

УДК: 544.165:57.016.7:615.21

Г. П. Косінська<sup>1,2</sup>, Л. М. Огніченко<sup>2</sup>, Т. А. Сідельникова<sup>1</sup>, Я. Ф. Бурдіна<sup>1</sup>,  
А. О. Ширікалова<sup>1</sup>, Г. Ф. Степанов<sup>1</sup>, А. В. Грекова<sup>1</sup>, І. Р. Грідіна<sup>1</sup>,  
О. П. Лебідь<sup>1</sup>, В. Є. Кузьмін<sup>2</sup>

<sup>1</sup>Одеський національний медичний університет

Валіховський провулок 2, Одеса, 65082, Україна; e-mail: nikang@ukr.net

<sup>2</sup>Фізико-хімічний інститут ім. О. В. Богатського НАН України

Люстдорфська дорога 86, Одеса, 65080, Україна; e-mail: ogni@ukr.net

## 2D QSAR АНАЛІЗ ПРОНИКАЮЧОЇ ЗДАТНОСТІ ОРГАНІЧНИХ СПОЛУК ВІДНОСНО ПОВЕРХНІ ГЕМАТОЕНЦЕФАЛІЧНОГО БАР'ЄРУ

Побудовано адекватні QSAR моделі «структура – показник проникнення речовин через гематоенцефалічний бар'єр (ГЕБ) LogPS» для органічних сполук різних класів. Проведено структурну та фізико-хімічну інтерпретацію моделей. Виявлено, що проникнення речовин через ГЕБ в значній мірі обумовлено гідрофобним чинником (30%). Визначено деякі фрагменти молекул, що сприяють або перешкоджають проникненню речовин через ГЕБ.

**Ключові слова:** Гематоенцефалічний бар'єр (ГЕБ), QSAR, симплексний підхід, RF, PLS

З усіх гістогематичних бар'єрів (гематоплацентарний, гематоофтальмічний та ін.) найбільшу увагу у вивченні будови, функціонування, зміни бар'єрних показників приділяється гематоенцефалічному бар'єру (ГЕБ). Гематоенцефалічний бар'єр -фізіологічний бар'єр між кровоносною та центральною нервовою системами. Головна функція гематоенцефалічного бар'єру підтримання гомеостазу мозку; крім того, він виконує також функцію високоселективного фільтра, через який з кровоносного русла в мозок надходять поживні речовини, а в зворотному напрямку виводяться продукти життєдіяльності нервової тканини [1]. Для успішної терапії ряду захворювань центральної нервової системи (ЦНС) (хвороба Альцгеймера, інсульти, пухлини мозку та ін.) необхідний швидкий і легкий доступ лікарських засобів з крові в мозок шляхом подолання гематоенцефалічного бар'єру.

Одним з параметрів, що характеризують проникнення речовин через гематоенцефалічний бар'єр, є величина поверхні проникнення (permeability surface) LogPS, де  $P$ ,  $\text{см}\cdot\text{с}^{-1}$  – проникність капілярів;  $S$ ,  $\text{см}^2/\text{г}$  площа поверхні капілярів [2–5].

Метою цього дослідження було побудова QSAR моделей з використанням симплексного підходу [6, 7] та аналіз впливу різних структурних факторів на проникнення речовин через гематоенцефалічний бар'єр для вибірки даних по LogPS.

### МАТЕРІАЛИ І МЕТОДИ ДОСЛІДЖЕННЯ

На початковому етапі роботи об'єктом дослідження стала вибірка з 89 органічних сполук з відомим показником LogPS [3]. В подальшому вибірку було розширено за рахунок додаткового пошуку даних з різних наукових джерел до 177 сполук [4].

