UDC 544.165

E.V. Varlamova, PhD in Chemistry, Junior Research Fellow,
V.E. Kuz'min, Doctor of Chemistry, Prof.,
A. V. Bogatsky Physico-Chemical Institute, Odessa,
N.N. Muratov, PhD in Chemistry, Assoc.Prof.,
V.A. Shapkin, PhD in Chemistry, Assoc.Prof.,
Odessa National Polytechnic University

INFLUENCE OF COMPONENTS STRUCTURE AND COMPOSITION OF BINARY MIXTURES OF ORGANIC COMPOUNDS ON TOXICITY TOWARDS *DAPHNIA MAGNA*

Introduction. Deleterious effects of synthetic organic compounds on the aquatic environment is the subject of thorough research during last few decades. Because of the possibility of joint action a mixture of several substances may cause a greater risk than individual substances. While present water quality standards are based on data of the toxicity of individual substances in the future monitoring should be carried out basing on the data of the joint action of complex multicomponent mixtures. It is well known that the waste water (industrial, agricultural, domestic wastewater) contains several chemicals simultaneously. Therefore, understanding and ability to predict the combined effects of mixtures of toxic substances on microorganisms and other aquatic life forms are essential components in the environment protection.

Literature review. All known at literature descriptions of mixtures can be divided into several groups [1]:

— Descriptors based on the distribution coefficient for mixtures [2...5]. Distribution coefficient (octanol — water, cyclohexane — water, chloroform — water, etc.) is used as descriptors. All studies using such descriptors were performed using very small training sets, and predictive ability has not been checked properly for any of models. These descriptors are very limited in use because they can be applied to mixtures of 1:1. Advantage of these descriptors is that the mixture is described as a single entity.

— Integral additive descriptors [6…7]. In this approach descriptors calculated for the mixtures components separately are summarized taking into consideration their mole fractions. These descriptors have not been tested on mixtures with non-additive properties, namely synergism or antagonism, i.e. the possibility of their use in such cases looks very doubtful. Among the advantages of the additive approach is the following: a simple and intuitive process of descriptors generating; independence of approach from the property under consideration and the possibility of its use (taking into account the disadvantage of it) for study of any activity or property.

— Integral non-additive descriptors [8…10]. The main feature of this approach is in describing the mixture itself, i.e. mixtures descriptors are not derived from its components descriptors taken separately. To achieve this, the mixture is represented using both components simultaneously. This avoids the limitations encountered in the calculation of descriptors by the additive scheme. However, this methodology does not allow to analyze (describe) mixtures different from 1:1, which is its significant disadvantage.

— Fragment non-additive descriptors [11]. In this approach, fragments of the different components of the mixture are considered in the same descriptor. These descriptors of the mixture do not

DOI 10.15276/opu.1.45.2015.28

ISSN 2076-2429 (print)
ISSN 2223-3814 (online)

contain disadvantages of previous methods. They can be used to simulate any interested property or activity, are useful for describing the effects of the joint action or interaction of components. However, it should be noted that mixtures in this approach are described by descriptors of the individual compounds, and the characteristics describing directly the mixture are absent.

Aims of the Research are the study of the influence of structure of aliphatic n-alcohols, phenols, sodium salts of α-alkoxycarbonylsulfuric acids and their mixtures on toxicity towards Daphnia magna using simplex representation of the molecular structure and the determination of molecular fragments with positive or negative influence on the toxicity.

Main Body. To describe the structure of the tested substances and mixtures the simplex representation of the molecular structure (SiRMS) is used [12]. Within SiRMS any molecule can be represented as a system of various simplexes (four atomic fragments of fixed composition and structure).

Descriptor representation of compounds mixtures cannot be defined as corresponding linear combination of descriptors of mixture components because in most cases the dependences "structure, composition — property" are not additive. Usage of the simplex descriptors characterizing the compounds mixture itself is the main difference from the usual simplex approach. Bounded simplexes describe only separated mixture components, when unconnected simplexes can describe both components separately and the whole mixture. Therefore, it is necessary to note which unbounded simplexes belong to the same molecule and which to different. In the last case, such unbounded simplexes reflect the structure not for one molecule but characterize a couple of different molecules. Actually these simplexes are the structural descriptors of substance mixture (Fig. 1). To distinguish these simplexes in the process of generating of descriptors they are assigned by special mark. This approach takes into account the composition of mixture, i.e. descriptors of the individual components (substances A and B) are weighted according to the mole fractions of components in the mixture, and mixture descriptors are multiplied by the minimum proportion of one of the components.

