

Scientific paper

Modelling the Correlation Between Molecular Electrostatic Potential and pK_a on Sets of Carboxylic Acids, Phenols and Anilines

Miha Virant,^{1,*} Sara Drvarič Talian,^{2,*} Črtomir Podlipnik¹
and Barbara Hribar-Lee¹

¹ Faculty of Chemistry and Chemical Technology, University of Ljubljana, Večna pot 113, SI-1000 Ljubljana, Slovenia

² National Institute of Chemistry, Hajdrihova 19, SI-1000 Ljubljana, Slovenia

* Corresponding author: E-mail: miha.virant@fkkt.uni-lj.si; sara.drvarictalian@ki.si

Received: 04-10-2016

For Cutting Edge 2017

Abstract

Calculations of molecular electrostatic potential were correlated with experimental pK_a values for different sets of acidic molecules (carboxylic acids, phenols, and anilines) to obtain linear relationships of variable quality. A single tri-parameter model function was constructed to describe the pK_a dependence on MEP maxima together with two automatically generated molecular descriptors, namely the counts of carboxylic acid and amine functional groups.

Keywords: molecular electrostatic potential, pK_a, quantum mechanical calculations, multivariable linear regression

1. Introduction

Any distribution of the electrical charge in a molecule creates an electrostatic potential $V(\mathbf{r})$ at each point of the surrounding space \mathbf{r} :¹

$$V(\mathbf{r}) = \sum_A \frac{z_A}{|\mathbf{R}_A - \mathbf{r}|} - \int \frac{\rho(\mathbf{r}') d\mathbf{r}'}{|\mathbf{r}' - \mathbf{r}|} \quad (1)$$

where z_A is the charge of the nucleus A located at \mathbf{R}_A , and $\rho(\mathbf{r})$ is the electronic density function of the molecule, that can be determined either computationally, or experimentally by diffraction methods.² This so-called molecular electrostatic potential (MEP) is a well-established tool for interpreting and predicting molecular reactivity.^{2–5}

One of the physico-chemical properties that have recently been shown to be correlated with the MEP is pK_a value which describes the acid strength of certain organic acids.^{6–7} The background of the correlation is due to the fact that MEP is a quantum-mechanical descriptor influenced by electronic or stereoelectronic properties of the groups in close vicinity to a particular acidic atom. Similarly, NAO (the sum of valence p natural atomic orbitals) has also been used for such calculations and it has been

proven that this descriptor is equivalent to MEP.⁸ Other examples of descriptors in QSAR models that have already been successfully used for determining pK_a values are topological⁹, atom type descriptors¹⁰, and group philicity¹¹.

The correlation was established by comparing the calculated MEP values via different theoretical methods with the experimentally determined pK_a values for the same sets of compounds. The success of the method consequently depends on the training set used in the program learning process, and the similarity of the studied compounds.⁷

A relatively good correlation between the maximum in MEP and the pK_a value was found for amines, anilines, carboxylic acids, alcohols, sulfonic acids, and tioles.¹² Furthermore, a good correlation was established recently between pK_a and the most positive value of the MEP of benzoic acids and phenols, and between the most negative value of MEP and the local ionization energies of these compounds.¹³ Nevertheless, all the published work has studied only one set of similar compounds and to the best of our knowledge, a model that could be applied to the wider array of differently functionalized molecules has not yet been presented.

Since the knowledge of the pK_a values is vital for predicting the reactivity of the compounds in question and is therefore of broad interest in drug design, toxicology studies, chemical synthesis etc., it is desirable to extend the current knowledge to other compounds.

This paper describes the correlation between pK_a values and MEP for certain tested carboxylic acids, phenols, and anilines and additional physico-chemical descriptors, which are introduced in a regression model to improve the correlation.

2. Methods

Calculations of molecular electrostatic potential maxima (V_{max}) were executed using Spartan '14 (v. 1.1.4) software. Energy minimizations of equilibrium geometries were performed on sets of 39 aliphatic carboxylic acids, 17 monosubstituted benzoic acids, 19 mono-substituted anilines, and 19 phenols (Given in Supporting Information, Table S1).

