

The use of QSAR methods for determination of n-octanol/water partition coefficient using the example of hydroxyester HE-1

Joanna Guziałowska-Tic^{1,*}

¹Opole University of Technology, Department of Environmental Engineering, Mikołajczyka 5 Street, 45-271 Opole, Poland

Abstract. According to the Directive of the European Parliament and of the Council concerning the protection of animals used for scientific purposes, the number of experiments involving the use of animals needs to be reduced. The methods which can replace animal testing include computational prediction methods, for instance, the quantitative structure-activity relationships (QSAR). These methods are designed to find a cohesive relationship between differences in the values of the properties of molecules and the biological activity of a series of test compounds. This paper compares the results of the author's own results of examination on the n-octanol/water coefficient for the hydroxyester HE-1 with those generated by means of three models: Kowwin, MlogP, AlogP. The test results indicate that, in the case of molecular similarity, the highest determination coefficient was obtained for the model MlogP and the lowest root-mean square error was obtained for the Kowwin method. When comparing the mean logP value obtained using the QSAR models with the value resulting from the author's own experiments, it was observed that the best conformity was that recorded for the model AlogP, where relative error was 15.2%.

1 Introduction

The chemical safety assessment of products for the purpose of their registration requires tests to determine their physico-chemical, toxicological and ecotoxicological properties [1]. Unfortunately, some of those tests involve the use of animals. After the coming into force of the Directive of the European Parliament and of the Council No. 2010/63/EU of 22 September 2010 concerning the protection of animals used for scientific purposes, the number of experiments involving the use of animals has to be

reduced so as to ease their pain and suffering in connection with the tests [2].

According to the provisions of the Directive, the target is to eliminate the unnecessary tests on animals. REACH requires that the registrants should restrict new tests with the use of vertebrates for the purpose of registration and use them only as a last resort [3]. First, it is necessary to collect and assess all existing data concerning the test substance and then, to detect any data gaps and consider the possibility of filling them with the results of *in vitro/ex vivo* tests or other, alternative concepts (Fig. 1).



Fig. 1. Strategies to avoid unnecessary animal testing [3].

* Corresponding author: j.guzialowska@po.opole.pl

This means that all available data obtained from various sources, including tests made *in vivo* (on live animals), *ex vivo* (on animal tissue) and *in vitro* (with the use of cultured cells or bacteria) need to be collected. Another option is to use information based on tests involving human exposure, predictions based on available information about substances with similar structures (e.g., “cross-sectional approach” or “chemical categories”), as well as expectations based on reliable quantitative prediction methods, such as quantitative structure-activity relationships (QSAR) [5].

The QSAR methods are used, first of all, in the development and designing of new medicines. However, it is becoming an ever more frequent practice to use them for determination of the toxicological and ecotoxicological properties of chemical compounds within the REACH system [6]. The staff of the European Chemicals Agency (ECHA) have presented the use of various – not only experimental – methods for the assessment of the ecotoxicity of substances. Based on the results, it was found that only in just over 1% of the substances were the QSAR methods used for the assessment of ecotoxicological properties [7-11].

The biological activity of a chemical compound is directly related to its physical properties. Lipophilicity is the most useful parameter in the assessment of biological activity of substances and in the predicting their toxic activity. It is essential in the distribution of chemical compounds taking place in the body. Specifically, focus is on how a given substance behaves when penetrating the lipophilic membranes which separate aqueous media. Lipophilicity is described by partition processes

taking place between two phases: non-polar (organic) and polar (typically water).

The polar/non-polar phase system in biological structures and in the natural environment is best illustrated by n-octanol and water as solvents. After mixing, the two solvents form two separate phases although, due to their mutual partial solubility, the resulting system comprises octanol saturated with water and water saturated with octanol.

