Quantification of $\sigma$–holes and their use as a descriptor for the theoretical calculation of pKa values for carboxylic acids

Guillermo Caballero-García 1, Gustavo Mondragón-Solórzano 1, Raúl Torres-Cadena 1, Marco Diaz-García 1, Jacinto Sandoval-Lira 1, Joaquín Barroso-Flores 1,*

1 Centro Conjunto de Investigación en Química Sustentable UAEM – UNAM, Carretera Toluca–Atlacomulco km 14.5, Unidad San Cayetano, Personal de la UNAM, Toluca 50200, Estado de México, Mexico; jbarroso@unam.mx

* Correspondence: jbarroso@unam.mx; Tel.: +52-722-2766610 Ext 7754

Abstract: Theoretical approaches to calculate pKa values for Brønsted acids is a challenging task that, most of the time, involves sophisticated and time-consuming methods. Therefore, heuristic approaches are efficient and appealing methodologies to approximate these values. Herein, by considering the electrostatic potential on acidic hydrogen atoms in a similar fashion that a $\sigma$–hole is defined, we calculated the maximum surface potential, $V_{\text{S,max}}$, and used it as a descriptor to correlate it with experimental acidity constants. These values were calculated using the CPCM implicit solvent model (water) with six different methods: five density functionals and the Møller–Plesset second order perturbation theory. Six different basis sets were combined with each method in order to benchmark a total of thirty-six levels of theory. Overall, 1080 calculations were performed and found to correlate with experimental data. The $\omega$B97X-D/6-31+G(d,p) level of theory stands as the best one for consistently reproduce the reported pKa values.

Keywords: pKa; Hydrogen Bond; $\sigma$–hole.

1. Introduction

The term ‘tetrel bond’ was first used by Frontera in 2013 to designate the directional, intermolecular and non-covalent interactions of electrostatic nature involving elements in group 14. These atoms behave as electrophiles through their interaction with either $n$ or $\pi$ electrons from Lewis bases [1,2]. Such interactions are similar to those observed for elements in groups 15 and 16, which are termed pnictogen and chalcogen bonds, respectively. The formation of these non-covalent interactions stems from a similar origin as halogen bonds, via the presence of $\sigma$–holes [3–5], a localized region of positive electrostatic potential on the surface of the group 17 atom and opposite to the internuclear axis of a covalent $\sigma$ bond, hence $\sigma$–holes. Therefore, tetrel bonds belong to a broader kind of directional non-covalent electrostatic interactions like halogen [6–8], chalcogen [9,10], and pnictogen bonds [11]. A stretch of this label has been applied to hydrogen bonding, despite the absence of $p$ electrons on hydrogen atoms and the high polarizability of their bonds [12,13].

Tetrel bonds are stabilizing interactions in nature [14,15] that are able to form cooperative networks [9,16–20], a feature that is used as a powerful tool for the design of crystal structures [21–23]. The stabilization arising from these interactions ranges from 1 kcal/mol to 50 kcal/mol [24]. The strength of the interaction increases as the tetrel atom increases its atomic number; the electronegativity of the atom bonded opposite to the tetrel bond; and the number of electron withdrawing groups bonded to the tetrel atom. Therefore, the formation and strength of a tetrel

© 2018 by the author(s). Distributed under a Creative Commons CC BY license.
bond depend closely on the polarization of the electron density surrounding the tetrel atom. These factors have been extensively investigated by Scheiner, who has further divided them into electronic [24,25] and steric [26] contributions.

Experimentally, tetrel bonds have been observed by means of H, 13C, and 207Pb NMR techniques [27–29], as well as through X-ray diffraction methods from which their effects in the formation of crystal structures is obtained by direct measurements [21,23,30], including systems where tetrel–tetrel bonds are formed [31]. However, their interesting electronic structure is the feature that has attracted most of the research. Several computational studies on their nature have been published so far, from their strict quantum treatment [32] to their charge transfer dynamics in the attoseconds regime [33] and the tunneling bond-breaking processes promoted by σ-holes [34]. Studies regarding the specific features for each kind of derivatives of the tetrel atoms are reported. These include silicon oxide based compounds such as silanol, siloxane and others [35–37]; tin and lead compounds [38]; unsaturated hydrocarbons [39,40], carbenes[41], anionic species [42–44]; and even single–electron tetrel bonds involving the -CH3 radical species [45]. Non–tetrel atoms such as halogens [44,46], or beryllium bonded complexes which are able to induce the formation of σ-holes [47,48] have also been studied.

