STATISTICAL ANALYSIS METHODS IN EXPLAINING THE RELATIONSHIP BETWEEN THE BIOCONCENTRATION FACTOR AND MOLECULAR DESCRIPTORS - APPLICATIONS TO SMALL POLYCYCLIC AROMATIC HYDROCARBONS

Marina Alexandra Tudoran, Ioan Taranu, Bogdan-Ovidiu Taranu

National Institute of Research and Development for Electrochemistry and Condensed Matter, Dr. Aurel Paunescu Podeanu Street No. 144, 300569 Timisoara, Romania
e-mail: r_alliy@yahoo.com

Abstract
The investigation of the chemical properties and potential environmental behavior of toxic compounds represents a subject of significant interest for researchers, as it allows them to develop strategies for protecting both human health and the integrity of the natural environment. In the present study, computational and statistical methods are employed to explore the relationship between the bioconcentration factor and surface area, polarizability, hydration energy, HOMO and LUMO energies, and the HOMO-LUMO gap, in a set of polycyclic aromatic hydrocarbon (PAH) molecules. This approach can be used as an alternative to existing QSAR methods to facilitate the assessment of bioaccumulation factors for a wider range of molecules, thereby extending the applicability of these existing procedures.

Introduction
The bioconcentration factor (BCF) can be defined as the ratio between the equilibrium concentration of a chemical in aquatic organisms and the corresponding concentration of the chemical dissolved in the surrounding aqueous environment: $\boldsymbol{BCF} = C_f / C_w$. In the calculation formula, C_f (mg/kg) represents the steady-state concentration of the chemical within the organism (or the tissue) in question, while C_w (mg/L) represents the steady-state concentration of the chemical in the aquatic environment in which the organism is located [1,2]. The BCF is an estimate of a xenobiotic's tendency to concentrate and accumulate in an aquatic organism and is one of the most important criteria in both ecotoxicological assessment and hazard assessment [2,3]. The parameter most frequently employed as a preliminary screening tool to indicate bioaccumulation potential is the log P [4]. A variety of QSAR models utilize a range of theoretical molecular descriptors, encompassing molecular weight, connectivity indices, topological, geometric, or quantum-chemical descriptors, in addition to soil adsorption coefficients, fragment constants, and linear solvation energy relationships descriptors [5,6].
Polycyclic aromatic hydrocarbons (PAHs) are compounds constituted by two or more fused benzene rings, with varying structural configurations. The arrangement of the benzene rings determines the physical, chemical, and toxicological characteristics of the molecules. The behavior of PAHs in aquatic environments is contingent upon their chemical properties [7]. Numerous studies have been conducted over time to understand the distribution, fate, and environmental effects of this type of molecules, which have significant environmental and public health implications due to their presence in all environmental components, resistance to biodegradation, bioaccumulation potential, and carcinogenic activity. PAHs are considered among the first recognized environmental carcinogens that do not readily break down under natural conditions and have a persistence that increases with their molecular weight. Through environmental interactions, they may also contribute to pathogenic mechanisms such as DNA mutations, oxidative stress, protein dysfunction, or mitochondrial damage [8-10]. In order to better understand the impact of these compounds on human health and ecosystems, further

research into the mechanisms and models of PAH toxicity remains of paramount importance. The long-term goal is to develop more effective environmental protection strategies and implement appropriate measures to reduce the potential hazards posed by aromatic compounds [11].

In light of these considerations, the present study aims to evaluate the relationship between the toxicity (expressed as the bioconcentration factor) of a number of selected PAH molecules and several physical-chemical parameters, employing both computational and statistical techniques.

Experimental

The studied PAH molecules were drawn using HyperChem 8.0.10 software [12] and optimized using the ab initio, STO-3G basis set. Their physical-chemical parameters, generated at the same level include solvent-accessible surface area (S_A), polarizability (α), hydration energy (E_H), energy of HOMO and LUMO orbitals, and HOMO-LUMO gap (ΔE_{H-L}). The bioconcentration factor (BCF) was considered as a toxicity parameter and computed with ADMETlab 3.0 [13]. Statistical methods were used to analyze the data: Kolmogorov-Smirnov Test, Pearson correlation, and simple linear regression. All statistical calculations were performed using IBM SPSS Statistics 21 software.