DOI: [https://doi.org/10.18524/2304-0947.2022.3\(83\).268607](https://doi.org/10.18524/2304-0947.2022.3(83).268607)

13

© Г. П. Косінська, Л. М. Огніченко, Т. А. Сідельникова, Я. Ф. Бурдіна, А. О. Ширікалова,  
Г. Ф. Степанов, А. В. Грекова, І. Р. Грідіна, О. П. Лебідь, В. Є. Кузьмін, 2022

Для розрахунку структурних дескрипторів використовували 2D симплексне представлення молекулярної структури. В рамках такого підходу кожна молекула розбивається на ряд симплексів – чотириатомних молекулярних фрагментів фіксованої топології структури. При генерації симплексів атоми в молекулі можуть диференціюватися в залежності від типу індивідуальності, а також з урахуванням фізико-хімічних характеристик даних атомів, наприклад, атомних зарядів, ліпофільності, електронегативності, здатності бути донором / акцептором Н зв'язку і тощо. Детально процедуру розрахунку дескрипторів описано в роботах [6, 7].

Для того, щоб встановити зв'язок між структурою досліджуваних сполук та показником LogPS, в даній роботі використовували наступні методи машинного навчання:

- метод часткових найменших квадратів (Partial Least Squares (PLS)) [8], який добре зарекомендував себе при роботі з великими масивами структурної інформації, оскільки ґрунтується на перетворенні великої кількості структурних параметрів в невелику кількість прихованих (латентних) змінних;
- метод випадкового лісу (Random Forest (RF)) [9]; нелінійна модель випадкового лісу являє собою ансамбль окремих дерев рішень.

Відповідно до принципів OECD [10], QSAR моделі повинні володіти адекватною прогностичною здатністю для сполук з невідомою активністю. Для цього пропонується оцінювати «область застосовності» (DA - Domain Applicability) моделі. В даній роботі для оцінки області застосовності (DA) PLS-моделей використовували еліпсоїдний метод [6] та діаграми Вілліямса («Williams plot») [11]. Для RF-моделей було використано локальний деревовидний метод оцінки DA на основі матриці відстаней, розрахованої з урахуванням важливості дескрипторів [12].

Дана робота є першим систематичним дослідженням впливу структури сполук різних класів на параметр LogPS. Як було вказано раніше, на початковому етапі роботи об'єктом дослідження стала вибірка з 89 органічних сполук з відомим показником LogPS. 14 молекул з даного набору були відібрані до вибірки зовнішнього тесту. Для цього молекули були впорядковані відповідно до значень їх LogPS, і потім кожна шоста молекула в цьому ряду була відібрана у зовнішній тест. Для вибірки з решти сімдесяти п'яти молекул була проведена процедура п'ятикратної крос-валідації [6] (внутрішній тест), тобто молекули даної вибірки були упорядковані згідно до значень їх властивостей і потім кожна п'ята молекула в даному ряду відібрана в тестову вибірку. Таким чином, було сформовано п'ять наборів, кожний з яких містив навчальну і тестову вибірку. Слід зазначити, що серед отриманих п'яти наборів кожна молекула потрапляла в тестову вибірку тільки один раз. Тестова вибірка (20% від загальної кількості молекул) виключалась з процесу побудови моделі. На даних навчальної вибірки будувалась модель, яка використовувалась для прогнозу значень LogPS сполук тестової вибірки.

Для виключення можливості помилкової кореляції було проведено процедуру *Y – Scrambling* – створення моделі методом випадкового підбору значень властивостей [6].

## РЕЗУЛЬТАТИ ТА ЇХ ОБГОВОРЕННЯ

В результаті дослідження отримано п'ять адекватних 2D PLS QSAR моделей (М1-М5), які ввійшли в основу консенсусної моделі М6. Статистичні показники побудованих моделей представлено в таблиці 1. Для консенсусної PLS-моделі М6 коефіцієнт детермінації для тестової вибірки ( $R^2_{\text{ts}}$ ) дорівнює 0.78.