The starting molecular conformations were formed using the Merck molecular force field (MMFF) optimization.¹⁴ Further, semi-empirical methods AM1, and PM6, ab initio Hartree-Fock method HF 6-31G*, and a density functional theory method B3LYP 6-31G* were used to calculate the MEP.¹⁵

Several isoelectronic densities (0.2, 0.02, 0.002, and 0.0002) were examined for a linear correlation between the maximum in MEP, V_{max} , and pK_a values obtained from the literature.^{16,17} The best correlation was obtained for the isoelectronic value of 0.02, which was used in the further calculations.

To improve the correlation between pK_a and calculated descriptors from MEP, CANVAS's ligfilter descriptors were included into the calculation of multiple linear regression analysis with CANVAS software from Schrödinger 2015 Suite. These descriptors denote the presence/absence of certain functional groups, which are important for the stability of the molecules studied. The model was constructed using V_{max} and two best suited ligfilter descriptors, which count the number of certain functional groups.

For the validation of the model, a group of 87 compounds was divided into two sets. The first set of compounds (named training set) was composed of 70% randomly selected compounds, which were used for building a model. The second one (testing set) contained the remaining 30% of compounds used for the validation of the prediction capacity of the model.

The best subset of descriptors for the construction of the model was selected with CANVAS's simulated annealing¹⁹ algorithm with performing maximum of 1000 MC steps, initial and final temperature which corresponds to standard deviation of y being 0.5 and 0.05 respectively.

3. Results and Discussion

The results for the correlation between V_{max} in MEP for the isoelectronic density 0.02 and pK_a for phenols (a), anilines (b), aliphatic (c), and benzoic acids (d) are given below. Multivariable analysis of the compounds studied is presented in subsection (e).

a) The correlation between V_{max} and pK_a for the chosen set of monosubstituted phenols is shown in Figure 1. The results obtained with different methods are shown with different symbols/colours.

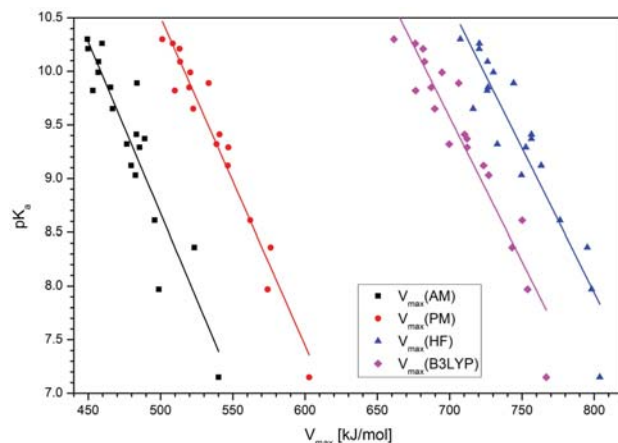


Figure 1. The correlation between the V_{max} calculated using different methods at IsoValue 0.02, and pK_a values^{16,17} of phenols.

Regardless of the used method for calculating the MEP, the correlation coefficient, R^2 , is around 0.8. If we disregard the values for nitrophenols calculated via PM6 method, the slopes of the fitted curves are approximately the same, indicating that the results are not dependent on the method of calculation.

b) The results for the anilines are shown in Figure 2. For the determination of the correlation for the set of anilines, we modelled the V_{max} on protonated molecules (ani-

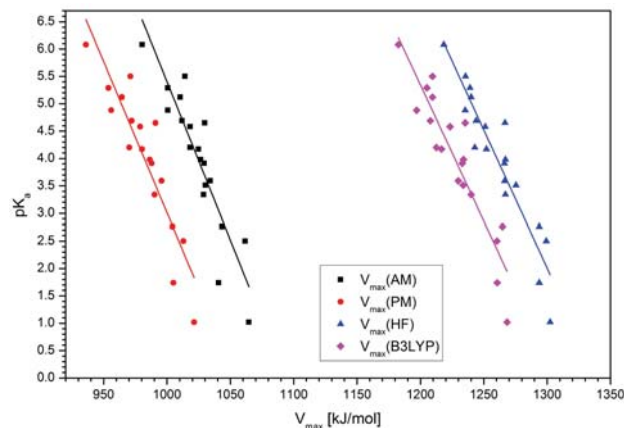


Figure 2. The correlation between the V_{max} calculated using different methods at IsoValue 0.02, and pK_a values^{16,17} of anilines.

