A STATISTICAL COMPARATIVE STUDY BETWEEN THE FUNCTIONS OF ALCOHOLS AND THIOLS USING QUANTUM CHEMISTRY METHODS

Marina Alexandra Tudoran*, Zoltan Szabadai

Renewable Energy and Electrochemistry Department, National Institute of R&D for Electrochemistry and Condensed Matter - INCEMC,
144 Aurel Păunescu-Podeanu, RO-300569, Timișoara, Romania
e-mail: alexandra.m.tudoran@gmail.com

Abstract
Using the QSAR method, the present work aims to comparatively study the relationship between solubility (data from literature) and several physical and chemical parameters (computationally calculated) for two classes of analogue compounds, i.e. alcohols and their corresponding thiols. The obtained results show that there are significant differences between the ways in which solubility can be computationally determined for these two classes.

Introduction
Alcohols represent a class of organic compounds in which a hydroxyl group is bonded to a carbon atom sp^3 hybridized. They are quite soluble in water up to 4 carbon atoms, both water solubility and their hydrophilic nature decreasing with increasing the hydrophobic alkyl group [1]. Thiols represents the sulphur analogues of alcohols which contain an SH group instead of the OH group, having two atoms and two lone pairs of electrons surrounding the atoms. Thiols are unable to form intermolecular hydrogen bonds, but they can form strong complexes with heavy metallic cathions [2, 3, 4]. In general terms, solubility represents the solute quantity which can be dissolved in a specific solvent in a given condition, characterizing a unionized substance dissolution in water [5, 6]. When a molecule in gas phase is transferred to aqueous environment, a change in energy occurs, leading to the premise that solubility is an important parameter in modulating molecular properties (e.g. redox potential, pK_a, thermodynamic stability, etc.) [7]. At the same time, water solubility can be explained based on the polarity of molecules, and has an important role in establishing the compound tendency to partition from a phase to another and in determining the distribution and the transport of molecules in environment. There are different ways of expressing the water solubility, either as gram per liter, either as mole per liter (noted with S) and conventionally converted in logarithmic values [5, 6, 8, 9, 10]. As a chemical descriptor, water solubility can be utilized in establishing the bio-concentration in aquatic organisms [10]. Thus it is of a fundamental importance in assessing the risk of molecules and in the process of design and synthesis of agro-chemicals (e.g. pesticides) [6, 10]. Although the experimental approach of determining the solubility returns very precise values, it is inefficient regarding the time and cost when a large data base of molecules is considered. A good alternative is represented by the in silico models, most of them based on the relation established between solubility and the chemical structure of the molecules through quantitative structure-properties relationships (QSPR) or quantitative structure-activity relationships (QSAR). These type of relations are established using simple or multiple linear regression procedure allowing different physical, chemical and structural descriptors to be part in the modelling process [6, 11].
The present study investigates comparatively the relationship between solubility (expressed in two different forms) and several physical and chemical descriptors (hydration energy, surface area and a new computed descriptor, generically named “topo-energy index”) generating the prediction equations for a series of alcohols and their corresponding thiols.

Experimental

The structures of molecules, considered as the working set, were generated by HyperChem 8.0.10 software [12] and pre-optimized using the MM+ force field. Then, the optimized structures were obtained by ab initio method with 6-31G** basis set. The RHF spin pairing was selected with 1e-008 as the accelerated convergence limit for SCF control. The geometry optimization was performed using the "Polak-Ribière (Conjugate gradient)" algorithm with a RMS gradient of 0.01 kcal/(Å mol). The HOMO distribution of the optimized molecular structure was obtained by plotting the 2D contours (carbon – cyan, hydrogen – white, oxygen – red, sulphur – yellow, positive isosurface – green wire mesh, and negative isosurface – violet wire mesh). Using the results obtained from HyperChem the Topo-energy Index was computed as:

$$\text{Topo-Energy Index} = \text{Hydration energy} / \text{Surface area}$$

The relationship between the response variable (solubility) and the explanatory variables (hydration energy, surface area and topo-energy index) was modeled using simple and multiple regression, using IBM SPSS program. For all the data in study the goodness of fit of normal distribution was assessed by applying Kolmogorov-Smirnov One - Sample Test.

Results and discussion

The HOMO orbitals of the optimized structures are presented in Figure 1. The ab initio results for hydration energy, surface area and the computed topo-energy index, and the solubility parameters taken from literatures [13] are given in Table 1. The Kolmogorov-Smirnov test for alcohols and thiols indicates that the *Exact. Sig* (2-tailed) > 0.05 for all of the parameters in study providing no evidence against the null hypothesis of normal distribution.