Unsaturated carbon species can also form tetrel–bonds involving their π electron density with various electron rich donors such as CO, CS [49] and OCS with nitrogen bases [50–52], σ–π bonded complexes [4,53], π–σ-hole [54] bonded complexes and cation–π interactions [55], and π electrons bearing carbon compounds with potential applications for CO2 capture [56–58]. Thus, the importance of the study of non–covalent interactions has large implications for crystal engineering [59], biochemical engineering and the understanding of chemical reactivity [60–62]. In our research group we have reported the chemical reduction of a trichloromethyl group into a methyl group via the attack of σ-holes on chlorine atoms by thiophenolate anions, a reaction mechanism which is extensible to other trichloromethyl compounds [63].

The hydrogen bond has become a paradigm among the toolbox of chemical concepts [64]. A strong, directional, non–covalent interaction that is responsible for a massive number of chemical phenomena [65–67]. The accurate prediction of pKa values for carboxylic acids by means of computational methods covers a wide range of potential applications from chemical design to drug development [68–70]. However, calculating the equilibrium constant for the deprotonation of a Brønsted acid implies the use of sophisticated and computationally intensive methods, such as G3MP2, to calculate solvation free energies for all the species involved in the associated thermodynamic cycle [71]. It is commonly regarded that the dissociation process is mainly controlled by electrostatic interactions, given the partial positive charge on the acid hydrogen. Since this interaction is not isotropic some parallel between hydrogen and σ-hole bonded systems arise. The formation of these directional interactions implies the presence of an electrostatic potential maximum located on the opposite side of the O–H σ–bond, which can be quantified by the maximum surface electrostatic potential, \(V_{S,max}\). By postulating herein that the first stages of the deprotonation process of any given carboxylic acid occur through an interaction akin a tetrel or halogen bond between the acid and a water molecule, R-COOH···OH2, we can use the \(V_{S,max}\) value as an efficient descriptor that strongly correlates with the measured pKa value of a carboxylic acid. Previously, the nucleophilicities and electrophilicities of Lewis acids and bases, respectively, have been derived from interpolation of their mutual dissociation energies [72].

2. Results

Thirty (30) different carboxylic acids with reported pKa values were selected from Lange’s Handbook of Chemistry [73], optimized and the surface electrostatic potential calculated (see methods section for full details). The structure of the acids are shown in Figure 1. The levels of theory used were obtained from the combination of the following functionals, ωB97X-D (A), B3LYP (B), LC-ωPBE (C), M06-2X (D) and PBE0 (E), as well as the Møller-Plesset second-order perturbation
theory, MP2 (F) and the following basis sets, 6-31+G(d,p) (1), 6-311++G(d,p) (2), cc-pVDZ (3), cc-pVTZ (4), aug-cc-pVTZ (5) and Def2-TZVP (6).

Figure 1. Thirty carboxylic acids comprising the chemical space under study.

In total, thirty-six levels of theory were used to calculate the electronic structure of the thirty carboxylic acids which comprise the chemical space under study for a total of 1,080 different wave functions upon which the maximum surface potential, $V_{S,max}$, was calculated and plotted against the experimental pK$_{a,exp}$ value. Simple linear regressions were performed to obtain the best fittings. The $V_{S,max}$ on each acidic hydrogen atom was used for the correlations, as an example, Figure 2 depicts the location of $V_{S,max}$ on the acid hydrogen atom for compound 14. This value is calculated on the isodensity surface $\rho = 0.001$ a.u. and is used as a descriptor for the magnitude of a $\sigma$-hole equivalent on hydrogen.
Figure 2. Maximum surface electrostatic potential, $V_{S,max}$, over the acidic hydrogen atom shown for compound 14 taking an isodensity value of 0.001 a.u. (isosurface not shown)

All correlation coefficients, slopes, and intercepts for all thirty-six levels of theory are collected in Table 1.

