

V.V. Vazhev<sup>1</sup>, B.G. Munarbaeva<sup>1</sup>, E.M. Yergaliyeva<sup>2</sup>, N.V. Vazheva<sup>2</sup>, M.A. Gubenko<sup>2</sup>

<sup>1</sup>Z. Aldamzhar Kostanay Social and Technical University, Kazakhstan;

<sup>2</sup>Kostanay State Pedagogical Institute, Kazakhstan  
(E-mail: v.vazhev@gmail.com)

## Modeling of acute aqueous toxicity of organic compounds for *Daphnia magna*

Methods for the assessment of the environmental hazard of chemicals include the detection of acute aquatic toxicity of compounds for various small organisms, these are test objects. Experimental determination of toxicity presents significant difficulties; therefore, computational methods are developed. The article is devoted to predicting the toxicity parameter of organic substances of different classes in relation to the test organism from the class of crustaceans *Daphnia magna*. The possibility of using QSAR / QSTR (Quantitative Structure Activity Relationships / Quantitative Structure Toxicity Relationships) method for estimating the toxicity index of pLC<sub>50</sub> using a large set of descriptors was investigated. To describe the structure of molecules, a set of 2644 descriptors generated by the Dragon 7 program was used. The calculations were carried out using the computer program Program Robustness Calculation (PROGROC). This program allows the number of descriptors to be used without preliminary selection exceeding the number of substances in the set. 546 Substances were used to construct the models, 170 compounds of which made up a control set, and 376 entered the training set. The obtained models are characterized by a high correlation between calculation and experiment. For a control sample of 170 compounds, the correlation coefficient is  $R = 0.952$  and the standard deviation is  $s = 0.45$ , for a training sample of 376 compounds,  $R = 0.971$  and  $s = 0.41$ . The model for all samples is characterized by the following statistical parameters: correlation coefficient is  $R = 0.966$  and standard deviation is  $s = 0.42$ .

**Keywords:** organic compounds, toxicity, modeling, QSAR/ QSTR, predicting, pLC<sub>50</sub>, *Daphnia magna*, PROGROC.

### Introduction

Determination of acute toxicity of chemical compounds in relation to test objects is one of the methods for assessing the environmental hazard of substances. *Daphnia magna* daphnoids are often used as the objects of investigation. They are small organisms of the crustacean class.

Toxicity is usually assessed by the quantitative parameter LC<sub>50</sub>, a semilethal concentration (mol/L), which causes half of the population to die within 24 (or 48) hours of exposure. The attractiveness of *Daphnia magna* as a test object is mainly due to the fact that it is grown without any special difficulties in the laboratory and is fairly stable in them, it gives a whole complex of test reactions and has a short life cycle that allows tracing the effects of toxic effects over a number of generations. Biotests on *Daphnia magna* are standardized in a number of countries. Nevertheless, like any experimental study, the definition of LC<sub>50</sub> is a costly procedure. Therefore, alternative methods based on quantitative structure-activity relationships, such as QSAR-methods (Quantitative Structure Activity Relationships), are developed. In such studies, the measure of toxicity is pLC<sub>50</sub> = -lg(LC<sub>50</sub>) — the toxicity parameter. The quality of calculations (modeling, prediction, estimation) is characterized by the correlation coefficient between the predicted and experimental values of the toxicity parameter  $R$  (or  $R^2$ ) and the standard deviation  $s$ .

Vighi and Calamari [1] carried out one of the first studies on several organotin compounds using physical and topological parameters. Equations with a significant correlation  $R^2 = 0.26-0.99$  and high prognostic ability were found. In [2] on the quantum-mechanical molecular descriptors, the toxicity models were constructed for 15 organophosphorus compounds with a correlation index of  $0.80 < R^2 < 0.82$ . In addition, the potential toxicity of 83 unstudied compounds was detected.

