the Principal Component Analysis (PCA) (Jolliffe, 2002) will be used to determine possible correlations between the applied treatments.

Calculation details

Quantum chemical computations were carried out using the Gaussian 16 program package (Frisch et al., 2016) using 6-31G, 6-31G* and 6-311++G(d,p) basis set (Hariharan et al., 1973). The calculations were performed at the ab initio MP2 (Møller-Plesset up to the second order) level where all electrons were correlated. Next, the density functional theory with hybrid meta exchange-correlation functional M0-62X (Zhao et al., 2008) and 3-parameter hybrid B3LYP functional (Becke, 1993; Lee et al., 1988; Vosko et al., 1980; Stephens et al., 1994) was used. B3LYP and M0-62X functionals are the most commonly used in calculations of inorganic and organic molecules. Additionally, all electrons were corelated in case of the MP2 approach.

Results and Discussions

Experimental gas-phase values for interatomic distances and dipole moments of studied diatomic molecules (Irikura, 2007; Huber et al., 2013; Nelson et al., 1967; Johnson III, 2022) were used to calculate the estimated experimental absolute partial charge (Exp) of the molecule; minimal value of 0.020e was found for CO, while maximal value of 0.878e was found for the LiF molecule. Physical constant e stands for elementary charge in this notation. The MUL, APT, NBO and MSK values of atomic partial charge were evaluated at DFT (B3LYP and M06-2X functionals) and MP2 levels. MUL charges were calculated at HF/MP2/6-311++G(d, p) level of theory in MP2 calculations as single point calculation throughout MP2, B3LYP and M06-X, using the 6-31G* basis set. MUL* values were computed in the same way as MUL' but applying the 6-31G basis set, as it is implied that a smaller basis set is more suitable for the calculations of electric properties of molecules.

The highest Exp value was found for the NaF molecule to be 0.878, whereas the lowest Exp value was 0.020 for the CO molecule. Experimentally acquired interatomic distances of these molecules are 1.93 Å and 1.13 Å, respectively. In case of the

MP2 approach, the highest values of atomic partial charges calculated are mostly for the NaF molecule obtained with MUL (0.838), MUL' (0.703), NBO (0.988), and MSK (0.874). The highest value of atomic partial charges determined using MUL* was for the NaCl (0.710), and with APT for the NO (1.194) molecule. On the other hand, the lowest values were computed for other molecules than CO: NO with MUL (0.025), CS with MUL* (0.019), NBO (0.069), and APT (0.031), BrCl with MUL', whereas MSK values showed lowest value of 0.000 for the CO molecule.

In case of the DFT-B3LYP approach, the highest value was found for NaF with MUL (0.757), NBO (0.978), and MSK (0.874), NaCl with MUL' (0.564) and MUL* (0.622), and for LiF with APT (0.860). The lowest values were obtained for CO with Mul (0.010), Mul* (0.013), and MSK (0.000), for BF with MUL' (0.097), and for BrCl with NBO (0.102) and with APT (0.103).

The highest values of atomic partial charges computed using the DFT-M0-62X method were found for the same molecules as with the DFT-B3LYP approach: for NaF at MUL 0.780, at MUL' 0.610, at NBO 0.978, and at MSK 0.874. A similar value of 0.618 obtained with MUL* was computed for the NaCl molecule and the value of 0.860 for the LiF molecule was found with the APT approach. The lowest values of atomic partial charges were computed for the CO molecule with the MUL (0.018) and MSK (0.000) approach. Secondly, the lowest value of 0.018 was calculated with MUL* for the BF molecule, while the BrCl molecule has the lowest values of atomic partial charges acquired with MUL' (0.029), NBO (0.095), and APT (0.103).

Calculated interatomic distances show similar values with mostly insignificant differences compared to experimentally obtained values of interatomic distances. For example, experimentally determined interatomic distance of the smallest HF molecule is 0.92 Å, and calculated interatomic distances of this molecule are 0.916 Å, 0.922 Å, and 0.918 Å at MP2, DFT-B3LYP and DFT-M0-62X levels of theory, respectively. Experimental value of interatomic distance for the largest molecule NaBr is equal to 2.50 Å. Yet, calculations at all considered levels of theory provide values of 2.528 Å at MP2, 2.526 Å at DFT-B3LYP, and 2.511 Å at DFT-M0-62X levels of theory. The largest differences in interatomic distances in a molecule were observed for the BrCl molecule, where experimental value is 2.32 Å while calculated values of interatomic distances are 2.164 Å, 2.192 Å, and 2.158 Å at MP2, DFT-B3LYP and M0-62X levels of theory, respectively.