Таблиця 1

Статистичні характеристики 2D PLS QSAR моделей

Table 1

Statistical characteristics of the 2D PLS QSAR models

Модель	Кількість дескрипторів	Навчальна вибірка			Тестова вибірка		Y-Scrambling	
		$R^2$	S	$Q^2$	$R^2_{\text{test}}$	$S_{\text{test}}$	$R^2_{\text{Y-Ser}}$	$Q^2_{\text{Y-Ser}}$
M1	36	0.92	0.36	0.82	0.88	0.37	0.18±0.02	0.08±0.02
M2	44	0.89	0.39	0.78	0.84	0.47	0.21±0.03	0.10±0.02
M3	23	0.84	0.46	0.79	0.72	0.62	0.15±0.02	0.04±0.02
M4	32	0.88	0.39	0.78	0.67	0.73	0.18±0.02	0.08±0.02
M5	27	0.88	0.39	0.82	0.81	0.57	0.08±0.02	0.08±0.02
Консенсусна модель М6		0.91			0.78	0.55		

де  $R^2$  – коефіцієнт детермінації,  $Q^2$  – коефіцієнт детермінації в умовах ковзного контролю (leave-one-out), S – стандартна помилка.

Як видно з таблиці 1, величини статистичних характеристик моделей, отриманих при використанні процедури *Y-Scrambling*, значно нижче за значенням, ніж у робочих моделей, що підтверджує не випадковість встановленого зв'язку між структурою сполук і величиною LogPS.

Побудовані за допомогою симплексного підходу моделі можуть бути легко інтерпретовані [6,7]. Для консенсусної PLS моделі М6 проведено аналіз внесків різних молекулярних фрагментів у властивість, що досліджується, були виділені деякі фрагменти молекул, що сприяють або перешкоджають проникненню речовин через ГЕБ (таблиця 2). Як видно з таблиці 2, негативно впливає на здатність молекул проникати через ГЕБ наявність полярних груп, а також донорів водневого зв'язку. Водночас наявність ароматичних фрагментів, атомів галогенів, а також інших фрагментів, що підвищують ліпофільність молекули, сприяє проникненню сполук через ГЕБ.

Використовуючи консенсусну PLS модель М6 було проведено зовнішнє тестування для чотирнадцяти молекул (таблиця 3). Отримані результати свідчать про високу прогностичну здатність консенсусної моделі ( $R^2_{\text{ex.ts}} = 0.82$ ).


Далі вибірку було розширено внаслідок додаткового пошуку даних з різних наукових джерел до 177 сполук. Побудовано QSAR RF-модель М7 (кількість дерев 500, кількість випадково відібраних змінних 70), яка характеризується задовільною прогностичною здатністю, що оцінена відповідно до внутрішньої процедури валі-

Таблиця 2

## Молекулярні фрагменти, які впливають на величину LogPS

Table 2

## The molecular fragments which increase and decrease LogPS

Сприяють LogPS:	
$\text{---F}$ ; $\text{---Cl}$ ; $\text{---Alk}$ ; $\text{---N(C}_2\text{H}_5)_2$ ; $\text{---}$ 	
Перешкоджають LogPS:	
$\text{---COOH}$ ; $\text{---S(=O)}_2\text{---}$ ; $\text{---C(=O)---}$ ; $\text{---C(=O)NH---}$ ; $\text{---NH}_2$ ; $\text{---OH}$	

дації RF моделей [9]. Коефіцієнт детермінації  $R^2_{\text{ob}}$  і середньоквадратична помилка прогнозу  $\text{RMSE}_{\text{ob}}$ , для «out-of-bag» вибірки дорівнюють 0.61 і 0.7, відповідно. Для навчальної вибірки отримана RF-модель **M7** має такі статистичні показники:  $R^2_{\text{ws}} = 0.97$ ,  $\text{RMSE}_{\text{ws}} = 0.26$ . На рис. 1 представлено залежність «спостережувані – розраховані» значення LogPS для навчальної вибірки. Всі молекули знаходяться в області застосовності RF-моделі **M7**.