The value of the n-octanol/water partition coefficient (logP) depends not only on the chemical structure of dissolved molecule of the compound (number and type of various functional groups, contribution of unsaturated sites, dipole moment value) but also – and to a high degree – on its size. The n-octanol/water partition coefficient, as measured for typical pollutants, is in a very wide range from 0.01 for high-polarity compounds to 10¹⁰ for highly hydrophobic substances. The span of values being so wide, the partition coefficient is expressed in the logarithmic form:

$$\log P = \log C_{okt} - \log C_w \quad (1)$$

where:

C_{okt} – molar concentration of substance in octanol (mol/L),

C_w – molar concentration of substance in water (mol/L).

Log Kow (log P) is an important parameter in the assessment of substances in for chemical safety, CLP classification (classification, labeling and packaging of substances and preparations) and PBT assessment (permanent/bioaccumulating/toxic) – Fig. 2.

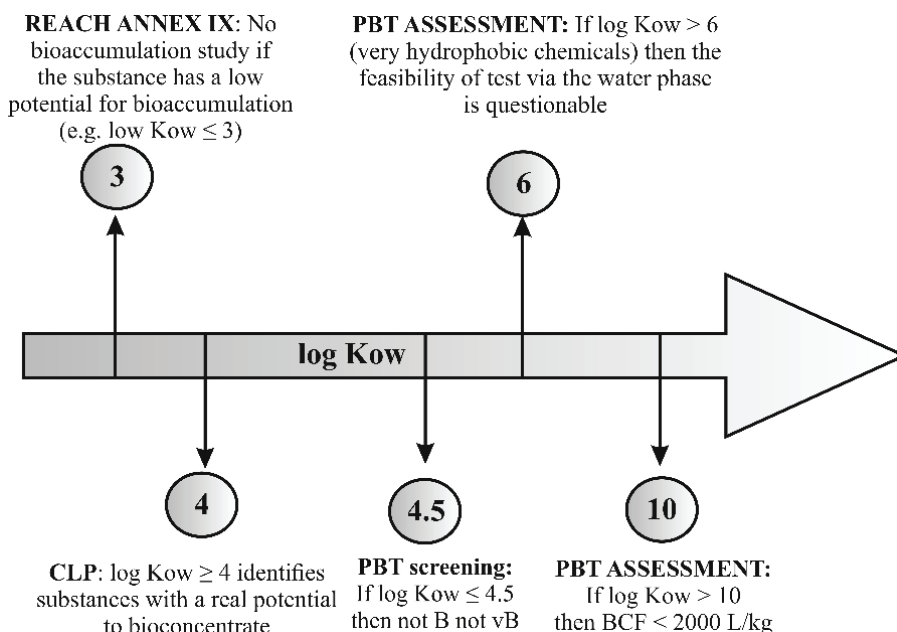


Fig. 2. Regulatory requirements for log Kow [12].

According to REACH, log Kow tests are required for all substances with production volumes of more than 1 Mg/year [1]. If $\log Kow \leq 3$, then (according to REACH, Annex IX), bioaccumulation tests are not required for substances with low bioaccumulative potential. For the purpose of CLP classification, $\log Kow \geq 4$ is the limit. The parameter is intended to identify substances with bioconcentration potential (BCF). The ability of substances to bioconcentrate is linked with their lipophilicity, in turn, is related to log Kow [12].

The use of QSAR methods for the assessment of the n-octanol/water partition coefficient using different models was reported earlier for polychlorinated biphenyls, thiophenols, benzene, ethanol, and other substances [12-17].

To assess how reliable the QSAR methods are in determination of the physico-chemical, toxicological and ecotoxicological properties of chemical substances, the results of tests on animals are compared with those performed using *in silico* methods. Models, enabling the assessment of the effect of chemical substances on the human health and natural environment are commonly available. Based on data from the ANTARES project, which promoted the use of numerical methods for determination of the properties of chemical substances, it was established that the following tools were applicable in the log Kow assessment: ToxPredict, VEGA, Epi Suite, SPARC, VCCLAB [18].

In this report, the results of own tests are compared with those obtained using different QSAR methods in log Kow determination for the hydroxyester HE-1. Three different models, provided by the VEGA tool *in silico*: Kowwin, MlogP, AlogP, were used.