Mutual correlation between individual values of atomic partial charges at various levels of theory was determined via PCA using the first (PC1) and second (PC2) principal components of each analysis since their coverage over the whole dataset of relations is 95.89 % for the B3LYP, 95.98 % for the M06-2X, and 92.37 % for the MP2 dataset. Chemical accuracy is then further estimated using linear regression between Exp and individual values. Exp values, in atomic unit, were calculated according to:

$$Exp = \left(\frac{p}{d}\right) / e \qquad (2)$$

Where Exp is the calculated value of atomic partial charge determined from experimentally obtained dipole moment p converted from Debeye to C·m, and experimentally determined interatomic distance d, and then divided by elementary charge e.

Tab. 1. Experimental gas-phase electric dipole moments (p), interatomic distances (d) taken from (Nelson et al., 1967; Johnson III, 2022), evaluated experimental absolute values of partial charges (atomic units) and Mulliken charges obtained from experimental geometry calculations.

Molecule	*p*/*D*	*d*/Å	Exp
HF	1.83	0.92	0.416
HCl	1.09	1.27	0.178
HBr	0.83	1.41	0.122
ClF	0.88	1.63	0.113
BrF	1.42	1.76	0.168
BrCl	0.52	2.32	0.047
NaF	8.12	1.93	0.878
NaCl	8.97	2.36	0.791
NaBr	9.09	2.50	0.756
LiF	6.28	1.56	0.836
LiCl	7.09	2.02	0.730
LiBr	7.23	2.17	0.693
BF	0.50	1.27	0.082
CO	0.11	1.13	0.020
SO	1.55	1.50	0.215
CS	1.96	1.53	0.266
NO	0.16	1.15	0.029

Out of all studied molecules, the most polar one is NaBr with experimentally obtained value of induced dipole moment of 9.09 *D* and the least polar molecule is CO with the dipole moment of 0.11 *D* as shown in Table 1. These two molecules have interatomic distances of 2.50 Å and 1.13 Å, respectively. Experimentally obtained data result in partial atomic charges of 0.756 for NaBr and 0.020 for CO.

Theoretical gas-phase interatomic distances and absolute values of Mulliken partial atomic charges are shown in Table 2. As it was mentioned before, individual theoretical values of interatomic distances slightly deviate from those calculated from experimental data. However, when Mulliken values of atomic partial charges are compared to Exp values, rather high deviations are observed. Yet, values of partial atomic charges found from theoretically optimized geometries show only small deviations from values obtained from single point calculations in experimental geometries.

Therefore, values of atomic partial charges are suggested to be mostly influenced by theoretical means of calculation, while they show little to no dependency on the geometry used; values of atomic partial charges obtained by various methods on selected levels of theory are shown in Table 3.

In comparison to the NaBr molecule (experimentally most polar) with its Exp value of atomic partial charge of 0.756, the MP2 level of theory provides the closest value of 0.762 with the MSK approach, while the highest deviation from the Exp values was observed for the NBO (0.924), followed by the MUL* approach (0.579). In case of the CO molecule, the smallest deviation from the Exp value (0.020) was determined again with the MSK approach (0.000), even though this value suggests that the CO molecule is not polar. The highest deviations were observed for the NBO (0.594) followed by the MUL' approach (0.307). Complete results of atomic partial charges with bond lengths of individual diatomic molecules obtained at the MP2 level of theory are shown in Table 3a.

Individual values of atomic partial charges vary in contrast to experimentally acquirable data. Even their differences depend on the studied molecule. This behaviour can be observed also for the B3LYP and M0-62X levels of theory. Compared to MP2 and DFT levels, MSK values of atomic partial charges show the least change, to the point they remain the same in most of the studied molecules. Results of the DFT-B3LYP level of theory for the NaBr molecule show the same value of 0.762 of atomic partial charge for the MSK approach, even though other approaches provide different values than those obtained at the MP2 level of theory. The highest difference in value from Exp was acquired with the MUL* approach (0.469), followed by the MUL' value (0.488) and the MUL value (0.566). In comparison to MP2 results in general, individual values of various approaches, aside from MSK values, changed together with their deviations from each Exp value. The most notable difference was observed for the radical NO molecule for APT partial charge at the MP2 level of computation; this behavior was not observed at the DFT level of theory.

Values obtained at the DFT-M0-62X level of theory show similar trends as those obtained at the DFT-B3LYP level of theory. Thus, MSK values show the smallest deviation from Exp values in most cases, while APT values also show rather small diffe-

Tab. 2. Comparison of gas-phase Mulliken values of partial atomic charges (atomic units) calculated using experimental (Exp) and optimized geometries at various levels of theory.