На основі аналізу структурних параметрів, які увійшли до RF-моделі **M7**, було виявлено, що проникнення речовин через ГЕБ в значній мірі обумовлено

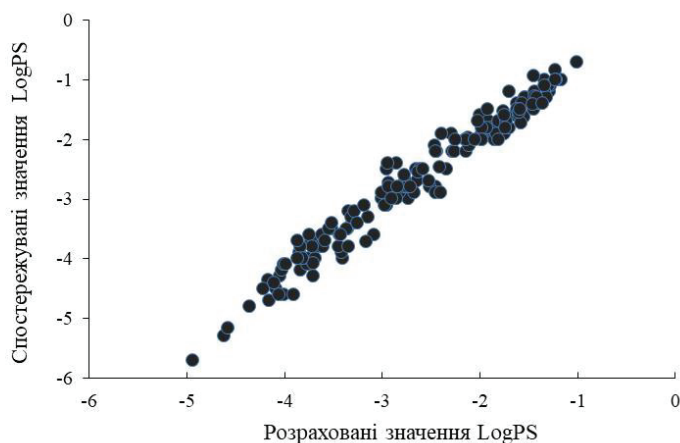


Рис. 1. Діаграма спостережувані – розраховані значення LogPS для навчальної вибірки (RF-модель **M7**)

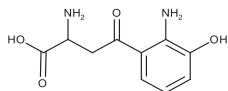

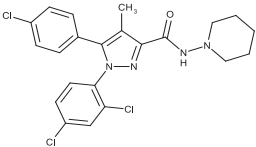
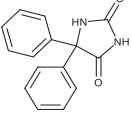
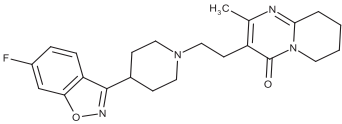
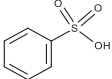
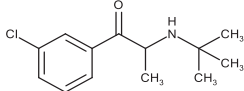
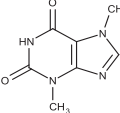
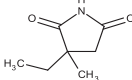
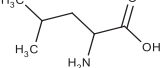
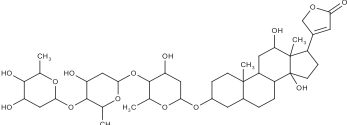

Fig. 1. Observed vs calculated values of LogPS for work set (RF-model **M7**)

Таблиця 3

Прогноз LogPS для зовнішньої тестової вибірки

Table 3

The predicted values of LogPS for external test set

Молекула	LogPS		Молекула	LogPS	
	Експ.	Прогноз		Експ.	Прогноз
 <i>3-hydroxykyurenine</i>	-4.49	-4.32	 <i>Amantadine</i>	-2.71	-2.45
 <i>Acompia</i>	-1.60	-1.23	 <i>Phenytoin</i>	-2.07	-2.75
 <i>Risperidone</i>	-1.81	-1.73	 <i>Benzenesulfonic acid</i>	-2.11	-1.72
 <i>Bupropion</i>	-1.52	-1.88	 <i>Theobromine</i>	-3.00	-3.25
 <i>Ethosuximide</i>	-2.47	-2.51	 <i>L-Leucine</i>	-4.08	-3.35
 <i>Digoxin</i>	-4.30	-4.65	 <i>L-Glycine</i>	-3.50	-3.70

гідрофобним чинником (30%) (рис. 2). Більший внесок ліпофільності можна пояснити хімічною природою ендотелію, що складається з фосfolіпідів, що пояснює зростання здатності молекул долати фізіологічні бар'єри зі збільшенням їх ліпофільності. Природа атомів, електростатичний фактор, здатність бути донором / акцептором потенційного водневого зв'язку і ван-дер-ваальсові взаємодії дають декілька нижчий внесок у зміну властивості – 24%, 16%, 16% і 14%, відповідно. Вочевидь, розподіл зарядів на атомах істотно впливає на проникнення речовин через полярні зони на поверхні ГЕБ, а ван-дер-ваальсові взаємодії цілком ймовірно визначають орієнтацію молекул гідрофобними центрами до поверхні ендотелію ГЕБ.