2 Methods and materials

2.1 Materials

The hydroxyester HE-1 (3-hydroxy-2,2,4-trimethylpentyl isobutyrate) is used as an additive to paints and varnishes and is obtained in a sequence of chemical reactions in which isobutyric aldehyde is the basic starting material in the aldol condensation process with the subsequent Cannizzaro and Tishchenko reaction [19] (Fig. 3).

The hydroxyester HE-1 (Fig. 3) is a hydrophobic organic solvent of which the main use is as coalescent in water-based architectural paint formulations [20].

HE-1 has a boiling point of 255°C which means, it is not a volatile organic compound (VOC). Its density is 0.9477 g/cm³, flash point is 128°C, autoignition temperature is 410°C.

2.2. Method to determine logP

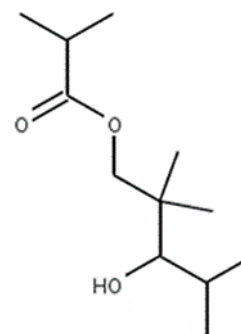


Fig. 3 Semi-structural formula for hydroxyester HE-1.

The n-octanol/water partition coefficient for the hydroxyester HE-1 was found to be 2.1 in own tests, performed in accordance with the OECD Test Guideline No. 117 [21].

LogP was found using the high-performance liquid chromatography apparatus HP 1090 with injector with loop capacity 20µl, UV detector with diode array, Purospher RP-18 column for HPLC, analytical balance, and measuring glass. The partition coefficient was determined using the following control substances: bromobenzene, thymol, diphenylamine, isopropylbenzene, benzyl benzoate, 1,2,4-trichlorobenzene, n-butylbenzene, dibenzyl and triphenylamine.

2.3 VEGA *in silico*

VEGA *in silico* renders accessible a number of QSAR models for the assessment of hazards, posed by chemical substances or for the development of individual models for research purposes. The platform is based on the CAESAR or T.E.S.T. data bases but it also uses other ones, for instance: EUTOXRISK, Toxbank, Caleidos, Orchestra, Antares-life, Life-prosil, LIFE-edesia.

The VEGA platform uses algorithms which are independent of the QSAR models and which enable the search for similar compounds. Such algorithms are used both for the identification of similar compounds and for the analysis of the role of descriptors (numerical value for the physicochemical parameters which characterize the specific ligand) and fragments for the chemicals searched for as well as for similar compounds.

The VEGA tool uses an algorithm which takes into account the molecular similarity of elements and molecules of chemical compounds in the aspect of relationship between their structure and properties [22]. In the VEGA tool, two compounds or elements are considered to be similar if their chemical structures and reactivities are similar in comparable conditions.

In this paper, the author used the latest version of VEGA *in silico* platform 1.1.3, available from www.vega-qsar.eu [23].

The input was the unambiguous notation of the structure of chemical compounds, based on the following character series (smiles) for HE-1: CC(C)C(C(C)(C)COC(=O)C(C)C)O.

The models Kowwin, MlogP and AlogP are discussed in this paper as ones used for the calculation of the n-octanol/water partition coefficient.

In the VEGA platform, the coefficients AlogP and MlogP were used as part of the Meylan model, implemented in the VEGA tool. Methods for the calculation of the n-octanol/water partition coefficient are based on similarities in the structure or properties of chemical compounds.

2.3.1 LogP model (Meylan/Kowwin) version 1.1.4

The model is based on the method of average fragment consistency (AFC), also known as Kowwin [24]. Generally, according to the method, an atom other than hydrogen is the core for a molecule fragment. Each fragment of a compound depends on the type of atoms which are connected with the core. The AFC method uses 250 correction coefficients for steric effects, H bonds, and influence of polar structural fragments. The correction coefficient is calculated from multiple linear regression (MLR). Regression analysis takes into consideration the size of the molecule fragment and its frequency in the structure as well as the correction coefficient and its frequency in the structure.