Molecule	Exp		B3LYP		M0-62X		MP2	
	d/Å	MUL	*d*/Å	MUL	*d*/Å	MUL	*d*/Å	MUL
HF	0.92	0.290	0.922	0.291	0.918	0.296	0.916	0.302
HCl	1.27	0.133	1.287	0.136	1.282	0.145	1.273	0.132
HBr	1.41	0.107	1.427	0.110	1.420	0.123	1.410	0.099
ClF	1.63	0.141	1.678	0.154	1.646	0.150	1.672	0.171
BrF	1.76	0.221	1.811	0.234	1.782	0.214	1.807	0.261
BrCl	2.32	0.011	2.192	0.025	2.158	0.035	2.164	0.031
NaF	1.93	0.757	1.954	0.757	1.936	0.780	1.972	0.838
NaCl	2.36	0.537	2.383	0.545	2.366	0.563	2.376	0.631
NaBr	2.50	0.558	2.526	0.566	2.511	0.585	2.528	0.648
LiF	1.56	0.619	1.583	0.627	1.576	0.647	1.595	0.702
LiCl	2.02	0.400	2.025	0.402	2.021	0.415	2.016	0.463
LiBr	2.17	0.423	2.179	0.426	2.179	0.441	2.179	0.484
BF	1.27	0.233	1.271	0.232	1.268	0.237	1.272	0.150
CO	1.13	0.011	1.128	0.010	1.122	0.018	1.139	0.105
SO	1.50	0.220	1.510	0.222	1.490	0.222	1.537	0.256
CS	1.53	0.248	1.539	0.241	1.528	0.196	1.539	0.049
NO	1.15	0.090	1.148	0.091	1.138	0.097	1.134	0.025

Tab. 3a. Comparison of gas-phase absolute values of partial atomic charges (atomic units) calculated using selected approaches at MP2 level of theory.

Molecule	MUL	MUL'	MUL*	*NBO*	*APT*	*MSK*
MP2						
HF	0.302	0.517	0.370	0.555	0.423	0.474
HCl	0.132	0.243	0.259	0.243	0.212	0.258
HBr	0.099	0.256	0.213	0.168	0.142	0.195
ClF	0.171	0.359	0.351	0.375	0.279	0.150
BrF	0.261	0.393	0.471	0.461	0.333	0.192
BrCl	0.031	0.008	0.149	0.115	0.084	0.045
NaF	0.838	0.703	0.646	0.988	0.888	0.874
NaCl	0.631	0.667	0.710	0.958	0.844	0.791
NaBr	0.648	0.590	0.579	0.946	0.819	0.762
LiF	0.702	0.661	0.611	0.975	0.871	0.835
LiCl	0.463	0.499	0.495	0.939	0.796	0.734
LiBr	0.484	0.441	0.353	0.924	0.767	0.725
BF	0.150	0.233	0.153	0.623	0.276	0.086
CO	0.105	0.307	0.180	0.594	0.135	0.000
SO	0.256	0.456	0.447	0.640	0.245	0.243
CS	0.049	0.023	0.019	0.069	0.031	0.243
NO	0.025	0.195	0.200	0.500	1.194	0.012

Tab. 3b. Comparison of gas-phase absolute values of partial atomic charges (atomic units) calculated using selected approaches with B3LYP functional at DFT level.

Molecule	MUL	MUL'	MUL*	*NBO*	*APT*	*MSK*
B3LYP						
HF	0.291	0.465	0.317	0.548	0.417	0.452
HCl	0.136	0.226	0.244	0.253	0.211	0.250
HBr	0.110	0.231	0.177	0.184	0.131	0.184
ClF	0.154	0.282	0.232	0.335	0.279	0.150
BrF	0.234	0.298	0.324	0.405	0.355	0.192
BrCl	0.025	0.006	0.141	0.102	0.103	0.045
NaF	0.757	0.552	0.471	0.978	0.858	0.874
NaCl	0.545	0.564	0.622	0.928	0.811	0.791
NaBr	0.566	0.488	0.469	0.909	0.777	0.762
LiF	0.627	0.526	0.446	0.961	0.860	0.835
LiCl	0.402	0.405	0.399	0.925	0.787	0.734
LiBr	0.426	0.348	0.230	0.907	0.752	0.725
BF	0.232	0.097	0.037	0.550	0.271	0.111
CO	0.010	0.174	0.013	0.478	0.255	0.000
SO	0.222	0.364	0.275	0.601	0.320	0.243
CS	0.241	0.098	0.047	0.172	0.127	0.243
NO	0.091	0.112	0.067	0.500	0.186	0.012

rences from Exp. MUL, MUL', MUL* and NBO approaches provide results with rather high deviations from Exp.