Table 1. Linear regression parameters obtained for the pKa vs. $V_{S,max}$ plots. Intercept units in kcal/mol

<table>
<thead>
<tr>
<th>Method</th>
<th>Slope</th>
<th>Intercept</th>
<th>$R^2$</th>
<th>Method</th>
<th>Slope</th>
<th>Intercept</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>-0.1890</td>
<td>15.6915</td>
<td>0.9241</td>
<td>D1</td>
<td>-0.1935</td>
<td>15.9819</td>
<td>0.9310</td>
</tr>
<tr>
<td>A2</td>
<td>-0.1901</td>
<td>15.3454</td>
<td>0.9227</td>
<td>D2</td>
<td>-0.1893</td>
<td>15.1661</td>
<td>0.9315</td>
</tr>
<tr>
<td>A3</td>
<td>-0.2072</td>
<td>15.5189</td>
<td>0.9123</td>
<td>D3</td>
<td>-0.2133</td>
<td>15.7926</td>
<td>0.9002</td>
</tr>
<tr>
<td>A4</td>
<td>-0.2021</td>
<td>15.8188</td>
<td>0.9174</td>
<td>D4</td>
<td>-0.2007</td>
<td>15.7719</td>
<td>0.9225</td>
</tr>
<tr>
<td>A5</td>
<td>-0.1982</td>
<td>15.8367</td>
<td>0.9161</td>
<td>D5</td>
<td>-0.1765</td>
<td>14.5112</td>
<td>0.8796</td>
</tr>
<tr>
<td>A6</td>
<td>-0.2045</td>
<td>15.8616</td>
<td>0.9168</td>
<td>D6</td>
<td>-0.2033</td>
<td>15.6156</td>
<td>0.9200</td>
</tr>
<tr>
<td>B1</td>
<td>-0.1830</td>
<td>14.7367</td>
<td>0.8996</td>
<td>E1</td>
<td>-0.1867</td>
<td>15.3189</td>
<td>0.9084</td>
</tr>
<tr>
<td>B2</td>
<td>-0.1833</td>
<td>14.3499</td>
<td>0.8993</td>
<td>E2</td>
<td>-0.1860</td>
<td>14.8108</td>
<td>0.8973</td>
</tr>
<tr>
<td>B3</td>
<td>-0.2077</td>
<td>14.8807</td>
<td>0.8952</td>
<td>E3</td>
<td>-0.2035</td>
<td>15.1196</td>
<td>0.8955</td>
</tr>
<tr>
<td>B4</td>
<td>-0.1983</td>
<td>14.9381</td>
<td>0.8999</td>
<td>E4</td>
<td>-0.2007</td>
<td>15.4529</td>
<td>0.9016</td>
</tr>
<tr>
<td>B5</td>
<td>-0.1895</td>
<td>14.6298</td>
<td>0.8886</td>
<td>E5</td>
<td>-0.1937</td>
<td>15.2129</td>
<td>0.8963</td>
</tr>
<tr>
<td>B6</td>
<td>-0.2017</td>
<td>14.8703</td>
<td>0.8223</td>
<td>E6</td>
<td>-0.2017</td>
<td>15.3350</td>
<td>0.8954</td>
</tr>
<tr>
<td>C1</td>
<td>-0.1834</td>
<td>15.2811</td>
<td>0.8972</td>
<td>F1</td>
<td>-0.1937</td>
<td>15.5261</td>
<td>0.9046</td>
</tr>
<tr>
<td>C2</td>
<td>-0.1808</td>
<td>14.8234</td>
<td>0.9051</td>
<td>F2</td>
<td>-0.2033</td>
<td>15.4111</td>
<td>0.9041</td>
</tr>
<tr>
<td>C3</td>
<td>-0.2136</td>
<td>15.8429</td>
<td>0.9037</td>
<td>F3</td>
<td>-0.2183</td>
<td>15.7690</td>
<td>0.9158</td>
</tr>
<tr>
<td>C4</td>
<td>-0.1839</td>
<td>14.8594</td>
<td>0.8935</td>
<td>F4</td>
<td>-0.2099</td>
<td>15.7708</td>
<td>0.8957</td>
</tr>
<tr>
<td>C5</td>
<td>-0.2005</td>
<td>15.9657</td>
<td>0.9187</td>
<td>F5</td>
<td>-0.1999</td>
<td>15.4754</td>
<td>0.9005</td>
</tr>
<tr>
<td>C6</td>
<td>-0.2078</td>
<td>15.9734</td>
<td>0.9171</td>
<td>F6</td>
<td>-0.2085</td>
<td>15.5996</td>
<td>0.8927</td>
</tr>
</tbody>
</table>