2.3.2 MlogP (Moriguchi logP) version 1.1.2

The MlogP model is based on the regression equation with 13 structural parameters [25,26]. The regression coefficients were assessed on the basis of 1230 organic compounds, including aliphatic, aromatic and heterocyclic compounds comprising the following elements: C, H, O, N, S, P, F, Cl, Br, I. The model variables include: the sum of lipophilic and hydrophilic atoms, N/O atom proximity effect, number of unsaturated bonds, number of polar aromatic substitutes, presence of ring-type structures, number of nitrogen groups, presence of intramolecular hydrogen bonds, amphoteric properties.

2.3.3 AlogP (Ghose-Crippen-Viswanadhan logP) version 1.1.2

The AlogP model is based on the regression equation, describing the hydrophobic properties of 120 atom types [27]. Compounds comprising such atoms as: C, H, O, N, S, Se, P, B, Si and halogens are considered. For each compound, logP is calculated as the sum of the number of all atoms, multiplied by their corresponding hydrophobicity constants.

2.4 Evaluation of models

The results were evaluated using the determination coefficient (R^2) and root-mean square error (RMSE), which are calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - y_{avg})^2} \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (3)$$

where:

y_i – experimental value,

\hat{y}_i – value, obtained using QSAR method,

y_{avg} – average value of all the values obtained using the QSAR methods,

n – number of substances tested.

The QSAR model is regarded as satisfactory when R^2 is equal to or as close to 1 as possible and RMSE is as close to 0 as possible.

3 Results and discussion

Using the VEGA *in silico* tool, the values of logP were calculated by means of the three QSAR models: Kowwin, MlogP, AlogP. Table 1 shows the compounds with –molecular similarity to HE-1 which were used in the logP calculations for the three different models. The data in Table 1 indicate that chemical similarity of the compounds to HE-1 varied between 0.805 and 0.878.

Then, using three models, the results of experiments (originating from different data bases) were compared with those obtained by means of the VEGA tool and substances with chemical similarity to HE-1 (Fig. 4 - 6). Determination coefficients (R^2) and root-mean square errors (RMSE) were then calculated for each of the three models. The R^2 was high for each of the three models: 0.9878 for MlogP, 0.9596 for AlogP, and 0.9718 for Kowwin.

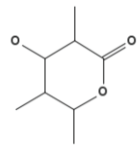
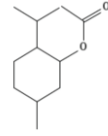
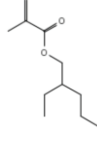
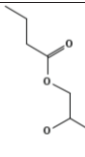
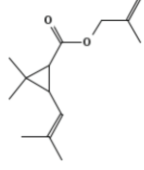
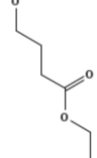
The RMSE was obtained from formula (3) for all the three QSAR models. For MlogP, RMSE was the highest: 1.03. For AlogP and Kowwin, RMSE was 0.72 and 0.39, respectively.

Taking into consideration molecular similarity, the results for R^2 and RMSE indicate that the Kowwin method is the optimum solution for determination of the n-octanol/water partition coefficient.

The average value of logP was then found for each of the three models (average of 6 values, obtained for compounds with molecular similarity) and compared with the value obtained in own experiments (logP = 2.1), as well as relative error was calculated (Table 2).

The results shown in Table 2 indicate that the best conformity between the results of own experiments and those originating from the QSAR models was obtained for the AlogP model.

Table 1. List of substances with chemical similarity to HE-1

CAS number of chemical compound	Smiles	Similarity	Semi-structural formula
50405-44-2	<chem>O=C1OC(C)C(C)C(O)C1C</chem>	0.878	
16409-45-3	<chem>O=C(OC1CC(C)CCC1C(C)C)C</chem>	0.829	
688-84-6	<chem>O=C(OCC(CC)CCCC)C(=C)C</chem>	0.825	
557-25-5	<chem>O=C(OCC(O)CO)CCC</chem>	0.82	
96358-98-4	<chem>O=C(OCC(=C)C)C1C(C=C(C)C)C1(C)C</chem>	0.816	
999-10-0	<chem>O=C(OCC)CCCO</chem>	0.805	