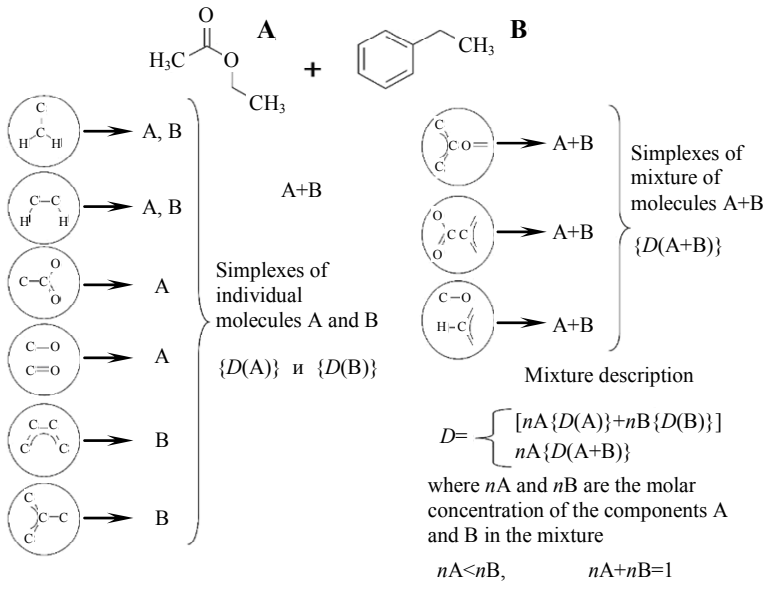


Fig. 1. Simplex descriptors of compounds mixture

To construct the models "structure — property" the method of partial least squares (PLS) is used [13], it has worked well when working with large data sets.

In this paper it is studied the effect of structure of aliphatic n-alcohols, phenols, sodium salts of α-alkoxycarbonylsulfuric acids and their mixtures on toxicity towards *Daphnia Magna* [14]. The toxicity of sodium salts of α-alkoxycarbonylsulfuricacids, phenols, aliphatic n-alcohols and their binary mixtures towards *Daphnia Magna* was used as investigated activity. Toxicity is defined as a decimal logarithm of EC_{50} (mmol/l). EC_{50} is the concentration of the toxic substance which leads to immobilization of 50 % of the test population.

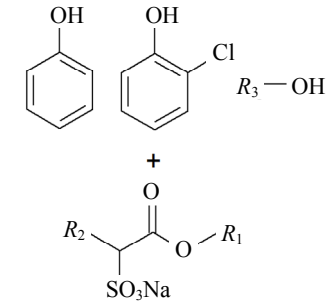
Results. At the initial stage of work for all molecules simplex descriptors are calculated (about 2 thousands), they are weighted for the mixtures according to the molar concentration of the mixture components. Differentiation of atoms in simplexes is based on the following characteristics:

— atom individuality;
— partial atomic charge;
— lipophilicity;
— electronic polarizability;
— possibility of atom to be a donor/acceptor of hydrogen in hydrogen bonding.

To test the predictive ability of the obtained models four strategies for the formation of training and test sets are used:

— compounds for the test set are selected by the minimal difference with the training one;
— compounds for the test set are selected by the maximal difference with the training one;
— compounds for the test set are selected randomly;
— all the studied compounds are included in the test set.

By 10 PLS-models are obtained for all four strategies. The quality of the obtained models is high enough — R^2=0,86…0,97, Q^2=0,74…0,96, R^2_{test}=0,86…0,99, so all of these models are included in the consensus model. This model has a high statistical characteristics (R^2=0,97, S=0,13). The ratio of the observed and predicted values is shown on Fig. 3.