Individual bond lengths are comparable at each level of calculation; however, MSK and APT charges show rather satisfactory results in comparison to the values calculated from experimental data. Each type of atomic partial charge shows only small changes in their values at various computational levels.

Assuming the basis set dependency of Mulliken values, MUL' values should be located between MUL

Tab. 3c. Comparison of absolute values of partial atomic charges (atomic units) calculated using selected approaches with M0-62X functional at DFT level.

Molecule / M0-62X	MUL	MUL'	MUL*	*NBO*	*APT*	*MSK*
HF	0.296	0.482	0.341	0.548	0.417	0.457
HCl	0.145	0.237	0.255	0.253	0.211	0.255
HBr	0.123	0.252	0.216	0.190	0.131	0.196
ClF	0.150	0.304	0.267	0.346	0.279	0.150
BrF	0.214	0.313	0.373	0.415	0.355	0.192
BrCl	0.035	0.029	0.114	0.095	0.103	0.045
NaF	0.780	0.610	0.541	0.978	0.858	0.874
NaCl	0.563	0.586	0.618	0.940	0.811	0.791
NaBr	0.585	0.516	0.521	0.925	0.777	0.762
LiF	0.647	0.577	0.511	0.964	0.860	0.835
LiCl	0.415	0.427	0.406	0.929	0.787	0.734
LiBr	0.441	0.377	0.295	0.912	0.752	0.725
BF	0.237	0.121	0.018	0.565	0.271	0.099
CO	0.018	0.192	0.052	0.487	0.255	0.000
SO	0.222	0.384	0.310	0.614	0.320	0.243
CS	0.196	0.092	0.044	0.172	0.127	0.243
NO	0.097	0.119	0.101	0.500	0.186	0.012

and MUL* values since they were calculated using medium basis set of 6-31G*. This is the case only with HCl, BrF, NaF, NaCl, NaBr, LiF, LiBr, BF, and CS molecules. In case of the BrCl molecule, its MUL' value is significantly lower than MUL and MUL* values, and for the rest of the studied molecules, MUL' values are higher than MUL and MUL* values. These inconsistencies suggest not only the basis set dependency of Mulliken values of atomic partial charges, but also recommend their calculation with several basis sets so the trend of Mulliken values can be estimated, more precisely.

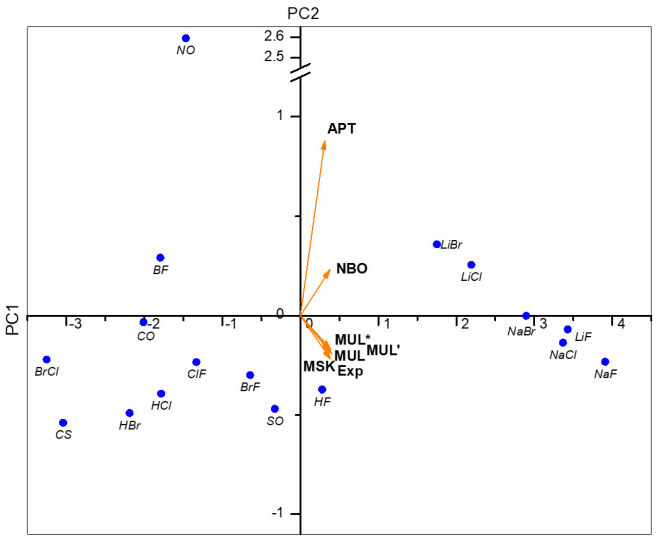
Additionally, most methods of atomic partial charges determination provide nearly identical results at the DFT level of theory regardless of the used functional. This fact implies independency of the studied population analyses from the functional used, or rather small influence of the functional on their results compared to the influence of the level of theory.

PCA was used to determine chemical accuracy of the used methods resulting in biplots (Fig. 1) and covariance matrices (Table 5). The main difference between PCA biplots was observed between MP2 and the remaining methods. Not only the relative ordering (top to bottom) and angles are different, but also the position of NO molecule is significantly different implying errors in the description of this radical molecule at the MP2 level of theory, which corresponds to data shown in Table 3a. Each dataset is divided into two groups: the first group consists of sodium and lithium halides and are located mostly in the first and fourth quadrant of a biplot; the second group consists of the other studied molecules, and it is located in the second and third quadrant with the exception of HF molecule.