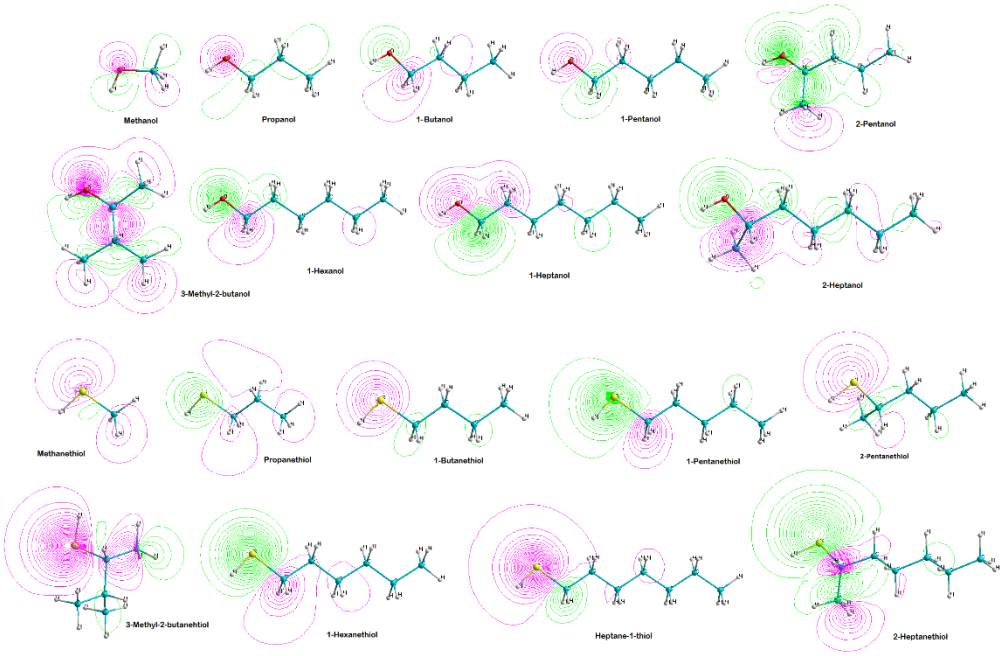

Figure 1. Electronic distribution of HOMO orbitals – alcohols (ball-and-stick model)

To determine whether hydration energy and surface area correlate with log S in case of alcohols, a Pearson test was conducted. There was a significant negative correlation between hydration energy (M = -3.934, SD = 1.611) and log S (M = -0.278, SD = 0.986), r = -0.773, $p < 0.05$, and a strong negative correlation between surface area (M = 271.610, SD = 58.871) and log S, r = -0.970, $p < 0.01$. A multiple regression was carried out to investigate whether hydration energy

and surface area could significantly predict log S. The results of the regression indicated that the model explained 97.3% of the variance and was a significant predictor of log S, $F(1, 6) = 109.347$, $p < 0.01$. Both variables contributed significantly to the model (for hydration energy: B = 0.235, $p < 0.05$, and for surface area: B = -0.022, $p < 0.01$). The final predictive model was:

$$\log S_{alcohol} = 6.604 + (0.235) \times \text{hydration energy} + (-0.022) \times \text{surface area}$$

In order to investigate how well the newly created parameter (i.e. the topo-energy index) predicts the log S value of alcohols, correlation and simple regression were conducted. There was a significant negative correlation between topo-energy index (M = -0.0167, SD = 0.013) and log S, r = -0.773, $p < 0.05$. It was found a significant regression equation (R^2 = 0.598, $F(1, 7) = 10.408$, $p < 0.05$), of the form:

$$\log S_{alcohol} = -1.255 + (-58.374) \times \text{topo-energy index}_{OH}$$

Topo-energy index significantly contributed to the model (B = -58.374, $p < 0.05$).

Table 1. Descriptors of alcohols and thiols

Molecule	**HE (kcal/mol)**	**SA (Å2)**	**TEI**	**logS (mol/L)**	**WS (g/L)**
METHANOL	-7.64	154.88	-0.0493285	1.21	519
PROPANOL	-4.90	221.76	-0.022096	0.81	391
1-BUTANOL	-4.27	250.49	-0.0170466	0.33	158
1-PENTANOL	-3.91	282.26	-0.0138525	-0.37	37.2
2-PENTANOL	-2.91	276.11	-0.0105393	-0.17	59.7
3-METHYL-2-BUTANOL	-2.85	267.03	-0.010673	-0.11	69.2
1-HEXANOL	-3.55	311.84	-0.011384	-1.1	8.22
1-HEPTANOL	-3.18	343.64	-0.00925387	-1.7	2.32
2-HEPTANOL	-2.19	336.48	-0.00650856	-1.4	4.54
METHANETHIOL	-1.41	178.51	-0.00789872	-0.38	20
PROPANETHIOL	-0.12	239.64	-0.000500751	-1.2	4.99
1-BUTANETHIOL	0.32	273.22	0.00117122	-1.5	3.06
1-PENTANETHIOL	0.68	301.84	0.00225285	-2.2	0.62
2-PENTANETHIOL	0.85	292.18	0.00290917	-2.2	0.64
3-METHYL-2-BUTANETHIOL	1	279.20	0.00358166	-2.2	0.61
1-HEXANETHIOL	1.05	334.20	0.00314183	-2.9	0.16
HEPTANE-1-THIOL	1.42	363.25	0.00390915	-3.4	0.056
2-HEPTANETHIOL	1.59	353.08	0.00450323	-3.3	0.061