Cassani and Kovarich [3] simulated the toxicity of 97 triazoles and benzotriazoles using the multiple linear regression method with a maximum  $R^2 = 0.83$ . Moosus and Maran [4] in the study of 253 compounds of different classes reached the maximum value  $R^2 = 0.74$  of all the models obtained. Toropov and Benfenati [5] have used topological indices as descriptors for predicting the toxicity of 262 pesticides. The best model had  $R^2 = 0.7822$  and  $s = 0.849$  for a training set of 220 substances and  $R^2 = 0.7388$ ,  $s = 0.941$  for a control set of 42 substances. Toropova et al. [6] used descriptors as functions of representing the structure of compounds in simplified molecular input line entry system (SMILES) and obtaining models with the highest value  $R^2 = 0.7322$ . Another article by Toropova et al. [7] used a hybrid representation of the structure of mole-

cules through SMILES and hydrogen-suppressed graph. The results for the obtained models are comparable with those of [6], while it was noted that hybrid descriptors are the best predictors.

Katritzky et al. [8] modeled the toxicity of  $EC_{50}$  against *Daphnia magna* for a series of 130 benzoic acids, benzaldehydes, phenylsulfonyl acetates, cycloalkane-carboxylates, benzanilides, and other esters. A general model with five descriptors, including all 130 compounds, was characterized by the following statistical parameters such as  $R^2 = 0.712$ ,  $s = 0.6$ . Kar, Roy [9] investigated a large group of 297 chemical compounds, predictors were three-dimensional (electronic and spatial) and two-dimensional topological and information content indices, as well as n-octanol/water partition coefficient. When dividing the set into a training set of 222 substances and a control set of 75, the maximum value  $R^2 = 0.75$  is reached. Kühne et al. [10] investigated a large collection of 1365 organic compounds with experimental data of toxicity and noted contradictions in the part of data obtained from various sources. They obtained local models for subsets of the entire set, with the maximum values  $R^2 = 0.85$  and  $s = 0.66$ .

Cassotti et al. [11] subjected the existing toxicity data sets to examination and formed a MICHEM set of 546 substances that were most spared from discrepancies in the experimental data from various sources. When this set was divided into a training set of 436 substances and a control set of 110, they obtained a model with  $R^2 = 0.78$ .

Commenting the current state of the prediction of acute toxicity of chemical compounds for *Daphnia magna*, it can be noted that for heterogeneous and large sets of substances, the quality of forecasting cannot be considered sufficiently successful.

Assuming that the quality of predicting can be improved by increasing the number of descriptors and improving the computational technology, in this work we used a large set of 2644 descriptors generated by the Dragon 7 program to predict the  $pLC_{50}$  for *Daphnia magna*. The  $pLC_{50}$  toxicity values for *Daphnia magna* are derived from the appendix to the article [11].

Calculations were performed using the PROGROC (PROGram ROBustness Calculation) computer program developed by us [12], which was successfully applied to predict some parameters of biological activity, in particular, toxicity of organic compounds for *Tetrahimena pyriformis* [13]. An important and unique feature of this program is the possibility of using without a preliminary selection the number of descriptors exceeding the number of substances.

A number of models were built, for one of them the results of  $pLC_{50}$  prediction are shown in Figures 1 and 2. The training set included 376 compounds, the control set — 170. The statistical parameters of the model are given in Table 1.

Table 1

Correlation between the experimental and calculated  $pLC_{50}$  values

Correlation	Full set	Training set	Control set
R	0.966	0.971	0.952
s	0.42	0.41	0.45
Amount of substances	546	376	170

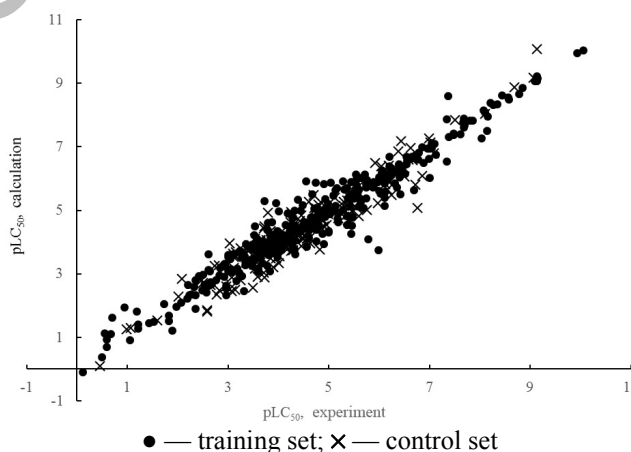


Figure 1. Correlation between the experimental values of  $pLC_{50}$  and calculated by molecular descriptors for the training and control sets

The frequency distribution of pLC<sub>50</sub> prediction errors using molecular descriptors is shown in a histogram (Fig. 2).