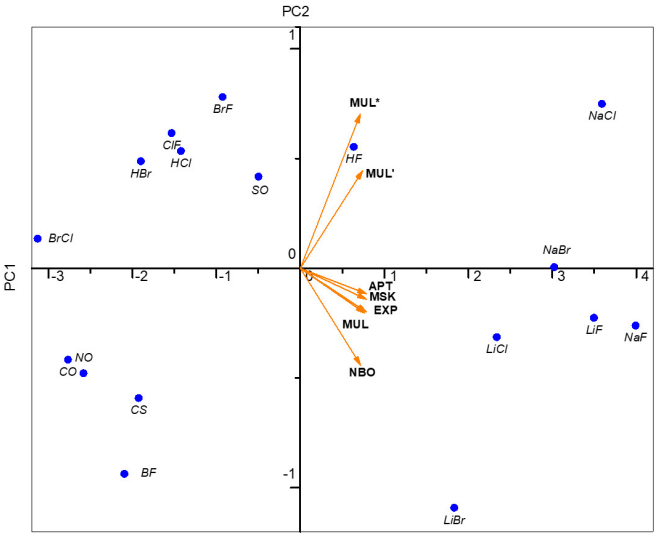
As mentioned before, differences observed between the MP2 and DFT biplots are mainly because of the APT charge value of NO, which is significantly displaced in comparison to other biplots. Such observation suggests that this radical molecule is not described well enough. It is also possible that APT method provides inaccurate values for radical molecules at the MP2 level of theory.

Biplots suggest that the most accurate method of partial atomic charges determination is MUL, followed by MSK. Mulliken method is expected to be rather accurate for MP2 geometries. However, by evaluation of covariance matrices (see Table 4), the most accurate method is MSK followed by APT at the DFT level of theory, probably due to an inconsistency in MUL data.

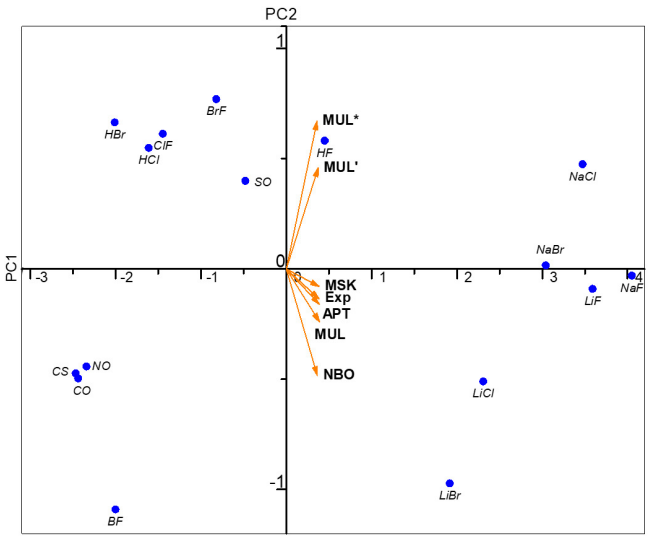
Table 5 shows the list of parameters for linear regression of every studied approach and Exp values. Linear regression was performed to assess the accuracy of individual approaches for theoretical determination of atomic partial charges. Chemical accuracy is determined from values of intercepts, slopes, and correlation coefficients. The closer



a)



b)



c)

Fig. 1. Primary component analysis for MP2 (a), B3LYP (b) and M06-2X (c) methods.

Tab. 4. Covariance table of PCA for individual methods of calculation provided by a computational tool (Jolliffe, 2002).

Group\Group	Exp	MUL	MUL'	MUL*	NBO	APT	MSK
				MP2			
Exp	1.000	0.950	0.818	0.804	0.816	0.659	0.995
MUL	0.950	1.000	0.913	0.900	0.873	0.656	0.946
MUL'	0.818	0.913	1.000	0.942	0.874	0.621	0.829
MUL*	0.804	0.900	0.942	1.000	0.803	0.624	0.811
NBO	0.816	0.873	0.874	0.803	1.000	0.756	0.798
APT	0.659	0.656	0.621	0.624	0.756	1.000	0.636
MSK	0.995	0.946	0.829	0.811	0.798	0.636	1.000
				B3LYP			
Exp	1.000	0.952	0.858	0.829	0.872	0.997	0.995
MUL	0.952	1.000	0.846	0.806	0.857	0.947	0.942
MUL'	0.858	0.846	1.000	0.911	0.809	0.872	0.877
MUL*	0.829	0.806	0.911	1.000	0.703	0.842	0.845
NBO	0.872	0.857	0.809	0.703	1.000	0.861	0.854
APT	0.997	0.947	0.872	0.842	0.861	1.000	1.000
MSK	0.995	0.942	0.877	0.845	0.854	1.000	1.000
				M06-2X			
Exp	1.000	0.956	0.865	0.842	0.868	0.960	0.996
MUL	0.956	1.000	0.869	0.836	0.869	0.934	0.950
MUL'	0.865	0.869	1.000	0.941	0.820	0.880	0.880
MUL*	0.842	0.836	0.941	1.000	0.733	0.846	0.858
NBO	0.868	0.869	0.820	0.733	1.000	0.951	0.853
APT	0.960	0.934	0.880	0.846	0.951	1.000	0.953
MSK	0.996	0.950	0.880	0.858	0.853	0.953	1.000