linium cations). As with the set of phenols, the value of R^2 being higher than 0.8 shows good correlation between calculated and experimental properties.

c) In the case of aliphatic acids, a poor correlation between V_{\max} and pK_a was obtained (Figure 3). The correlation coefficient R^2 ranged from 0.395 to 0.588, depending on the method of calculation, the semi-empirical PM6 method having the lowest score. Poor correlations could be the consequence of structural diversity of the test set, and its larger pK_a range span.¹⁵ However, to indisputably prove this speculation, further calculations on a larger set of systematically substituted aliphatic acids would be required.

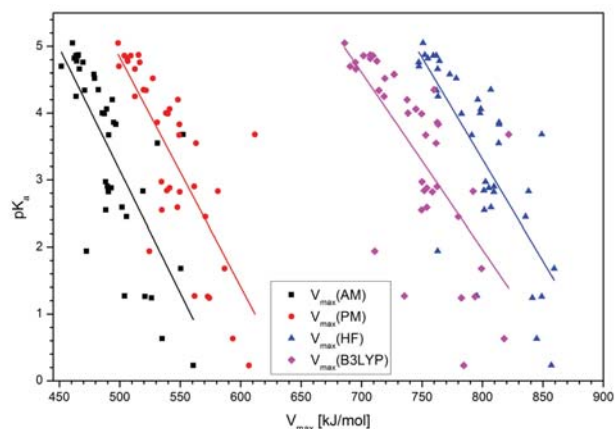


Figure 3. The correlation between the V_{\max} calculated using different methods at IsoValue 0.02, and pK_a values^{16,17} of aliphatic acids.

d) The results for the correlation between V_{\max} and pK_a for the benzoic acids studied is shown in Figure 4. As seen, the values of V_{\max} and pK_a for the set of mono-substituted benzoic acids are poorly correlated. We speculate that the solvent effect, which was not specifically accounted for in the calculations, could be the reason. An evident deviation from linearity can be observed in values for *p*-methoxybenzoic, which could be due to the stronger reso-

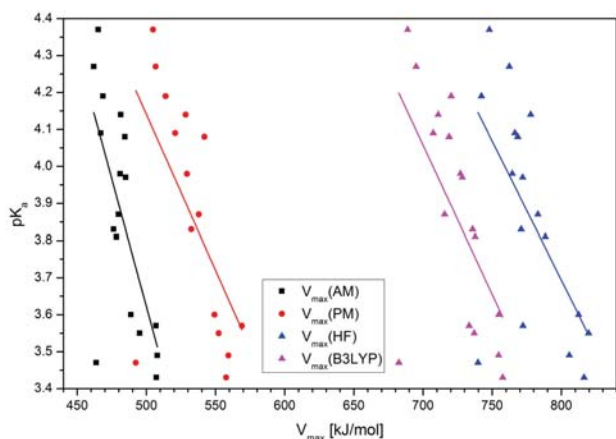


Figure 4. The correlation between the V_{\max} calculated using different methods at IsoValue 0.02, and pK_a values^{16,17} of benzoic acids.

nance effect from substituent in *para* position influencing the MEP in a non-linear fashion corresponding to pK_a . The relatively large scatter in the studied pK_a range, therefore, suggests a non-linear response of the V_{\max} to the electronic effects of some substituents.

As evident from the Figure 5, the obtained linear correlations for the applied molecules can be classified into three groups. By incorporating additional parameters, it should be possible to construct a single model to predict pK_a values for all four molecule sets.