R_1=C_nH_{2n+1}, n=1, 2, 4...7;
R_2=C_mH_{2m+1}, m=7, 8, 10, 12...14;
R_3=C_kH_{2k+1}, k=5, 6, 8, 9

Fig. 2. Studied compounds

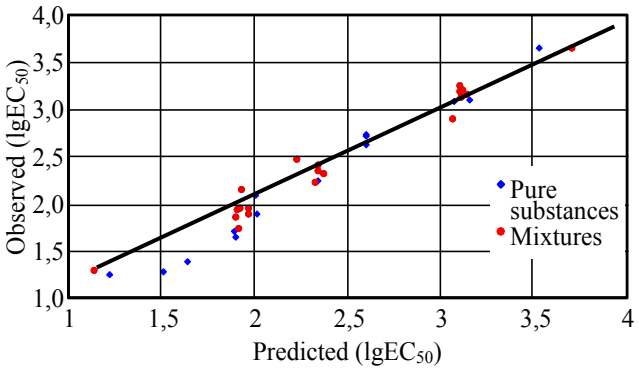


Fig. 3. The ratio of the observed and predicted values of the toxicity towards Daphnia Magna for the tested compounds and mixtures.

Further the inverse problem is solved. The contribution of each atom in the activity is determined for each molecule. It is found that increasing of carbon chain length in the alkyl substituent increases the toxicity. Likely it is connected with the growth of lipophilicity.

Analysis of the relative influence of some physical and chemical characteristics on the change of the studied properties, which are estimated from the resulting QSAR-models, showed that electrostatic factors (42 %) and the factors caused by atom individuality (30 %) have the largest contribution to the change of toxicity of the tested compounds.

Conclusions. QSAR-analysis of the toxicity of binary mixtures of sodium salts of α-alkoxycarbonylsulfuric acids with phenols and alcohols towards *Daphnia Magna* shows the possibility of reliable prediction of the specified property (R^2_{test}= 0,86…0,99). It is found that increasing of the chain length of the alkyl substituents increases the toxicity. Likely it is connected with the increase of compounds lipophilic.

Література

1. Existing and developing approaches for QSAR analysis of mixtures / E.N. Muratov, E.V. Varlamova, A.G. Artemenko [et al.] // Molecular Informatics. — 2012. — Vol. 31, Issue 3-4. — PP. 202 — 221.

ISSN 2076-2429 (print)
ISSN 2223-3814 (online)

2. Development of quantitative structure activity relationships in toxicity prediction of complex mixtures / H.-X. Yu, Z.-F. Lin, J.-F. Feng [et al.] // Acta Pharmacologica Sinica. — 2001. — Vol. 22, Issue 1. — PP. 45 — 49.
3. Wei, D.B. QSAR-based toxicity classification and prediction for single and mixed aromatic compounds / D.B. Wei, L.H. Zhai, H.-Y. Hu // SAR and QSAR in Environmental Research. — 2004. — Vol. 15, Issue 3. — PP. 207 — 216.
4. Prediction of mixture toxicity with its total hydrophobicity / Z. Lin, H. Yu, D. Wei [et al.] // Chemosphere. — 2002. — Vol. 46, Issue 2. — PP. 305 — 310.
5. Quantification of joint effect for hydrogen bond and development of QSARs for predicting mixture toxicity / Z. Lin, P. Zhong, K. Yin [et al.] // Chemosphere. — 2003. — Vol. 52, Issue 7. — PP. 1199 — 1208.
6. Quantitative structure-activity relationships for joint toxicity of substituted phenols and anilines to Scenedesmus obliquus / C. Wang, G. Lu, Z. Tang, X. Guo // Journal of Environmental Sciences. — 2008. — Vol. 20, Issue 1. — PP. 115 — 119.
7. Application of QSPR to mixtures / S. Ajmani, S.C. Rogers, M.H. Barley, D.J. Livingstone // Journal of Chemical Information and Modeling. — 2006. — Vol. 46, Issue 5. — PP. 2043 — 2055.
8. Computer-based QSARs for predicting mixture toxicity of benzene and its derivatives / L. Zhang, P.-J. Zhou, F. Yang, Z.-D. Wang // Chemosphere. — 2007. — Vol. 67 Issue 2. — PP. 396 — 401.
9. Characterization of mixtures Part 1: Prediction of infinite-dilution activity coefficients using neural network-based QSPR models / S. Ajmani, S.C. Rogers, M.H. Barley [et al.] // QSAR & Combinatorial Science. — 2008. — Vol.27, Issue 11-12. — PP. 1346 — 1361.
10. Characterization of mixtures. Part 2: QSPR models for prediction of excess molar volume and liquid density using neural networks / S. Ajmani, S.C. Rogers, M.H. Barley [et al.] // Molecular Informatics. — 2010. — Vol. 29, Issue 8-9. — PP. 645 — 653.
11. Quantitative structure–property relationship (QSPR) modeling of normal boiling point temperature and composition of binary azeotropes / V.P. Solov'ev, I. Oprisiu, G. Marcou, A. Varnek // Industrial & Engineering Chemistry Research. — 2011. — Vol. 50, Issue 24. — PP. 14162 — 14167.
12. Kuz'min, V.E. Hierarchical QSAR technology based on the Simplex representation of molecular structure / V.E. Kuz'min, A.G. Artemenko, E.N. Muratov // Journal of Computer-Aided Molecular Design. — 2008. — Vol. 22, Issue 6-7. — PP. 403 — 421.
13. A PLS kernel algorithm for data sets with many variables and fewer objects. Part 1: Theory and algorithm / S. Rännar, F. Lindgren, P. Geladi, S. Wold // Journal of Chemometrics. — 1994. — Vol. 8, Issue 2. — PP. 111 — 125.
14. Defining the toxic mode of action of ester sulphonates using the joint toxicity of mixtures / G. Hodges, D.W. Roberts, S.J. Marshall, J.C. Dearden // Chemosphere. — 2006. — Vol. 64, Issue 1. — PP. 17 — 25.