Tab. 5. List of intercepts (A), slopes (B), and determination coefficients R^2 of linear regression for correlations of individual theoretical absolute values of atomic partial charges and Exp values.

Data	B3LYP			M0-62X			MP2		
	A	B	R^2	A	B	R^2	A	B	R^2
MUL	–0.047	1.407	0.905	–0.043	1.369	0.913	0.001	1.184	0.903
MUL'	–0.118	1.594	0.736	–0.131	1.526	0.748	–0.114	1.264	0.670
MUL*	–0.035	1.537	0.687	–0.057	1.466	0.708	–0.103	1.305	0.646
NBO	–0.155	0.922	0.761	–0.154	0.911	0.753	–0.115	0.824	0.666
APT	–0.102	1.077	0.921	–0.102	1.077	0.921	0.081	0.596	0.434
MSK	–0.019	1.010	0.993	–0.020	1.009	0.992	–0.018	1.005	0.991

values of intercepts to 0 and slopes to 1 are, the more accurate the theoretical approach is, which is further supported by the correlation coefficient value.

Some theoretical approaches show higher accuracy than others, yet the worst correlation coefficient is observed for APT charge calculated at the MP2 level of theory. This is caused by inclusion of APT charge for the NO molecule. If data obtained for the NO molecule are disregarded in this case, the intercept value changes from 0.081 to –0.035 and the slope value changes from 0.596 to 0.962. The observed shift in values towards 0 for intercept and towards 1 for slope results in the determination coefficient R^2 changing to 0.913. Thus, APT values are reasonably accurate at the MP2 level of theory unless radicals are considered. With this observation in mind, all methods provide consistent results in terms of chemical accuracy. Thus, the most accurate results were obtained with the MSK method, followed by APT, MUL, NBO and least accurate results were observed with the MUL* method. Figure 2 shows