The obtained linear model is shown in Figure 3 for method (A) only, the plots with the rest of the methods (B) – (F) are presented in the Supporting Information section (Figures S 1 to S 12).
Figure 3. Linear correlations between pK\textsubscript{a}\textsubscript{exp} against \(V_{S,\text{max}}\) for DFT method (A) with the six basis sets (1) through (6).

To further analyze the obtained models, a comparison between experimental and calculated pKa values is made by calculating \(\Delta pK_a = pK_{a,\text{exp}} - pK_{a,\text{calc}}\). Figure 4 shows these plots for results obtained with the functional (A), the corresponding \(\Delta pK_a\) plots for the other levels of theory are collected in the Supporting Information section (Figures S 1 – S 12).
3. Discussion

3.1. Computational method: DFT or ab initio?

From all the tested levels of theory, the highest $R^2$ correlation coefficients (Table 1) between $V_{5,\text{max}}$ and $pK_{a,\text{exp}}$ values are obtained consistently with the $\omega$B97X-D functional (A), whereas the lowest are obtained with the B3LYP functional (B). The latter method, albeit being one of the most popular ones to model organic molecules, could inadequately be describing the surface potential due to the lack of dispersion effects in the functional, which is not the case for $\omega$B97X-D that includes implicit dispersion. A similar performance to that of B3LYP was observed for the PBE0 functional (E), and just slightly improved for LC-$\omega$PBE (C). The latter functional was thought to yield much better results due to the long range correlation term, however that was not the case.

The M06-2X functional (D) also shows to be properly describing the surface electrostatic potentials, as shown in the high correlation coefficients. This is plausibly because of the dispersion terms included in its formulation, which also applies for the $\omega$B97X-D functional.

Although the M06-2X functional is widely used and regarded as probably the best functional to model organic reactions, yields a larger discrepancy in the $\Delta pK_a$ plots than those obtained with
ωB97X-D (Figures 4 and S 7). In sharp contrast with all the other tested functionals (A, B, C and E), the Minnesota functional has the largest values of ΔpKa (Figure S 8), which indicates that may not be an adequate functional to estimate acidity constants using linear regressions.

As a benchmark comparison standard, the Møller-Plesset second-order perturbation theory, MP2 (F), was included in the study, not only to assess its accuracy, but to compare DFT and at least one wave function method as well. Despite of performing similarly to ωB97X-D yielding low error bars (ΔpKa, Figure S 12), the correlation coefficients for the MP2 levels of theory were lower than those for ωB97X-D. This shows that, for the case of modeling surface electrostatic potentials, a computationally expensive method may not always be preferred, as very similar or even better results can be obtained with a less demanding approach in just a fraction of time.

Amongst the set of models obtained for (A), the A1 level of theory has the highest correlation ($R^2 = 0.9241$) corresponding to the ωB97X-D/6-31+G(d,p).

3.2. Basis set: Is larger better?

Most of the reported benchmarks to model organic molecules deal with the DFT functional. However, little attention is payed to the basis set, or more precisely, to the proper functional/basis set combination. It is quite a common paradigm among computational chemists that the larger the basis set the better; therefore, using a CBS method as a benchmark standard, but how much is enough.

Four out of the six methods yielded the strongest $V_{\text{SSmax}}$-pKa correlations when using the relatively medium size 6-31+G(d,p) basis set. This was not the case for the LC-ωPBE functional, which required the largest basis set (5) under study for a good correlation coefficient. Surprisingly, the M06-2X functional presented the largest ΔpKa deviations when combined with the largest basis set aug-cc-pVTZ (Figure S 8).

In the case of the MP2 calculations (Figure S 11), increasing the so-called quality of the basis set may not be beneficial in all cases. When comparing the split-valence Pople’s basis sets, practically the same correlation was found with the double-ζ set and the corresponding triple-ζ quality one, 0.9046 versus 0.9041, respectively. On the other hand, the Dunning-Huzinaga basis showed a deemed correlation when increasing the set size from cc-pVDZ to cc-pVTZ, 0.9158 and 0.8957, respectively. However, ΔpKa deviations are practically consistent among the MP2 levels of theory.