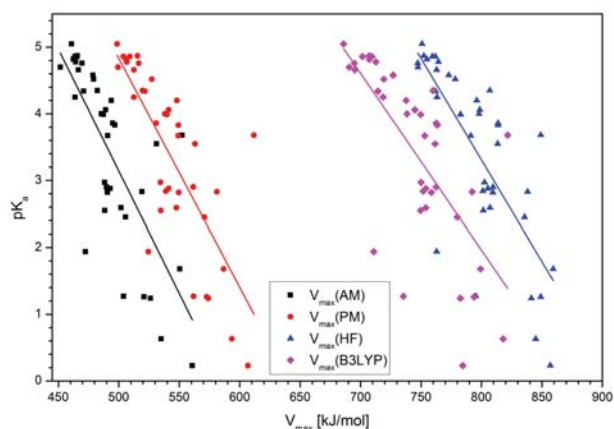


Figure 5. Experimental pK_a values^{16,17} as a function of V_{\max} (AM1) for all molecule sets.

e) To improve the obtained correlations we made an attempt to include some software-generated descriptors into the multivariable analysis. Using CANVAS, we determined ligfilter descriptors to be the parameters of choice. The inclusion of a number of carboxylic acids or carboxylates (N_{COO}) and a number of quaternary ammonium groups ($N_{\text{NH4+}}$) together with calculated V_{\max} provides the possibility of unifying the correlation between chemical descriptors and pK_a for different types of molecules (Figure 6).

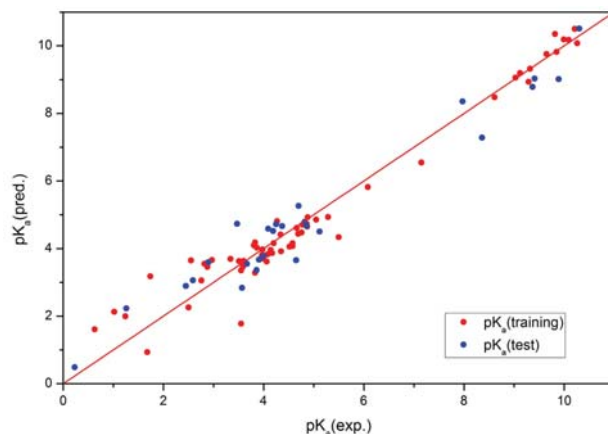


Figure 6. Predicted versus experimental pK_a values^{16,17} for training and testing sets of molecules.

Some deviations are present in the case of dicarboxylic aliphatic acids. As a result, they were excluded from the set for model construction. Using a more complicated calculation method does not notably improve the correlation. R^2 has the same value (0.96) for the models with V_{\max} from AM1, HF or B3LYP methods. A slight distinction between models can be seen from the quality of prediction for a test set of molecules. The Q^2 parameter for the models with V_{\max} from AM1, HF or B3LYP methods has the values 0.95, 0.92 and 0.92, respectively. Calculated parameters and the equation of the model for the simplest computational method are given in Table 1. The data for models obtained via B3LYP, and HF method is presented in Table S2 (see Supporting Information) for comparison. The implemented descriptors were used as a correction factor to enable the construction of a single model for different sets of compounds.

Table 1. Coefficients of the model for pK_a prediction with V_{\max} (AM1) and its statistical parameters.

Model	Coefficient	Standard deviation
N_0	30.2	± 1.5
V_{\max} (AM1)	-0.0438	± 0.0032
N_{COO}	-5.16	± 0.18
$N_{\text{NH}_4^+}$	6.18	± 0.59

$$pK_{a(\text{pred.})} = -0.0438 V_{\max} - 5.16 N_{\text{COO}} + 6.18 N_{\text{NH}_4^+} + 30.2$$

Set of compounds	Validation parameter
training	$R^2 = 0.96$
testing	$Q^2 = 0.95$

4. Conclusions

Good correlations were obtained between the maxima of molecular electrostatic potential and pK_a values for anilines and phenols. With the inclusion of certain physico-chemical descriptors, it was possible to unify the correlation between chemical descriptors and pK_a for different types of molecules with a single model function. The correlation is satisfying even with the simplest quantum mechanical calculation methods (semi-empirical method AM1).