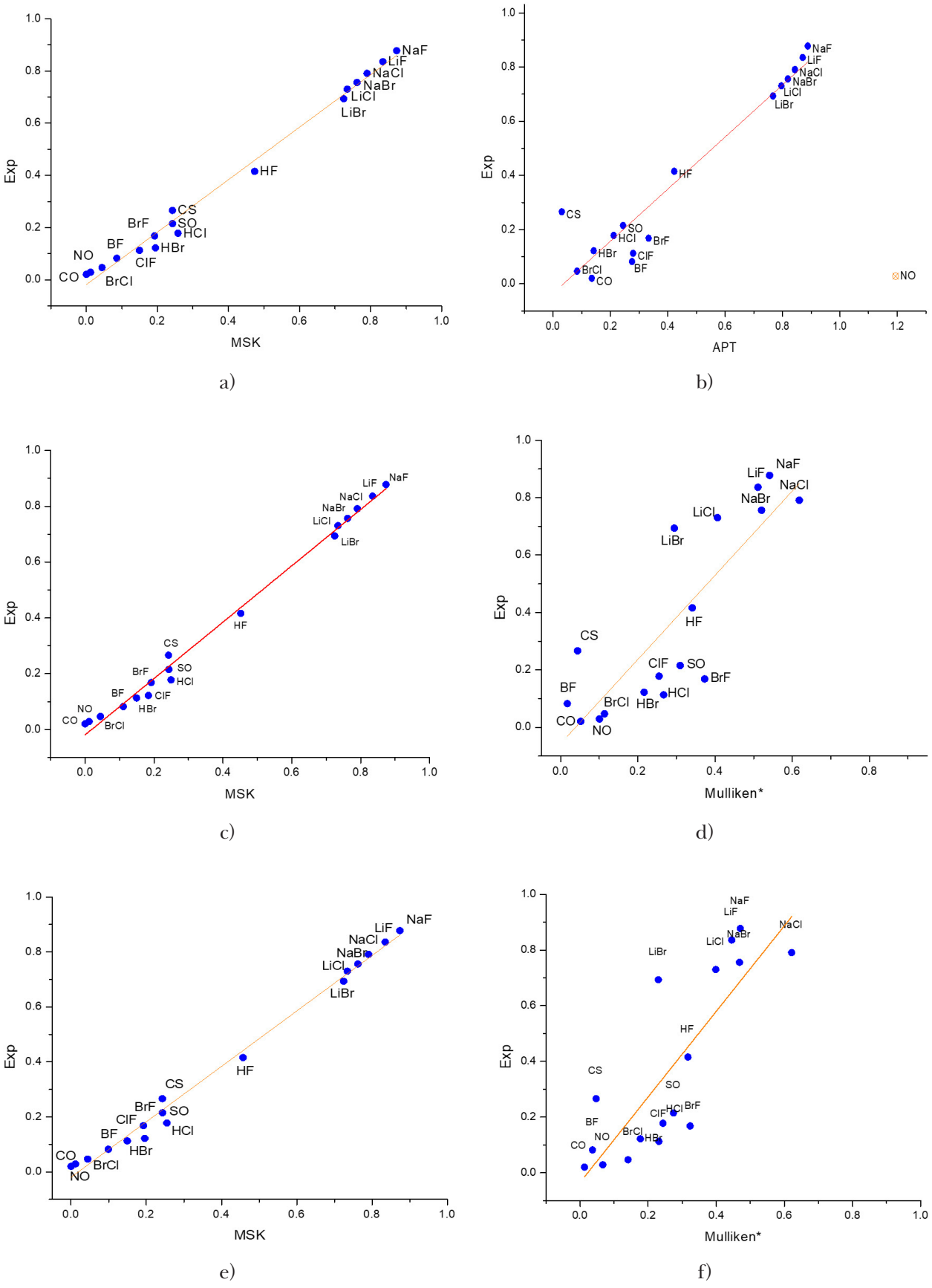


Fig. 2. Selected dependences of theoretical absolute values of partial atomic charges on experimental values. Selection is based on best and worst R^2 factor. Hollow symbols indicate data excluded from linear regression. for: a) MSK (MP2), b) APT(MP2), c) MSK(B3LYP), d) MUL*(B3LYP), e) MSK(M0-62X) and f) MUL*(M06-2X).

linear regression for the most and the least accurate results, whereas linear regression in Fig. 2b has a satisfactory course but it is acquired by exclusion of data for the NO molecule.

Conclusions

In this study, partial atomic charges were evaluated at different levels of theory and were further compared to their values obtained from experimentally acquired parameters.. To find the most accurate method of partial atomic charge determination, the principal component analysis was used. Furthermore, PCA determined the most accurate method tested.

MSK scheme provided results consistently and most accurately at the DFT and MP2 levels of calculation. MUL population analysis performed at the 6-311++G** basis set provided consistent but less accurate results than MSK, while NBO values are consistent and less accurate at the DFT and MP2 computational levels. Last evaluated method, APT, provides rather accurate results at the DFT level, but at the MP2 level of calculation, it showed most significant deviations from the obtained experimental values for radical molecules. The least accurate results were obtained with the Mulliken population analysis computed using the 6-31G* (MUL') and 6-31G basis set (MUL*) at the DFT level. The order of the least accurate methods at the MP2 level changes to MUL*, APT, and MUL', where MUL* is the least accurate method.

A slight dependency on the used functional has been found from further study of results of various methods of atomic partial charges determination. The difference between results is significant in the selection of the level of theory that calculations are to be performed at.

Each of the used methods showed the same trends but the calculated values should be checked, by another computational method, which is also recommended when using the Mulliken population analysis with smaller basis sets. Principal component analysis is rather suitable for the determination of trends and dependencies of chemical and physical parameters.

Correlations of individual theoretical approaches with the experiments using linear regression have confirmed MSK to be the most accurate approach at all levels of theory. It was also determined that the APT approach performed at the MP2 level of theory provides rather accurate results in case radical molecules are not included in the studied set.

Acknowledgment

This work was supported by the Slovak Research and Development Agency (APVV-19-0024) and VEGA 1/0461/21.

References

Becke AD (1993) J. Chem. Phys., 98: 5648—5652.

Besler BH, Merz KM, Kollman PA (1990) Journal of Computational Chemistry, 11(4): 431—439.

Biarge J, Herranz J, Morcillo J (1961) Annals of the Royal Spanish Society of Physics and Chemistry, 57, p 81.

Breneman MC, Wiberg BK (1990) Journal of Computational chemistry 11(3): 361.

Cagardová D, Matúška J, Poliak P, Lukeš V (2019) J. Phys. Chem. C.