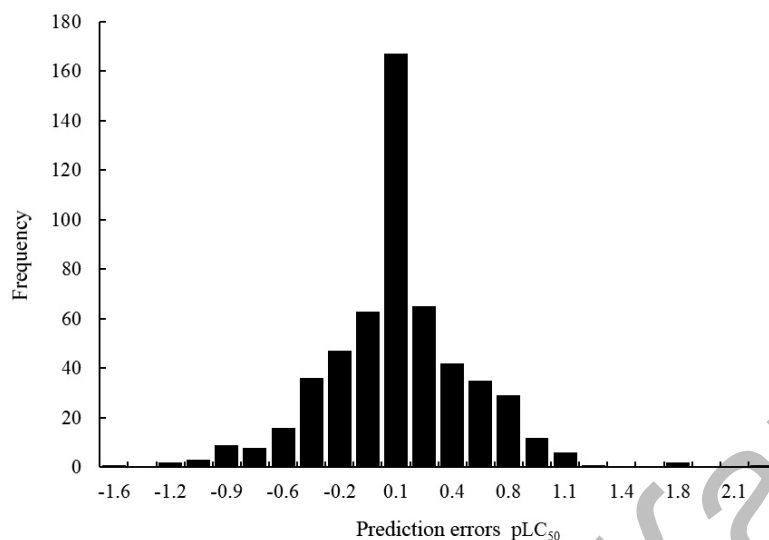


Figure 2. Histogram of frequency distribution of pLC<sub>50</sub> prediction errors

The histogram of prediction error frequency distribution has edge emissions; the Table 2 shows the characteristics of the connections corresponding to these emissions.

Table 2

**Compounds with significant discrepancies between experimental [11] and calculated values of pLC<sub>50</sub>**

Compound	CAS*	Experiment	Calc.	Residual	Set
Acrylic aldehyde	107-02-8	5.99	3.74	2.25	Training
N-methylaniline	100-61-8	5.79	4.08	1.71	Training
Pyraclostrobin	175013-18-0	6.76	5.07	1.69	Control
Pendimethalin	40487-42-1	3.72	5.30	-1.57	Training
Bis(2-ethylhexyl)phthalate	117-81-7	4.55	5.92	-1.37	Training
9H-xanthene	92-83-1	3.95	5.22	-1.27	Training
Pyrethrin 2	121-29-9	7.37	8.59	-1.22	Training
1,2,4,5-Tetramethylbenzene	95-93-2	5.45	4.27	1.18	Training
Benzo[b]naphtho[1,2-d]furan	205-39-0	4.74	5.88	-1.15	Training
2-(6-Methoxy-2-naphthyl)propanoic acid	22204-53-1	3.79	4.93	-1.14	Control
Anthracene	120-12-7	6.70	5.63	1.07	Training
Methylcyclohexane	108-87-2	4.82	3.77	1.05	Control
1-Methyl-2-pyrrolidinone	872-50-4	4.91	3.93	0.98	Training
1-Propanol	71-23-8	0.94	1.95	-1.02	Training
1-Phenoxy-2-propanol	770-35-4	2.61	3.62	-1.01	Training

Note: CAS — registration number for Chemical Abstracts.

It is very likely that the listed in the table, or influencing them through links in the compound model, have errors in the experimental determination of the toxicity parameter.

### Conclusion

In conclusion, we can note that the results obtained by us are superior to the literature data on the quality of predicting the toxicity parameter towards *Daphnia magna*. So, for example, in the most successful model [11], obtained on the same set of substances, but with a smaller sample size, the correlation coefficient is 0.88 ( $R^2 = 0.78$ ).