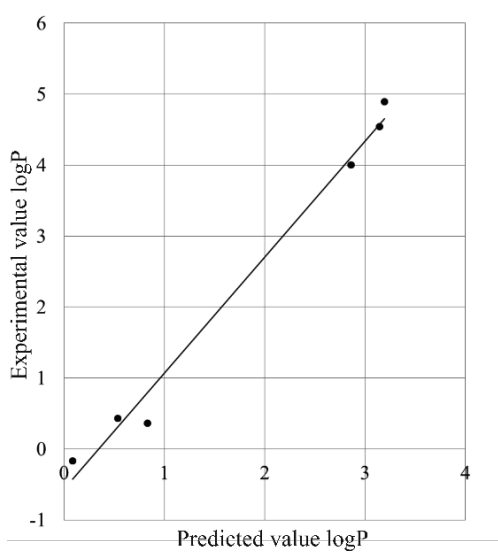


Fig. 4 Comparison between experimental results with those obtained using the MlogP model for substances with molecular structures similar to that HE-1

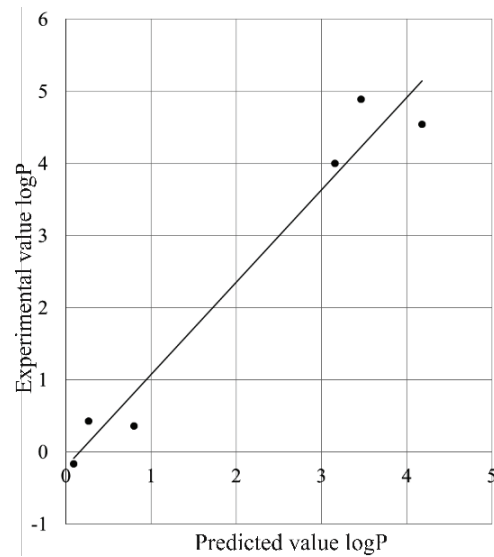


Fig. 5 Comparison between experimental results with those obtained using the AlogP model for substances with molecular structures similar to that HE-1

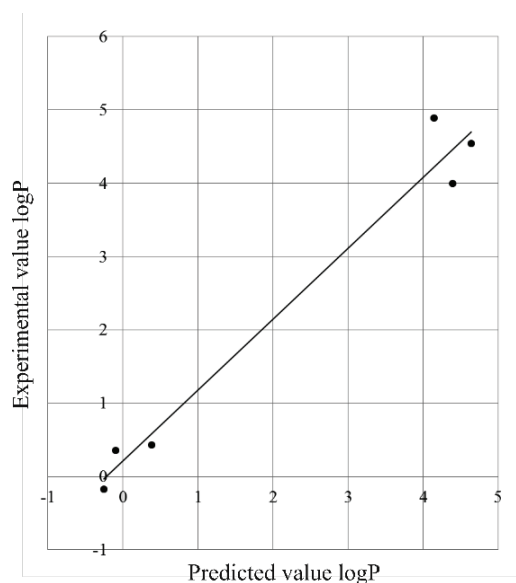


Fig. 6 Comparison between experimental results with those obtained using the Kowwin model for substances with molecular structures similar to that HE-1

Table 2 Comparison of average logP values provided by different QSAR models with the value obtained in own experiments for HE-1

Method	logP	Relative error, %
Own experiments	2.1	-
MlogP	2.75	30.9
AlogP	2.42	15.2
Kowwin	3	42.8

4 Conclusions

A comparison between the results of experiments on the calculation of the n-octanol/water partition coefficient for the additive to paints and varnishes – hydroxyester HE-1, and those obtained using the VEGA *in silico* tool was presented. Three models: logP: MlogP, AlogP and Kowwin were compared.

The results indicate that, for the substance with molecular similarity to HE-1, the highest determination coefficient was obtained for the model MlogP. When comparing the root-mean square error, the lowest value was obtained for the model Kowwin.

Taking into consideration the mean logP value, obtained using the QSAR models with those obtained in own experiments, it was observed that the optimum conformity (lowest relative error) was recorded for the model AlogP.