Cagardová D, Matúška J, Michalík M, Poliak P, Lukeš V (2020) Synthetic Metals, 260: 116263.

Cagardová DM, Matúška J, Michalík M, Lukeš V (2022) Materials Today Communications, 33: 104688.

Frisch MJ, Head-Gordon M, Pople JA (1990) Chem. Phys. Lett., 166: 275—280.

Gaussian 16, Revision B.01, Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Petersson GA, Nakatsuji H, Li X, Caricato M, Marenich AV, Bloino J, Janesko BG, Gomperts R, Mennucci B, Hratchian HP, Ortiz JV, Izmaylov AF, Sonnenberg JL, Williams-Young D, Ding F, Lipparini F, Egidi F, Goings J, Peng B, Petrone A, Henderson T, Ranasinghe D, Zakrzewski VG, Gao J, Rega N, Zheng G, Liang W, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Throssell K, Montgomery JA Jr, Peralta JE, Ogliaro F, Bearpark MJ, Heyd JJ, Brothers EN, Kudin KN, Staroverov VN, Keith TA, Kobayashi R, Normand J, Raghavachari K, Rendell AP, Burant JC, Iyengar SS, Tomasi J, Cossi M, Millam JM, Klene M, Adamo C, Cammi R, Ochterski JW, Martin RL, Morokuma K, Farkas O, Foresman JB, Fox DJ. Gaussian, Inc., Wallingford CT, 2016.

Gross KC, Seybold PG, Hadad CM (2009) Int. J. Quantum Chem., 109: 3679—3684.

Hariharan PC, Pople JA (1973) Theor. Chim. Acta., 28: 213—222.

Huber KP, Herzberg G (2013) Molecular Spectra and Molecular Structure. Springer New York, NY.

Johnson III RD (2022, May 22), NIST Computational Chemistry Comparison and Benchmark Database, http://cccbdb.nist.gov/.

Jolliffe IT (2002) Principal Component Analysis in: Springer Series in Statistics, Springer New York, NY.

Irikura KK (2007) Journal of Physical and Chemical Reference Data, 36: 389—397.

Klein E, Lukeš V (2006) Chem. Phys., 330(3): 515—525.

Klein E, Lukeš V (2007) Journal of Molecular Structure: THEOCHEM, 805(1-3): 153—160.

Lee C, Yang W, Parr RG (1988) Phys. Rev. B, 37: 785—789.

Mehmood A, Janesko BG (2017) Angew. Chem. Int., 56: 6878—6881.

Michalik M, Poliak P, Klein E, Lukeš V (2018a) Chem. Phys. Let., 709: 71—76.

Michalik M, Poliak P, Lukeš V (2018b) Acta Chimica Slovenica, 65(1): 23—33.

Mulliken RS (1955) The Journal of Chemical Physics, 23(10): 1833—1840.

Nelson Jr RD, Lide Jr DR, Maryott AA (1967) National Standard Reference Data Series — National Bureau of Standards 10, U.S. Department of Commerce.

Person W, Newton J (1974) J. Chem. Phys. 61, 1040—1049.
Raczyńska ED, Hallman M, Kolczyńska K, Stepniewski (2010) Symmetry, 2: 1485—1509.
Reed AE, Weinstock RB, Weinhold F (1985), J. Chem. Phys., 83(2): 735—746.
Richter WE, Duarte LJ, Bruns RE (2021) J. Chem. Inf. Model., 61(8): 3881—3890.
Rostkowski M, Paneth P (2007) Pol. J. Chem., 81: 711—720.
Singh UC, Kollman PA (1984) J. Comp. Chem., 5: 129—145.
Stephens PJ, Devlin FJ, Chabalowski CF, Frisch MJ (1994) J. Phys. Chem, 98: 11623—11627.
Štellerová D, Michalík M, Lukeš V (2022) Phytochemistry, 203: 113387.
Svobodová Vařeková R, Geidl S, Ionescu CM, Skřehota O, Kudera M, Sehnal D, Bouchal T, Abagyan R, Huber HJ, Koča J (2011) J. Chem, Inf. Model, 51(8): 1795—1806.
Thakkar AJ (2021) Quantum Chemistry (Third Edition): A concise introduction for students of physics, chemistry and materials science, Institute of Physics Publishing, Bristol, United Kingdom.
Ugur I, Marion A, Parant SP, Jensen JH, Monard G (2014) J. Chem. Inf. Model., 54: 2200—2213.
Vosko SH, Wilk L, Nusair M (1980) Can. J. Phys., 58: 1200—1211.
Williams DE, Yan JM (1988) Adv. Atomic Mol. Phys, 23: 87.