References

1. Muratov, E.N., Varlamova, E.V., Artemenko, A.G., Polishchuk, P.G. and Kuz'min, V.E. (2012). Existing and developing approaches for QSAR analysis of mixtures. *Molecular Informatics*, 31(3-4), 202-221.
2. Yu, H.-X., Lin, Z.-F., Feng, J.-F., Xu, T.-L. and Wang, L.-S. (2001). Development of quantitative structure activity relationships in toxicity prediction of complex mixtures. *Acta Pharmacologica Sinica*, 22(1), 45-49.
3. Wei, D.B., Zhai, L.H. and Hu, H.-Y. (2004). QSAR-based toxicity classification and prediction for single and mixed aromatic compounds. *SAR and QSAR in Environmental Research*, 15(3), 207-216.
4. Lin, Z., Yu, H., Wei, D. Wang, G., Feng, J. and Wang, L. (2002). Prediction of mixture toxicity with its total hydrophobicity. *Chemosphere*, 2002, 46(2), 305-310.
5. Lin, Z., Zhong, P., Yin, K., Wang, L. and Yu, H. (2003). Quantification of joint effect for hydrogen bond and development of QSARs for predicting mixture toxicity. *Chemosphere*, 52(7), 1199-1208.
6. Wang, C., Lu, G., Tang, Z. and Guo, X. (2008). Quantitative structure-activity relationships for joint toxicity of substituted phenols and anilines to Scenedesmus obliquus. *Journal of Environmental Sciences*, 20(1), 115-119.
7. Ajmani, S., Rogers, S.C., Barley, M.H. and Livingstone, D.J. (2006). Application of QSPR to mixtures. *Journal of Chemical Information and Modeling*, 46(5), 2043-2055.
8. Zhang, L., Zhou, P.-J., Yang, F. and Wang, Z.-D. (2007). Computer-based QSARs for predicting mixture toxicity of benzene and its derivatives. *Chemosphere*, 67(2), 396-401.
9. Ajmani, S., Rogers, S.C., Barley, M.H., Burgess, A.N. and Livingstone, D.J. (2008). Characterization of mixtures Part 1: Prediction of infinite-dilution activity coefficients using neural network-based QSPR models. *QSAR & Combinatorial Science*, 27(11-12), 1346-1361.

10. Ajmani, S., Rogers, S.C., Barley, M.H., Burgess, A.N. and Livingstone, D.J. (2010). Characterization of mixtures. Part 2: QSPR models for prediction of excess molar volume and liquid density using neural networks. *Molecular Informatics*, 29(8-9), 645-653.
11. Solov'ev, V.P., Oprisiu, I., Marcou, G. and Varnek, A. (2011). Quantitative structure–property relationship (QSPR) modeling of normal boiling point temperature and composition of binary azeotropes. *Industrial & Engineering Chemistry Research*, 50(24), 14162-14167.
12. Kuz'min, V.E., Artemenko, A.G. and Muratov, E.N. (2008). Hierarchical QSAR technology based on the Simplex representation of molecular structure. *Journal of Computer-Aided Molecular Design*, 22(6-7), 403-421.
13. Rännar, S., Lindgren, F., Geladi, P. and Wold, S. (1994). A PLS kernel algorithm for data sets with many variables and fewer objects. Part 1: Theory and algorithm. *Journal of Chemometrics*, 8(2), 111-125.
14. Hodges, G., Roberts, D.W., Marshall, S.J. and Dearden, J.C. (2006). Defining the toxic mode of action of ester sulphonates using the joint toxicity of mixtures. *Chemosphere*, 64(1), 17-25.