Results and discussion

The PAH molecules in the working set are classified into four categories: linear catacondensates (*LC*), angular catacondensates (*AC*), angular pericondensates (*AP*), and cluster pericondensates (*PC*). Their surface area ranges from 341.37 to 553.95 (M = 466.056, SD = 60.860), the polarizability values range from 21.15 to 41.34 (M = 34.098, SD = 6.117), while the hydration energy values range from -3.89 to -1.63 (M = -3.159, SD = 0.5366). The highest recorded bioconcentration factor value (M = 2.386, SD = 0.265) is observed in the case of phenanthrene (3.091), while the lowest is observed for benzo(ghi)perylene (2.085). The physical-chemical and toxicity parameters are given in Table 1. Figure 1 shows the HOMO and LUMO distributions for several molecules in the working set belonging to the four classes.

Table 1. Molecular descriptors and the toxicity parameter of the selected PAHs

Name	Type	SA	α	E_H	E_{HOMO}	E_{LUMO}	ΔE_{H-L}	BCF
Anthracene	LC	365.80	22.80	-2.69	-5.516	4.785	-10.301	3.069
Phenanthrene	AC	358.05	22.80	-2.59	-6.229	5.652	-11.880	3.091
Phenalene	CP	341.37	21.15	-1.63	-5.741	5.480	-11.221	2.601
Tetracene	LC	427.12	28.98	-3.09	-4.905	4.104	-9.010	2.254
Benz(a)anthracene	AC	421.50	28.98	-3.02	-6.222	5.541	-11.764	2.495
Chrysene	AC	415.03	28.98	-2.94	-5.864	5.209	-11.073	2.732
7H-Benzo(de)anthracene	AP	399.80	27.34	-2.19	-6.109	5.246	-11.355	2.973
Triphenylene	AC	421.75	28.98	-2.90	-6.209	5.609	-11.818	2.296
Pyrene	PC	375.25	25.50	-2.40	-5.506	4.787	-10.293	2.524
Pentacene	LC	491.89	35.16	-3.49	-5.472	4.871	-10.343	2.161
Benzo(a)naphthacene	AC	484.98	35.16	-3.42	-5.016	4.215	-9.232	2.270
Benzo(b)chrysene	AC	482.60	35.16	-3.37	-5.353	4.600	-9.953	2.359
Dibenz(a,h)anthracene	AC	480.00	35.16	-3.36	-5.621	4.958	-10.579	2.313
Pentaphene	AC	486.06	35.16	-3.45	-5.587	4.899	-10.486	2.203
Picene	AC	473.90	35.16	-3.30	-5.825	5.193	-11.018	2.455
Dibenz(a,c)anthracene	AC	475.08	35.16	-3.37	-5.658	4.940	-10.599	2.293
Perylene	AP	425.08	31.68	-2.75	-5.063	4.320	-9.382	2.198
Hexacene	LC	553.95	41.34	-3.89	-4.162	3.288	-7.449	2.095
Hexaphene	AC	533.12	41.34	-3.85	-5.095	4.289	-9.384	2.126

Benzo(a)pentacene	AC	551.90	41.34	-3.82	-4.580	3.727	-8.306	2.215
Dibenzo(a,j)naphthacene	AC	548.10	41.34	-3.76	-5.119	4.327	-9.446	2.362
Dibenzo(a,l)naphthacene	AC	543.31	41.34	-3.76	-5.122	4.327	-9.450	2.323
Dibenzo(a,h)pyrene	AP	488.67	37.87	-3.24	-4.787	3.996	-8.783	2.270
Dibenzo(a,e)pyrene	AP	485.57	37.87	-3.22	-5.277	4.550	-9.827	2.224
Benzo(pqr)picene	AP	490.96	37.87	-3.19	-5.158	4.410	-9.567	2.312
Naphtho(2,3-a)pyrene	AP	495.71	37.87	-3.25	-4.854	4.059	-8.912	2.376
Benzo(c)picene	AC	533.51	41.34	-3.66	-5.667	4.990	-10.657	2.407
Dibenzo(a,c)naphthacene	AC	541.06	41.34	-3.77	-5.091	4.281	-9.372	2.232
Anthanthrene	PC	447.50	34.39	-2.72	-4.774	3.940	-8.714	2.266
Benzo(ghi)perylene	AC	443.06	34.39	-2.68	-5.690	4.506	-10.197	2.085

