QSAR ANALYSIS OF TOXICITY FOR A SERIES OF AROMATIC COMPOUNDS

Marina Alexandra Tudoran, Zoltan Szabadai

National Institute of R&D for Electrochemistry and Condensed Matter - INCEMC,
144 Aurel Păunescu-Podeanu, RO-300569, Timişoara, Romania
e-mail: r_alliy@yahoo.com

Abstract
Aromatic compounds are particularly interesting to scientists due to environmental and human health concerns. QSAR models are increasingly used to investigate the relationship between the toxicity of a pollutant and its structural properties, on the assumption that pollutants similar in structure and physicochemical properties are likely to be similar in toxicity. Based on the above mentioned, this paper investigates the relationship between electron affinity, ionization potential, and the toxicity of some aromatic compounds. Correlation and linear regression are used as QSAR analysis methods.

Introduction
Polycyclic aromatic hydrocarbons (PAHs) are aromatic compounds consisting of carbon and hydrogen atoms, highly stable, hydrophobic and lipophilic, characterised by their property of bioaccumulating, and considered harmful pollutants for the environment. Depending on the number of aromatic rings, PAHs are classified into two groups: low molecular weight compounds with less than four aromatic rings, and high molecular weight compounds with four or more aromatic rings. Studies have shown that polycyclic aromatic hydrocarbons have different environmental impacts depending on their molecular weight. Thus, while low molecular weight molecules are volatilised or easily decomposed in the environment, high molecular weight molecules with a large number of benzene rings are more chemically stable, exhibit high toxicity and persistence in the environment, thus remaining in soil, sediment, and water for a longer period of time [1-3].
The number and orientation of the benzene rings also affect the physical and chemical properties of PAH molecules, which in turn affect their mobility and distribution between different environmental compartments. Acute toxicity, which increases with increasing molecular size (up to 4-5 benzene rings), has also been shown to be influenced by molecular structure [2, 4, 5]. Predicting the toxicity of chemical compounds, however, remains a major challenge for the scientific community. Defined as the extent to which a chemical compound (or mixture) can harm, injure or kill a living organism, toxicity can be estimated using a range of sophisticated in vitro or in vivo experimental techniques. Because these techniques are expensive, time-consuming and raise ethical concerns (by using of animals or animal tissues), researchers have developed in silico (computer-based) methods to address these issues. One of the most widely used in silico methods is the quantitative structure-activity/property relationship (QSAR/QSPR) method, which is based on the assumption that chemical molecules that are similar in structure will have similar activities and properties. Thus, based on the relationship between chemical structures and activities/properties of known compounds, models can be constructed to predict the activities/properties of unknown (but structurally similar) chemicals using multivariate statistical techniques [6].

Starting from these considerations, this current work is intended to presents the relationship between LC_{50} (data from literature), electron affinity and ionization potential respectively

(computationally calculated) for a series of aromatic compounds using QSAR analysis. The obtained results show that the toxicity of benzene and linear PAHs is different from the toxicity of angular and clustered PAHs.

Experimental

The molecules from the working set were generated by HyperChem 8.0.10 software [7] and pre-optimized using the MM+ force field. Then, using Gaussian View 6.0.16 and Gaussian 3.0 (DFT 3-21G, B3LYP method), we generate the optimized structures and calculate the energy. The electron affinity (*EA*), considered as the energy difference resulted by an electron addition to a neutral molecule, was calculated with the formula: $EA = (E_{Neutral} - E_{Anion})$, with $E_{Neutral}$ – the energy of neutral form, and E_{Anion} – the energy of anion. The ionization potential (*IP*), considered as the energy necessary for an electron removal from a neutral molecule, was calculated with the formula: $IP = (E_{Cation} - E_{Neutral})$, with E_{Cation} – the energy of cation [8]. The distribution type of the data was determined using Kolmogorov-Smirnov Test. The association between LC_{50} (*Daphnia magna*) – as dependent variable, and the ionization potential and electron affinity – as independent variables, was assessed using Pearson correlation (*2-tailed*). Linear regression was computed in order to determine the line of the best fit for the data. All statistical calculations were performed using Origin 9 program.