Рис. 2. Вплив фізико-хімічних чинників на показник LogPS досліджуваних сполук (RF-модель M7)

Fig. 2. Relative influences of some physical-chemical factors on LogPS of investigated compounds (RF-model M7)

Таким чином, розроблено адекватні й інтерпретовані моделі «структура – LogPS», що володіють високою прогнозуючою здатністю. Виявлено, що проникнення речовин через ГЕБ в значній мірі обумовлено гідрофобним чинником (30%). Визначено деякі фрагменти молекул, що сприяють або перешкоджають проникненню речовин через ГЕБ. Негативно впливає на здатність молекул проникати ГЕБ наявність полярних груп, а також донорів/акцепторів водневого зв'язку ( $-\text{COOH}$ ,  $-\text{NH}_2$ ,  $-\text{OH}$ ,  $-\text{S}(\text{O})_2-$ ,  $-\text{CO}-$ ,  $-\text{NH}-\text{CO}-$ ). Водночас наявність ароматичних фрагментів, атомів галогенів ( $-\text{F}$ ,  $-\text{Cl}$ ), а також інших фрагментів (алкіл,  $-\text{N}(\text{C}_2\text{H}_5)_2$ ), що підвищують ліпофільність молекули, сприяє проникненню сполук через ГЕБ. Побудовані при виконанні даної роботи QSAR моделі ввійшли в експертну систему для прогнозування показників проникнення органічних речовин через ГЕБ.

## СПИСОК ЛІТЕРАТУРИ

1. *Partridge W.M.* Blood-Brain Barrier delivery // *Drug Discovery Today*. – 2007. – Vol. 12, N1–2. – P. 4–61. <http://doi.org/10.1016/j.drudis.2006.10.013>
2. *Dagenais C., Avdeef A., Tsinman O., Dudley A.* P-Glycoprotein Deficient Mouse in situ Blood-Brain Barrier Permeability and its Prediction using an in combo PAMPA Model // *Eur. J. Pharm. Sci.* – 2009. – Vol. 38, N2. – P. 121–137. <http://doi.org/10.1016/j.ejps.2009.06.009>
3. *Murakami H., Takanaga, Matsuo H., Ohtani H., Sawada Y.* Comparison of blood-brain barrier permeability in mice and rats using in situ brain perfusion technique // *Am. J. Phys. Heart Circ. Physiol.* – 2000. – P. 1022–1029. <http://doi.org/10.1152/ajpheart.2000.279.3.H1022>
4. *Lanevskij K., Japertas P., Didziapetris R., Petrauskas A.* Ionization-Specific prediction of Blood-Brain Permeability // *J. Pharm. Sci.* – 2009. – Vol. 98, N1. – P. 122–134. <http://doi.org/10.1002/jps.21405>
5. *Suenderhauf C., Hammann F., Huwyler J.* Computational Prediction of Blood-Brain Barrier Permeability Using Decision Tree Induction Huwyler // *Mol.* – 2012. – Vol. 17, N9. – P. 10429–10445. <http://doi.org/10.3390/molecules170910429>
6. *Kuz'min V.E., Artemenko A. G., Muratov E. N., Polischuk P. G., Ognichenko L. N., Liahovsky A. V., Hromov A. I., Varlamova E. V.* Virtual screening and molecular design based on hierarchical QSAR technology. In: *Recent Advances in QSAR Studies*. Eds. Puzyn T, Leszczynski J, Cronin M. – London: Springer, 2010. – P. 127–176. [http://doi.org/10.1007/978-1-4020-9783-6\\_5](http://doi.org/10.1007/978-1-4020-9783-6_5)
7. *Kuz'min V., Artemenko A., Ognichenko L., Hromov A., Kosinskaya A., Stelmakh S., Sessions Z., Muratov E.* Simplex representation of molecular structure as universal QSAR/QSPR tool // *J. Struct. Chem.* – 2021. – Vol. 32, N4. – P. 1365–1392. <http://doi.org/10.1007/s11224-021-01793-z>
8. *Rannar S., Lindgren F., Geladi P., Wold S.* A PLS kernel algorithm for data sets with many variables and fewer objects. Part 1: Theory and algorithm // *J. Chemom.* – 1994. – Vol. 8, N2. – P. 111–125. <https://doi.org/10.1002/cem.1180080204>
9. *Breiman L.* Random Forests // *Machine Learning*. – 2001. – Vol. 45. – P. 5–32. <http://doi.org/10.1023/A:1010933404324>
10. *Gramatica P.* Principles of QSAR models validation: internal and external // *QSAR Comb. Sci.* – 2007. – Vol. 26, N5. – P. 694–701. <http://doi.org/10.1002/qsar.200610151>
11. *Meloun M., Militku J., Hill M.* Crucial problems in regression modelling and their solutions // *Analyst.* – 2002. – Vol. 127. – P. 433–450. <http://doi.org/10.1039/b110779h>
12. *Polischuk P.G., Muratov E.N., Artemenko A.G., Kolumbin O.G., Muratov N.N., Kuz'min V.E.* Application of Random Forest method to QSAR prediction of aquatic toxicity // *J. Chem. Inf. Mod.* – 2009. – Vol. 49. – P. 2481–2488. <http://doi.org/10.1021/ci900203n>