АНОТАЦІЯ / АННОТАЦИЯ / ABSTRACT

К.В. Варламова, В.Є. Кузьмін, Н.Н. Муратов, В.А. Шапкін. **Вплив структури компонентів та складу бінарних сумішей органічних сполук на токсичність по відношенню до *Daphnia Magna*.** Симплексне подання молекулярної структури використано для консенсусного QSAR-аналізу токсичності по відношенню до Daphnia Magna натрієвих солей α-алкоксикарбонілсульфокислот, аліфатичних н-спиртів, фенолів та їх бінарних сумішей. Структуру суміші представлено як з використанням дескрипторів індивідуальних сполук, що входять до складу суміші, так і нових специфічних параметрів суміші, так званих незв'язних симплексів. Досліджувана вибірка складалася з 15 індивідуальних сполук і 20 сумішей. Як досліджувану активність використано логарифм EC_{50} (мкмоль/л). Метою даного дослідження є визначення молекулярних фрагментів, що надають позитивний і негативний вплив на токсичність, і отримання QSAR-моделей, здатних адекватно прогнозувати токсичність нових сполук та сумішей з їх структури і складу. Отримана адекватна консенсусна модель на основі сорока кращих QSAR-моделей (R^2=0,86…0,97; Q^2=0,74…0,96; $R^2_{тест}$=0,86…0,99), які отримані з використанням різних навчальних і тестових вибірок.

Ключові слова: QSAR, симплексне подавання сумішей сполук, дескриптори суміші, токсичність, *Daphnia magna*.

Е.В. Варламова, В.Е. Кузьмин, Н.Н. Муратов, В.А. Шапкин. **Влияние структуры компонентов и состава бинарных смесей органических соединений на токсичность по отношению к *Daphnia Magna***. Симплексное представление молекулярной структуры использовано для консенсусного QSAR-анализа токсичности по отношению к Daphnia Magna натриевых солей α-алкоксикарбонилсульфокислот, алифатических н-спиртов, фенолов и их бинарных смесей. Структура смеси представлена как с использованием дескрипторов индивидуальных соединений, входящих в состав смеси, так и новых специфических параметров смеси, так называемых несвязных симплексов. Исследуемая выборка состояла из 15 индивидуальных соединений и 20 смесей. В качестве исследуемой активности использован логарифм EC_{50} (мкмоль/л). Целью данного исследования является определение молекулярных фрагментов, оказывающих положительное и отрицательное влияние на токсичность, и получение QSAR-моделей, способных адекватно прогнозировать токсичность новых соединений и смесей из их структуры и состава. Получена адекватная консенсусная модель на основе сорока лучших QSAR-моделей (R^2=0,86…0,97; Q^2=0,74…0,96; $R^2_{тест}$=0,86…0,99), которые получены с использованием различных обучающих и тестовых выборок.

Ключевые слова: QSAR, симплексное представление смесей соединений, дескрипторы смеси, токсичность, *Daphnia magna*.

E.V. Varlamova, V.E. Kuz'min, N.N. Muratov, V.A. Shapkin. **Influence of components structure and composition of binary mixtures of organic compounds on toxicity towards *Daphnia Magna.*** Simplex representation of molecular structure is employed for consensus QSAR analysis of toxicity towards Daphnia magna of sodium salts of α-alkoxycarbonylsulfuric acids, aliphatic n-alcohols, phenols and their binary mixtures. The structure of a mixture is represented both using descriptors of individual compounds included in the mixture and using novel specific mixture parameters termed unconnected simplexes. The studied dataset included 15 single compounds and 20 mixtures. The logarithm of EC_{50} (mmole/l) is used as a target function. Aims of the research are the determination of molecular fragments with positive or negative influence on the toxicity and developing QSAR models capable to predict properly the toxicity of new compounds and mixtures from their structure and composition. Successful consensus model based on forty best QSAR models (R^2=0,86…0,97; Q^2=0,74…0,96; R^2_{test}=0,86…0,99) is obtained using different training and test sets.

Keywords: QSAR, simplex representation of compounds mixtures, mixture descriptors, toxicity, *Daphnia magna*.

Received December 8, 2014