## References

- 1 Vighi M. QSARs for organotin compounds on *Daphnia magna* / M. Vighi, D. Calamari // Chemosphere. — 1985. — Vol. 14. — P. 1925–1932.
- 2 Zvinavashe E. Quantitative structure-activity relationship modeling of the toxicity of organothiophosphate pesticides to *Daphnia magna* and *Cyprinus carpio* / E. Zvinavashe, T. Du, T. Griff, H.H.J. van den Berg, A.E.M.F. Soffers, J. Vervoort et al. // Chemosphere. — 2009. — Vol. 75. — P. 1531–1538.
- 3 Cassani S. *Daphnia* and fish toxicity of (benzo)triazoles: validated QSAR models, and interspecies quantitative activity-activity modelling / S. Cassani, S. Kovarich, E. Papa, P.P. Roy, L. van der Wal, P. Gramatica // Journal of Hazardous Materials. — 2013. — Vol. 258–259. — P. 50–60.
- 4 Moosus M. Quantitative structure-activity relationship analysis of acute toxicity of diverse chemicals to *Daphnia magna* with whole molecule descriptors / M. Moosus, U. Maran // SAR and QSAR in Environmental Research. — 2011. — Vol. 22. — P. 757–774.
- 5 Toropov A.A. QSAR models for *Daphnia* toxicity of pesticides based on combinations of topological parameters of molecular structures / A.A. Toropov, E. Benfenati // Bioorganic & Medicinal Chemistry. — 2006. — Vol. 14. — P. 2779–2788.
- 6 Toropova A.P. QSAR models for toxicity of organic substances to *Daphnia magna* built up by using the CORAL freeware / A.P. Toropova, A.A. Toropov, E. Benfenati, G. Gini // Chemical Biology & Drug Design. — 2012. — Vol. 79. — P. 332–338.
- 7 Toropova A.P. CORAL: QSAR modeling of toxicity of organic chemicals towards *Daphnia magna* / A.P. Toropova, A.A. Toropov, S.E. Martyanov, E. Benfenati, G. Gini, D. Leszczynska et al. // Chemometrics and Intelligent Laboratory Systems. — 2012. — Vol. 110. — P. 177–181.
- 8 Katritzky A.R. Quantitative structure-activity relationship (QSAR) modeling of EC<sub>50</sub> of aquatic toxicities for *Daphnia magna* / A.R. Katritzky, S.H. Slavov, I.S. Stoyanova-Slavova, I. Kahn, M. Karelson // Journal of Toxicology and Environmental Health, Part A. — 2009. — Vol. 72. — P. 1181–1190.
- 9 Kar S. QSAR modeling of toxicity of diverse organic chemicals to *Daphnia magna* using 2D and 3D descriptors / S. Kar, K. Roy // Journal of Hazardous Materials. — 2010. — Vol. 177. — P. 344–351.
- 10 Kühne R. Read-across prediction of the acute toxicity of organic compounds toward the water flea *Daphnia magna* / R. Kühne, R.-U. Ebert, P.C. von der Ohe, N. Ulrich, W. Brack, G. Schüürmann // Molecular Informatics. — 2013. — Vol. 32. — P. 108–120.
- 11 Cassotti M. Validation and extension of a similarity-based approach for prediction of acute aquatic toxicity towards *Daphnia magna* / M. Cassotti, V. Consonni, A. Mauri, D. Ballabio // SAR and QSAR in Environmental Research. — 2014. — Vol. 25(12). — P. 1013–1036.
- 12 Vazhev V.V. Estimation of flash points and molecular masses of alkanes from their IR spectra / V.V. Vazhev, M.K. Aldabergenov, N.V. Vazheva // Petroleum Chemistry. — 2006. — Vol. 46. — No. 2. — P. 136–139.
- 13 Важев В.В. Оценка токсичности органических соединений для *Tetrahymena pyriformis* методом QSAR / В.В. Важев, М.А. Губенко, Н.В. Важева, Э.М. Ергалиева, Е.А. Самсолюк // Вестн. Караганд. ун-та. Сер. Химия. — 2016. — № 3(83). — С. 39–44.