Results and discussion

The optimized structures of the aromatic compounds are presented in Figure 1. The physical-chemical parameters computed using Gaussian program and the LC_{50} computed with ADMETlab 2.0 [9] are given in Table 1. For all of the parameters in study the Kolmogorov-Smirnov test returned $p > 0.05$ (0.39477 for electron affinity, 0.71116 for ionization potential, and 0.13844 for LC_{50}), meaning that the data is normally distributed.

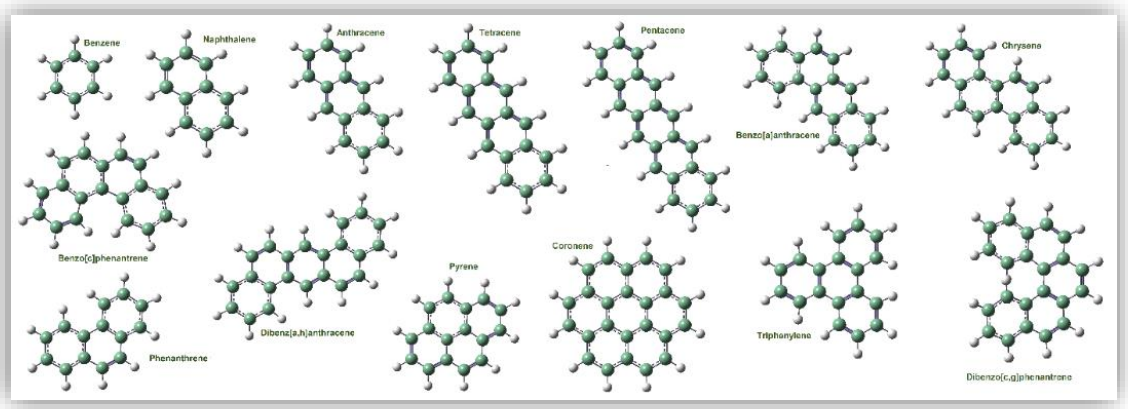

Figure 1. Optimized structures of neutral form of the molecules in study (ball-and-stick model)

According to the Pearson correlation results, there was a significant correlation in positive direction between LC_{50} and electron affinity ($r = 0.782$, $p < 0.01$), and a significant correlation in negative direction between LC_{50} and ionization potential ($r = -0.779$, $p < 0.01$).

Table 1. Values of electron affinity, ionization potential, and LC_{50} for the molecules in study

NAME	EA (A.U.)	IP (A.U.)	LC_{50} (*Daphia magna*)

Benzene	-0.078	0.335	3.378
Naphthalene	-0.029	0.285	4.597
Anthracene	0.003	0.253	5.256
Tetracene	0.026	0.232	5.688
Pentacene	0.042	0.217	6.085
Phenanthrene	-0.017	0.272	5.742
Benzo[c]phenanthrene	-0.001	0.264	6.060
Dibenzo[c,g]phenantrene	-0.001	0.259	6.465
Benzo[a]anthracene	0.008	0.250	5.945
Coronene	0.009	0.253	6.092
Pyrene	-0.003	0.258	5.924
Chrysene	-0.001	0.258	6.060
Triphenylene	-0.024	0.279	5.890
Dibenz[a,h]anthracene	0.009	0.248	6.151

The relationship between electron affinity and LC_{50} of aromatic molecules is described by equation (1), and plotted in Figure 2a.

$$\textbf{\textit{MODEL 1}}: LC_{50} = 5.75719 + 22.23965 \times \mathrm{EA} \quad (1)$$

In Figure 2a, two areas of distribution of molecules in the working set can be seen. Benzene and PAHs molecules with linear structure follow a linear distribution, while PAHs molecules with angular and clustered structure follow a growth/sigmoidal (MichaelisMenten) type distribution.

Table 2. Summary of simple linear regression with LC_{50} (*Daphnia magna*) as dependent variable and electron affinity (***MODEL 1***) and ionization potential (***MODEL 2***) respectively as independent variables (N = 14).