Povzetek

V članku smo predstavili linearno zvezo med izračunanimi vrednostmi molekulskega elektrostatskega potenciala (MEP) in eksperimentalnimi pK_a vrednostmi. Obravnavali smo sete različnih molekul (karboksilne kisline, fenole in aniline), za katere smo dobili dobre linearne korelacije z maksimumi MEP. Z uvedbo avtomatično generiranih molekularskih deskriptorjev, natančneje števila karboksilnih skupin in števila amino skupin, smo uspeli sestaviti enoten triparametrski model za izračun pK_a , ki je zajemal vse obravnavane molekule.

5. References

- J. S. Murray, K. Sen (Eds.): Molecular electrostatic potentials: concepts and applications, Elsevier, Amsterdam, Netherlands, **1996**.
- P. Politzer, D. G. Truhlar (Eds.): Chemical application of atomic and molecular electrostatic potentials, Plenum, New York, **1981**.
- E. Scrocco, J. Tomasi, *Adv. Quantum Chem.* **1978**, *11*, 115–193. [https://doi.org/10.1016/S0065-3276\(08\)60236-1](https://doi.org/10.1016/S0065-3276(08)60236-1)
- P. Sjöberg, P. Politzer, *J. Phys. Chem.* **1990**, *94*, 3959–3961. <https://doi.org/10.1021/j100373a017>
- P. Politzer, P. R. Laurence, K. Jayasuriya, *Environ. Health Persp.* **1985**, *61*, 191–202. <https://doi.org/10.1289/ehp.8561191>
- U. A. Chaudry, P. L. A. Popelier, *J. Org. Chem.* **2004**, *69*, 233–241. <https://doi.org/10.1021/jo0347415>
- J. Cerar, Č. Podlipnik, *Acta Chim. Slov.* **2008**, *55*, 999–1008. <https://doi.org/10.1063/1.3251124>
- Liu, S.; Schauer, C. K.; Pedersen, L.G. *J. Chem. Phys.* **2009**, *131*, 164107.
- Milletti, F.; Storchi, L.; Sforza, G.; Cruciani, G. *J. Chem. Inf. Model.* **2007**, *47*, 2172–2181. <https://doi.org/10.1021/ci700018y>
- Xing, Li.; Glen, R. C. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 796–805. <https://doi.org/10.1021/ci010315d>
- Parthasarathi, R.; Padmanabhan, J.; Elango, M.; Chitra, K.; Subramanian, V.; Chattaraj, P. K. *J. Phys. Chem. A* **2006**, *110*, 6540–6544. <https://doi.org/10.1021/jp055849m>
- S. Liu, L. G. Pedersen, *J. Phys. Chem. A* **2009**, *113*, 3648–3655. <https://doi.org/10.1021/jp811250r>
- Y. Ma, K. C. Gross, C. A. Hollingsworth–239.
- T. A. Halgren, *J. Comp. Chem.* **1996**, *17*, 490–519. [https://doi.org/10.1002/\(SICI\)1096-987X\(199604\)17:5/6<490::AID-JCC1>3.0.CO;2-P](https://doi.org/10.1002/(SICI)1096-987X(199604)17:5/6<490::AID-JCC1>3.0.CO;2-P)
- A. R. Leach: Molecular Modelling, principles and applications; Pearson Education Limited, Essex, United Kingdom, **2001**.
- A. Albert, E. P. Serjeant: Ionization constants of acids and bases, Methuen, London, United Kingdom, **1962**.
- D. D. Perrin, (Ed.): Dissociation constants of organic bases in aqueous solution, Butterworths, London, United Kingdom, **1965**.
- W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery: Numerical recipes in C, Cambridge University Press, Cambridge, United Kingdom, **1992**, pp 680–706.
- P. J. M. van Laarhoven, E. H. L. Aarts (Eds.): Simulated annealing: theory and application, Springer Science and Business Media. Dordrecht, Netherlands. **1987**.