Following the same procedure, we investigated the relationship between log S and hydration energy and surface area in case of thiols. There was a strong and negative relationship between log S (M = -2.142, SD = 0.995) and hydration energy (M = 0.598, SD = 0.917), r = -0.955, $p < 0.01$, and also between log S and surface area (M = 290.569, SD = 57.884), r = -0.985, $p < 0.01$. The obtained R^2 = 0.973 indicated that 97.3% of the variance in log S is due to the influence of hydration energy and surface area, the effect size being large. The results of ANOVA were significant $F(1, 6) = 106.420$, $p < 0.01$. While surface area contributed significantly to the model (B = -0.014, $p < 0.05$), hydration energy did not (B = -0.187, $p = 0.471$). The obtained equation for the regression line was:

$$\log S_{thiols} = 2.066 + (-0.187) \times \text{hydration energy} + (-0.014) \times \text{surface area}$$

Topo-energy index (M = -0.002, SD = 0.004) calculated for thiols was strong and negative correlated with log S, r = -0.877, $p < 0.01$, according to Pearson correlation analysis. In this case, regression analysis showed that there was 76.9% of the variance explained by the model and the model is significant, $F(1, 7) = 23.253$, $p < 0.05$. The regression equation was:

$$\log S_{thiol} = -1.811 + (-228.154) \times \text{topo-energy index}_{SH}$$

Also in this case, topo-energy index (B = -228.154, $p < 0.05$) makes a significant contribution in predicting log S.

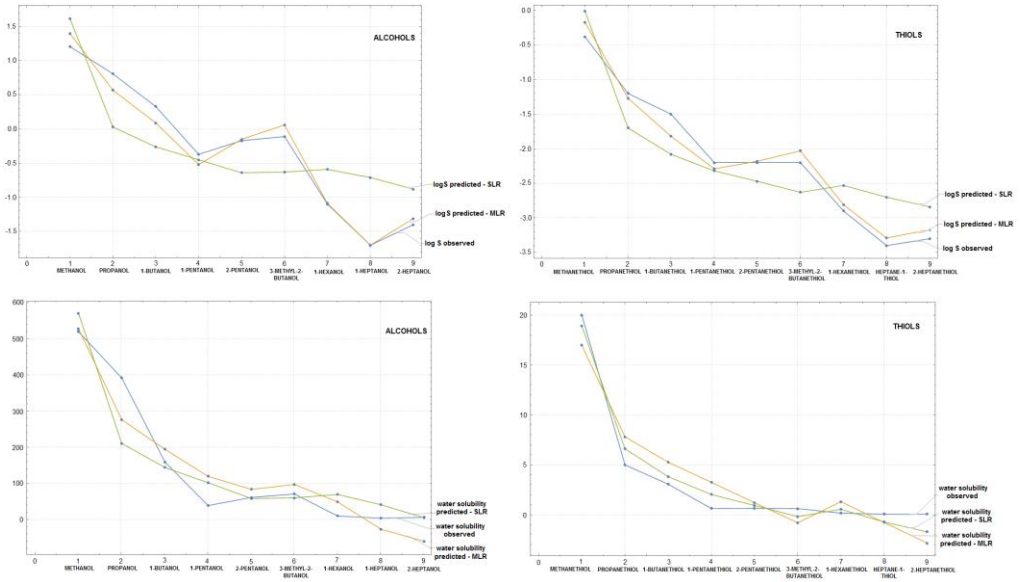

Figure 2. - Observed vs Predicted values of Log S and Water solubility for alcohols and thiols

Comparatively, we wanted to determine the influence of hydration energy, surface area and topo-energy index on water solubility for the same alcohols and thiols. In case of alcohols, Pearson correlation coefficients indicated that there is a strong negative relationship between water solubility (M = 138.798, SD = 188.299) and hydration energy, r = -0.909, $p < 0.01$, and surface area respectively, r = -0.928, $p < 0.01$. According to the results of the linear regression, hydration energy and surface area explained 89.7% of the variance, the model being significant ($F(1, 6) = 26.009$, $p < 0.05$, $R^2 = 0.897$). The hydration energy (B = -47.398, p = .199) and the surface area (B = -1.821, p = .089) have a similar contribution in predicting water solubility. The regression equation was:

$$\text{Water solubility}_{alcohol} = 446.909 + (-47.398) \times \text{hydration energy} + (-1.821) \times \text{surface area}$$