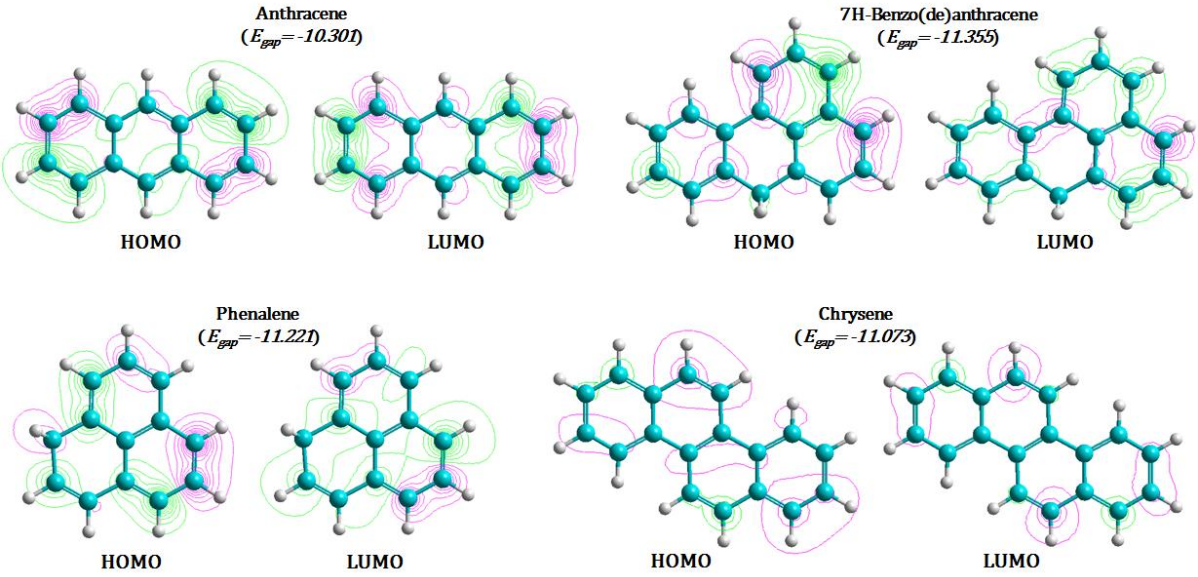

Figure 1. HOMO and LUMO distribution for several PAH molecules

The normality Kolmogorov-Smirnov Test returned $p > 0.05$ for all parameters, meaning that the data are normally distributed. As indicated by the Pearson analysis (Table 2) and following Schcher's interpretation [14], there is a moderate negative correlation observed between the bioconcentration factor and the surface area, the HOMO energy, and the HOMO-LUMO gap. Additionally, a strong negative correlation is observed between the bioconcentration factor and polarizability. Conversely, the bioconcentration factor shows a moderate positive correlation with hydration energy and LUMO energy. All correlations are statistically significant at the $p < 0.01$ level.

Table 2. Mean, standard deviation, and Pearson correlation coefficients for the studied variables ($N = 30$)

Variable	*M*	*SD*	S_A	α	E_H	E_{HOMO}	E_{LUMO}	ΔE_{H-L}
BCF	*2.386*	*0.265*	**-0.648****	**-0.709****	**0.501****	**-0.478****	**0.494****	**-0.488****
S_A	*466.05*	*60.860*	-	0.977**	-0.915**	0.550**	-0.575**	0.565**
α	*34.098*	*6.117*	-	-	-0.843**	0.551**	-0.585**	0.571**
E_H	*-3.159*	*0.536*	-	-	-	-0.436*	0.458**	-0.449**
E_{HOMO}	*-5.375*	*0.510*	-	-	-	-	-0.984**	0.996**
E_{LUMO}	*4.636*	*0.585*	-	-	-	-	-	-0.997**
ΔE_{gap}	*-10.012*	*1.089*	-	-	-	-	-	-

$^{**}p < 0.01$, $^{*}p < 0.05$

The relationship between the bioconcentration factor and the selected physical-chemical parameters was further investigated using simple linear regression, the results being presented in Table 3. The effect size was calculated using Cohen approach [15], after the formula:

$$f^2 = \frac{R^2}{1 - R^2} \quad (1)$$

Table 3. Results of simple regression analysis with the bioconcentration factor as dependent variable ($N = 30$)