In terms of the difference between experimental and correlated pKa values the A1 level of theory yields the smallest ΔpKa deviations with most of the differences kept under 0.5 pKa units, showing that, for this case, the larger the basis set size may not always be the better.

3.3. Limitations: A final remark

From the thirty-six levels of theory tested in this study, four compounds presented the largest ΔpKa deviations: 6, 7, 12 and 19. This possibly be due to strong delocalization effects from nearby π bonds to the $\sigma^{\text{O-H}}$ orbital in the acidic hydrogen atom or intramolecular hydrogen bonding with Lewis basic motifs (Figure 5). For such kind of compounds, further improvements are required in the methodology for our linear models.

Figure 5. Electronic effects to be considered in further improvements of the method.
So far, the applicability domain of these regression is limited by the pKa data used to construct the models (0.5 < pKa < 5.0), caution must be taken when using the linear models presented herein for molecules outside this range.

4. Materials and Methods

Geometry optimizations and wave function printouts for the 30 carboxylic acids were performed using the Gaussian 09 rev. E01 suite of programs [74] at each of the different levels of theory (see text). All calculations included the CPCM implicit solvation model (water). Frequency analyses were done at the end of each geometry optimization at the same level of theory in order to verify that the found geometries corresponded to energy minima. Ultrafine integration grid was used in all the calculations.

The maximum surface potential calculations were performed on the wave function files with the MultiWFN program, version 3.3.8 [75] using an isodensity value of 0.001 a.u. All the computed values are collected in the Supporting Information (Tables S 1 to S 6).

5. Conclusions

With our calculations, VS,max has proven not only to be a suitable descriptor of the magnitude of a σ–hole, but a proper quantity that correlates with the pKa value of carboxylic acids as well. By means of DFT calculations with the use of a simple implicit solvent model (CPCM), the value of VS,max can be calculated and the equations obtained herein can be used to estimate pKa values without the need for a full thermodynamic cycle calculation, thus, avoiding long computations of solvation free energies and other costly quantities.

The ωB97X-D/6-31+G(d,p) level of theory (A1) yielded the lowest ΔpKa values, standing as the best choice for estimating the pKa of any given acid through the calculation of VS,max. Hence, we highly recommend this level of theory for geometry optimization and wave function file print. Care must be taken, as the pKa value sought after should be between 0.5 and 5.0 pH units.

Further testing is needed for these regression models to become universal. Inclusion of intramolecular hydrogen bonding as well as highly delocalizing features within the chemical space are key features to be considered in future improvements of the model. Our proposed descriptor is also dependent of the isodensity value for the definition of the surface upon which it is calculated, and it is highly recommended to keep the value suggested by Bader et al. [76] of ρ = 0.001 a.u.

Supplementary Materials: The following are available online, Table S1-S6: Calculated VS,max values for carboxylic H atoms to the different levels of theory studied, Table S7: Reported pKa values for carboxylic acids studied. Figure S1-S5: Correlation of pKa exp vs VS,max. Figure S6-S10: Difference between the experimental and calculated pKa values (ΔpKa exp-cal).


Funding: This research received no external funding.

Acknowledgments: We thank DGTIC – UNAM for granting us access to their supercomputing facilities known as ‘Miztli’. We also thank Ms. Citlalit Martinez for keeping our local computational facilities running properly. JSL thanks DGAPA – UNAM for the postdoctoral scholarship.

Conflicts of Interest: The authors declare no conflict of interest.
References


17. Solimannejad, M.; Orojloo, M.; Amani, S. Effect of cooperativity in lithium bonding on the strength of
halogen bonding and tetrel bonding: \((\text{LiCN})_n \cdots \text{ClYF}_3\) and \((\text{LiCN})_n \cdots \text{YF}_3\text{Cl}\) (Y = C, Si and \(n=1-5\)) complexes as a working model. *J. Mol. Model.* 2015, 21, doi:10.1007/s00894-015-2722-1.


<table>
<thead>
<tr>
<th>Number</th>
<th>Reference</th>
</tr>
</thead>
</table>