В.В. Важев, Б.Г. Мунарбаева, Э.М. Ергалиева, Н.В. Важева, М.А. Губенко

### ***Daphnia magna* үшін органикалық қосылыстардың өткір сулы улылығын модельдеу**

Химиялық заттардың экологиялық қауіпсіздігін бағалау әдістерінің біріне әртүрлі ұсақ ағзалар — тест-объектілер үшін қосылыстардың өткір сулы улылығын анықтау жатады. Улылықты эксперименттік анықтау едәуір қиындықтар тудырады, сондықтан есептеу әдістері дамытылуда. Мақала әртүрлі класс органикалық заттарының улылық параметрін шаян тәрізділер класының өкілі болып табылатын *Daphnia magna* тәсілдік организмне қатысты болжауға арналған. Дескрипторлардың үлкен жиынтығын пайдаланып, QSAR/QSTR (Quantitative Structure Activity Relationships/Quantitative Structure Toxicity Relationships) әдісін pLC<sub>50</sub> улылық көрсеткішін бағалау үшін қолдану мүмкіндігі зерттелген. Молекулалардың құрылысын сипаттау үшін Dragon 7 бағдарламасы генерациялайтын 2644 дескрипторлар жиынтығы қолданылды. Есептеулер Program Robustness Calculation (PROGROC) компьютерлік бағдарламасы көмегімен жүргізілді. Бұл бағдарлама саны жиынтықтағы заттар санынан артық болатын дескрипторларды алдын ала іріктемей пайдалануға мүмкіндік береді. Модельдерді құру үшін 546 зат қолданылды, олардың ішінен 170 қосылыс бақылау іріктемесін, ал 376-ы жаттығу іріктемесін құрды. Алынған модельдер есептеу мен эксперимент арасындағы корреляцияның жоғары болуымен сипатталады. Бақылау іріктемесі үшін корреляция коэффициенті R = 0,952 және стандартты ауытқу s = 0,45, жаттығу іріктемесі үшін R = 0,971 және s = 0,41. Толық іріктеме үшін жасалған модель келесі статистикалық параметрлермен сипатталды: корреляция коэффициенті R = 0,966 және стандартты ауытқу s = 0,42.

*Кілт сөздер:* органикалық қосылыстар, улылық, болжау, QSAR/ QSTR, модельдеу, pLC<sub>50</sub>, *Daphnia magna*, PROGROC.

В.В. Важев, Б.Г. Мунарбаева, Э.М. Ергалиева, Н.В. Важева, М.А. Губенко

## Моделирование острой водной токсичности органических соединений для *Daphnia magna*

К числу методов оценки экологической опасности химических веществ относится выявление острой водной токсичности соединений для различных мелких организмов — тест-объектов. Экспериментальное определение токсичности представляет значительные трудности, поэтому развиваются вычислительные методы. Статья посвящена прогнозированию параметра токсичности органических веществ разных классов по отношению к тестовому организму из класса ракообразных *Daphnia magna*. Исследована возможность применения метода QSAR/QSTR (Quantitative Structure Activity Relationships/ Quantitative Structure Toxicity Relationships) для оценки показателя токсичности  $pLC_{50}$  с использованием большого набора дескрипторов. Для описания структуры молекул использован набор из 2644 дескрипторов, генерируемых программой Dragon 7. Расчеты выполнены с помощью компьютерной программы Program Robustness Calculation (PROGROC). Данная программа позволяет использовать без предварительного отбора число дескрипторов, превышающее число веществ в наборе. Для построения моделей использовано 546 веществ, из которых 170 соединений составили контрольную выборку, а 376 — вошли в тренировочную выборку. Полученные модели характеризуются высокой корреляцией между расчетом и экспериментом. Для контрольной выборки коэффициент корреляции  $R = 0,952$  и стандартное отклонение  $s = 0,45$ , для тренировочной выборки  $R = 0,971$  и  $s = 0,41$ . Модель для полной выборки характеризуется следующими статистическими параметрами: коэффициент корреляции  $R = 0,966$  и стандартное отклонение  $s = 0,42$ .