	Intercept		**Slope**		**Statistics**
	Value	*SE*	*Value*	*SE*	*Adj. R^2*
MODEL 1	5.75719	0.14002	22.23965	5.11828	0.57902
MODEL 2	11.61867	1.39023	-22.74868	5.28673	0.57399

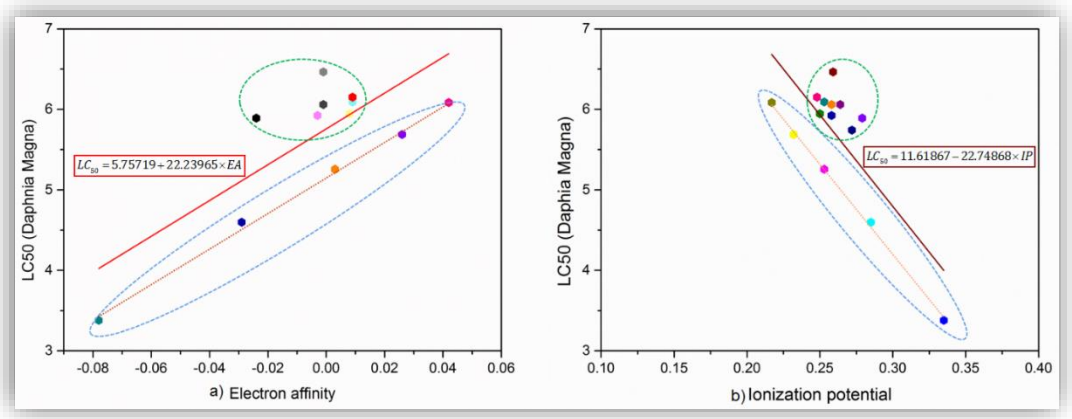

Figure 2. LC_{50} values of the studied polycyclic aromatic hydrocarbons as a function of a) electron affinity (***MODEL 1***) and b) ionization potential (***MODEL 2***).

The relationship between electron affinity and LC_{50} of aromatic molecules is described by equation (2), and plotted in Figure 2b.

$$\textbf{\textit{MODEL 2}}: LC_{50} = 11.61867 - 22.74868 \times \mathrm{IP} \quad (2)$$

In Figure 2b, the two areas of distribution are maintained as in privous case. Thus, benzene and PAHs molecules with linear structure follow a linear distribution, while PAHs molecules with angular and clustered structure follow a distribution described by the ratio of two polynomial functions (Rational0).

Conclusion

Linear regression was employed to produce the best-fit line through which the electron affinity and ionization potential of a series of PAHs molecules could be related to their LC_{50} values. For the associated regression lines, close values of the Adj.-R^2 coefficient were obtained (0.57902 for electronic affinity and 0.57399 for ionization potential, respectively), implying that the toxicity of an unknown aromatic compound with similar structure as the working set can be predicted with about 60% accuracy if either electronic affinity or ionization potential values are known. In addition, it was found that, in the case of the studied set, the association between toxicity and electron affinity and the ionization potential is strongly influenced by the way the benzene nuclei are arranged in molecules. Thus, in terms of the manifestation of toxicity in relation to the studied parameters, in case of molecules with angular and clustered structure, other variables play a role which in case of linear molecules either play no role or have a constant value.

References

[1] H. Ha, K. Park, G. Kang, S. Lee, Ecotoxicology 28 (2019) 333.
[2] H.I. Eldos, N. Zouari, S. Saeed, M.A. Al-Ghouti,Arab. J. Chem. 15(7) (2022) 103918.
[3] M. Honda, N. Suzuki, IJERPH 17(4) (2020) 1363.
[4] A. Alalaiwe, Y.K. Lin, C.H. Lin, P.W. Wang, J.Y. Lin, J.Y. Fang, Chemosphere 255 (2020) 126955.
[5] M.J. Munoz, J.V. Tarazona, Bull. Environ. Contam. Toxicol. 50 (1993) 363.
[6] A. Karim, V. Riahi, A. Mishra, M.H. Newton, A. Dehzangi, T. Balle, A. Sattar, Acs Omega, 6(18), (2021) 12306.
[7] HyperChem™ Professional, Hypercube, Inc., 1115 NW 4th Street, Gainesville, Florida 32601, USA.
[8] J.B. Foresman, Æ. Frisch, Exploring Chemistry With Electronic Structure Methods: A Guide to Using Gaussian 2nd Edition, Gaussian, Pittsburg, 1996, pp. 142-143.
[9] https://admetmesh.scbdd.com/service/evaluation/cal