References

- Regulation (EC) No. 1907/2006 of the European Parliament and of the Council (2006)

- Directive of the European Parliament and of the Council No. 2010/63/UE (2010)
- European Chemical Agency. How to use alternatives to animal testing to fulfil your information requirements for REACH registration (2016)
- W. Liliensblum, W. Dekant, H. Foth, T. Gebel, J.G. Hengstler, R. Kahl, P.-J. Kramer, H. Schweinfurth, K.-M. Wollin, Arch. Toxicol. **82**, 211 (2008)
- H.M. Patel, M.N. Noolvi, P. Sharma, V. Jaiswal, S. Bansal, S. Lohan, S.S. Kumar, S. Dhiman, V. Bhardwaj, Med. Chem. Res. **23**, 12 (2014)
- A. Cherkasov, E.N. Muratov, D. Fourches, A. Varnek, I.I. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y.C. Martin, R. Todeschini, V. Consonni, V.E. Kuzmin, R. Cramer, R. Benigni, Ch. Yang, J. Rathman, L. Terfloth, J. Gasteiger, A. Richard, A. Tropsha, J. Med. Chem. **57**, 12, (2014)
- M.A. Sobanska, R. Cesnaitis, T. Sobanski, B. Versonnen, V. Bonnomet, J.V. Tarazona, W. De Coen, W., Sci. Total Environ. **470-471**, 1225 (2014)
- J.V. Tarazona, M.A. Sobanska, R. Cesnaitis, T. Sobanski, V. Bonnomet, B. Versonnen, W. De Coen, W., Sci. Total Environ. **472**, 137 (2014)
- R. Cesnaitis, M.A. Sobanska, B. Versonnen, T. Sobanski, V. Bonnomet, J.V. Tarazona, W. De Coen, W. Sci. Total Environ. **475**, 116 (2014)
- B. Versonnen, J.V. Tarazona, R. Cesnaitis, M.A. Sobanska, T. Sobanski, V. Bonnomet, W. De Coen, W., Sci. Total Environ., **475**, 123 (2014)
- J. Guziałowska-Tic, W.J. Tic, Przem. Chem. **95**, 8 (2016)
- C.I. Cappelli, E. Benfenati, J. Cester, Environ. Res. **143**, 26 (2015)
- M. Rucki, M. Tichý, QSAR Comb. Sci. **25**, 4 (2006)
- A.A. Toropov, A.P. Torpova, C.I. Cappelli, E. Benfenati, Fluid Phase Equilib. **397**, 44 (2015)
- Shi J.-Q., J. Cheng, F.-Y. Wang, A. Flamm, Z.-Y. Wang, X. Wang, S.-X. Gao, Ecotoxicol. Environ. Saf. **78**, 134 (2012)
- F.-L. Zhang, X.-J. Yang, X.-L. Xue, X.-Q. Tao, G.-N. Lu, Z. Dang, J. Chem. Article ID 740548 (2013)
- R.Z. Chen, W.Z. Tang, M. Sillanpää, Environ. Process. **3**, 73 (2016)
- www.antaes-life.eu (22.03.2017)
- W. J.Tic, Advances in Chemical Engineering and Science. **6**, 4, (2016)
- Research work, funded from the resources for science in the years 2007-2010 as the development project Technological Initiative I No. 13856.
- Test No. 117: Partition Coefficient (n-octanol/water), HPLC Method (2004)
- M. Floris, A. Manganaro, O. Nicolotti, R. Medda, G. F. Mangiatordi, E. Benfenati, J. Cheminform. **6**, 39 (2014)
- www.vega-qsar.eu (25.03.2017)
- W.M. Meylan, P.H. Howard, J. Pharm. Sci. **84**, 83 (1995)
- I. Moriguchi, S. Hirono, Q. Liu, I. Nagakome, Y. Matushita, Chem. Pharm. Bull. **42**, 976 (1994)
- I. Moriguchi, S. Hirono, I. Nagakome, H. Hirano, Chem. Pharm. Bull. **40**, 127 (1992)
- A.K. Ghose, V.N. Viswanadhan, J.J. Wendolski, J. Phys. Chem. A **102**, 3762 (1998)