**Ключевые слова:** органические соединения, токсичность, моделирование, QSAR/QSTR, прогнозирование,  $pLC_{50}$ , *Daphnia magna*, PROGROC.

### References

- Vighi, M., & Calamari, D. (1985). QSARs for organotin compounds on *Daphnia magna*. *Chemosphere*, 14, 1925–1932.
- Zvinavashe, E., Du, T., Griff, T., van der Berg, H.H.J., Soffers, A.E.M.F., & Vervoort, J., et al. (2009). Quantitative structure-activity relationship modeling of the toxicity of organothiophosphate pesticides to *Daphnia magna* and *Cyprinus carpio*. *Chemosphere*, 75, 1531–1538.
- Cassani, S., Kovarich, S., Papa, E., Roy, P.P., van der Wal, L., & Gramatica, P. (2013). *Daphnia* and fish toxicity of (benzo)triazoles: validated QSAR models, and interspecies quantitative activity-activity modelling. *Journal of Hazardous Materials*, 258–259, 50–60.
- Moosus, M., & Maran, U. (2011). Quantitative structure-activity relationship analysis of acute toxicity of diverse chemicals to *Daphnia magna* with whole molecule descriptors. *SAR QSAR Environ. Res.*, 22, 757–774.
- Toropov, A.A., & Benfenati, E. (2006). QSAR models for *Daphnia* toxicity of pesticides based on combinations of topological parameters of molecular structures. *Bioorg. Med. Chem.*, 14, 2779–2788.
- Toropova, A.P., Toropov, A.A., Benfenati, E., & Gini, G. (2012). QSAR models for toxicity of organic substances to *Daphnia magna* built up by using the CORAL freeware. *Chem. Biol. Drug Des.*, 79, 332–338.
- Toropova, A.P., Toropov, A.A., Martyanov, S.E., Benfenati, E., Gini, G., & Leszczynska, D., et al. (2012). CORAL: QSAR modeling of toxicity of organic chemicals towards *Daphnia magna*. *Chemom. Intell. Lab. Syst.*, 110, 177–181.
- Katritzky, A.R., Slavov, S.H., Stoyanova-Slavova, I.S., Kahn, I., & Karelson, M. (2009). Quantitative structure-activity relationship (QSAR) modeling of  $EC_{50}$  of aquatic toxicities for *Daphnia magna*. *J. Toxicol. Environ. Health, Part A.*, 72, 1181–1190.
- Kar, S., & Roy, K. (2010). QSAR modeling of toxicity of diverse organic chemicals to *Daphnia magna* using 2D and 3D descriptors. *J. Hazard. Mater.*, 177, 344–351.
- Kühne, R., Ebert, R., von der Ohe, P.C., Ulrich, N., Brack, W., & Schüürmann, G. (2013). Read-across prediction of the acute toxicity of organic compounds toward the water flea *Daphnia magna*. *Mol. Inform.*, 32, 108–120.
- Cassotti, M., Consonni, V., Mauri, A., & Ballabio, D. (2014). Validation and extension of a similarity-based approach for prediction of acute aquatic toxicity towards *Daphnia magna*. *SAR QSAR Environ. Res.*, 25(12), 1013–1036.
- Vazhev, V.V., Aldabergenov, M.K., & Vazheva, N.V. (2006). Estimation of flash points and molecular masses of alkanes from their IR spectra. *Petroleum Chemistry*, 46(2), 136–139.
- Vazhev, V.V., Gubenko, M.A., Vazheva, N.V., Yergaliieva, E.M., & Samsoniuk, Ye.A. (2016). Otsenka toksichnosti orhanicheskikh soedinenii dlia *Tetrahymena pyriformis* metodom QSAR [Estimation of the toxicity of organic compounds to *Tetrahymena pyriformis* by QSAR method]. *Vestnik Karahandinskoho universiteta. Seriya Khimiia — Bulletin of the Karaganda University. Ser. Chemistry*, 3(83), 39–44 [in Russian].