Independent Variable	R^2	f^2	B	SE_B	*β*	*t*	*p*
Surface area solvent accessible	0.453	0.828	-0.003	0.001	-0.673	-4.818	< 0.001
Polarizability	0.502	1.008	-0.031	0.006	-0.708	-5.310	< 0.001
Hydration energy	0.327	0.485	0.283	0.077	0.571	3.684	0.001
HOMO energy	0.319	0.468	-0.294	0.081	-0.565	-3.621	0.001
LUMO energy	0.336	0.506	0.262	0.070	0.579	3.761	0.001
HOMO – LUMO gap	0.332	0.497	-0.140	0.038	-0.576	-3.729	0.001

The highest R^2 value was obtained for polarizability, followed by surface area; conversely, the lowest values corresponded to HOMO energy and hydration energy. Thus, polarizability accounts for approximately 50.2% of the variance of the bioconcentration factor, while surface area is responsible for 45.3% of the variation. In contrast, the HOMO energy accounts for only 31.9% of the variance, while the hydration energy accounts for 32.7%.

As indicated by the Cohen criteria [15], the effect size of hydration energy, HOMO energy, and HOMO-LUMO gap was moderate ($f^2 < 0.5$), large in the case of LUMO energy ($f^2 \approx 0.5$) and very large in the case of surface area ($f^2 \approx 0.8$). The highest value was obtained for polarizability ($f^2 = 1.008$), indicating that this parameter exerts the greatest influence on the bioconcentration factor. In contrast, the lowest effect size reported for HOMO energy indicates that this parameter exerts minimal influence on the bioconcentration factor.

Conclusion

The results of the correlation analysis indicate that an increase in the hydration energy and LUMO energies is associated with an increase in the value of the bioconcentration factor. Conversely, an increase in the polarizability, surface area, HOMO energy, and the HOMO-LUMO gap is predicted to result in a reversal of this effect. Regression analysis results indicate that polarizability may serve as an effective predictor when incorporated individually into a regression equation. Furthermore, the findings reveal that energy-related parameters contribute

to approximately 30% of the variation in the bioconcentration factor. As indicated by the effect size, it may be considered that the LUMO energy, surface area, and polarizability are the primary factors contributing to toxicity in the case of the studied molecules, meaning that these variables must be addressed in subsequent analyses. This study constitutes a preliminary investigation into the identification of the most suitable predictors for incorporation into multivariate analysis. Consequently, a novel predictive model can be developed to estimate the bioconcentration factor, thereby offering a potential alternative to existing models in literature.

References

[1] F. Lunghini, G. Marcou, P. Azam, R. Patoux, M.H. Enrici, F. Bonachera, D. Horvath, A. Varnek, *SAR QSAR* Environ. Res. 30 (2019) 507.
[2] T. Ivanciuc, O. Ivanciuc, D.J. Klein, Mol. Divers. 10 (2006) 133.
[3] P. Gramatica, E. Papa, QSAR Comb. Sci. 22 (2003) 374.
[4] A. Gissi, A. Lombardo, A. Roncaglioni, D. Gadaleta, G.F. Mangiatordi, O. Nicolotti, E. Benfenati, Environ. Res. 137 (2015) 398.
[5] C. Zhao, E. Boriani, A. Chana, A. Roncaglioni, E. Benfenati, Chemosphere 73 (2008) 1701.
[6] Y. Wang, Y., J.J., Wen, J., Li, W.C., He, L.M., Qin, Su, Y.H. Zhao, Environ. Toxicol. Pharmacol. 38 (2014) 388.
[7] H. Li, Y. Ran, Sci. World J. 2012 (2012) 632910.
[8] A.K. Haritash, C.P. Kaushik, J. Hazard. Mater. 169 (2009) 1.
[9] E. Stogiannidis, R. Laane, Rev. Environ. Contam. T. (2014) 49.
[10] M.M. Ferreira, Chemosphere 44 (2001) 125.
[11] S. Wang, C. Li, L. Zhang, Q. Chen, S. Wang, Environ. Res. (2023) 117901.
[12] HyperChem™ Professional, Hypercube, Inc., 1115 NW 4th Street, Gainesville, Florida 32601, USA.
[13] https://admetlab3.scbdd.com/server/evaluation.
[14] P. Schcher, L.A. Schwarte, Anesth. Analg. 126 (2018) 1763.
[15] J.E. Cohen, Statistical Power Analysis for the Behavioral Sciences, Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey, USA, 1988.