As for the topo-energy index, Pearson correlation coefficient indicated a strong and negative relationship between variables, r = -0.918, $p < 0.01$. It was found a significant regression equation (R^2 = 0.842, $F(1, 7) = 37.301$, $p < 0.01$), with topo-energy index being a significant predictor of the model (B = -13233.415, $p < 0.01$). The obtained equation was:

$$\text{Water solubility}_{alcohol} = -82.762 + (-13233.415) \times \text{topo-energy index}_{OH}$$

In case of thiols, there was a strong and negative correlation between water solubility (M = 3.355, SD = 6.463) and hydration energy, r = -0.927, $p < 0.001$, and surface area respectively, r = -.846, $p < 0.01$. The results of the regression indicated that there was a collective significant effect between predictors and water solubility (R^2 = 0.876, $F(2, 6) = 21.193$, $p < 0.05$). It was found that while hydration energy significantly predicted water solubility (B = -9.343, $p < 0.05$), surface area did not (B = 0.047, p = 0.415). The final predictive model was:

$$\text{Water solubility}_{thiol} = -4.588 + (-9.343) \times \text{hydration energy} + (0.047) \times \text{surface area}$$

In the same way, topo-energy index and water solubility (M = 3.356, SD = 6.463) of thiols were strong and negative correlated, r = -0.983, $p < 0.01$. The results of the regression indicated that the model explained 96.7% of the variance and was significant, $F(1, 7) = 205.550$, $p < 0.001$. The final predictive model was:

$$\text{Water solubility}_{thiol} = 5.770 + (-1663.090) \times \text{topo-energy index}_{SH}$$

It was found that topo-energy index significantly predicted water solubility (B = -1663.09, $p < 0.01$). Figure 2 presents comparatively the observed values and the values obtained using the prediction equations for log S and water solubility for alcohols and thiols.

Conclusion

We used linear regression analysis in order to construct a prediction equation that permits the estimation of log S and water solubility based on the knowledge of the hydration energy, surface area (as coefficients in multiple regression) and a new created topo-energy index (as coefficient in single regression), the dependent variables being significantly correlated with all predictors. In case of log S, we determined that for alcohols molecules, its values can be better predicted using multiple regression, with surface area having a higher contribution in the prediction equation compared to hydration energy. The same direction was maintained also in case of thiols, with the specification that in the prediction equation only the surface area had a significant contribution.

In case of water solubility of alcohols, although R^2 values were very close for both single and multiple regression, we determined that hydration energy and surface area combined are better predictors compared to topo-energy index. However, only surface area is significant in the prediction equation. We obtained a different result in case of water solubility of thiols, in this case topo-energy index being a better and a highly significant predictor.

Acknowledgements

This work was supported by the National Program NUCLEU PN 19 22 "Unpollutant and innovative technologies for health, environment protection and energetic efficiency", Project Code PN 19 22 01 01, Contract No. 40N/2019 "Advanced technologies for materials dedicated to the energy sector".

References

[1] D.R. Bloch, Organic Chemistry Demystified. McGraw-Hill, (2006), pp. 295.
[2] D. Klein, Organic Chemistry, Wiley, (2017), pp. 583.
[3] P.Y. Bruice, Organic Chemistry, Pearson Education, Inc., (2016), pp. 494.
[4] J.G. Smith, Organic Chemistry, McGraw-Hill Education, (2017), pp. 364.
[5] S.D. Sarker, L. Nahar, Chemistry for Pharmacy Students, Wiley, (2007), pp. 1-15.
[6] J. Ali, P. Camilleri, M.B. Brown, A.J. Hutt, S.B. Kirton, J. Chem. Inf. Model., 52 (2012) 420−428.
[7] A. Pomogaeva, D.M. Chipman, J. Chem. Theory Comput., 10 (2014) 211−219
[8] V. Gelmboldt, L. Ognichenko, I. Shyshkin, V. Kuzmin, Struct. Chem., 32 (2021) 309-319.
[9] J. Dai, B. Sun, A. Zhang, K. Lin, L. Wang, Bull. Environ. Contam. Toxicol. 61 (1998) 591-599
[10] H. Birch, A.D. Redman, D.J. Letinski, D.Y. Lyon, P. Mayer, Anal. Chim. Acta, 1086 (2019) 16-28
[11] A.A. Toropov, A.P. Toropova, M. Marzo, E. Benfenati Use of the index of ideality of correlation to improve aquatic solubility model. J. Mol. Graph. Model., 96 (2020) 107525
[12] HyperChem™ Professional, Hypercube, Inc., 1115 NW 4th Street, Gainesville, Florida 32601, USA.
[13] https://hmdb.ca/