Стаття надійшла до редакції 14.11.2022

**A. P. Kosinskaya<sup>1,2</sup>, L. N. Ognichenko<sup>2</sup>, T. A. Sidelnikova<sup>1</sup>, Ya. F. Burdina<sup>1</sup>,  
A. A. Shyrykalova<sup>1</sup>, G. F. Stepanov<sup>1</sup>, A. V. Grekova<sup>1</sup>, I. R. Gridina<sup>1</sup>, E. P. Lebed<sup>1</sup>,  
V. E. Kuz'min<sup>2</sup>**

<sup>1</sup>Odessa National Medical University; 2 Valikhovskiy lane, 65082, Odessa, Ukraine;  
[nikang@ukr.net](mailto:nikang@ukr.net)

<sup>2</sup>A.V. Bogatsky Physico-Chemical Institute of NAS of Ukraine; 86 Lustdorfskaya Doroga  
Str., 65080, Odessa, Ukraine; [ogni@ukr.net](mailto:ogni@ukr.net)

## 2D QSAR ANALYSIS OF BLOOD-BRAIN BARRIER PERMEABILITY FOR ORGANIC COMPOUNDS

The aim of this work is to analyze the influence of the structure of organic compounds on their ability to penetrate the blood-brain barrier (BBB) using the values of LogPS (logarithm of the product of permeability and surface area of metabolism, which is a measure of the transfer of compounds from the blood to the brain and reflects the degree of penetration of the substance into the brain without relative binding to proteins). The study of the permeability of BBB is necessary both for the development of new drugs for which the CNS is a biointerference, and for the development of effective ways of treating diseases of the brain.

At the initial stage of the study, the object of this work was the database of 89 organic compounds with a known value LogPS. Subsequently, the database was expanded to 177 compounds by additional data retrieval from various scientific sources. For all investigated compounds the structural parameters were calculated using the SiRMS method, which is based on the simplex representation of the molecular structure. QSAR models were developed using the “Partial Least Squares method” (PLS) and Random Forest (RF).

The adequate «structure – LogPS» 2D-QSAR models were obtained for organic compounds of different classes. Physico-chemical and structural interpretation of the of the obtained QSAR models was carried out. Among various physical and chemical factors, the hydrophobic factors have the greatest influence on BBB permeability (30%). The study proved that polar groups and donor/acceptor of H-bond (–COOH, –NH<sub>2</sub>, –OH, –S(O)<sub>2</sub>–, –CO–, –NH–CO–) negatively affect the penetration ability of compounds through BBB. Contrary, aromatic fragments, halogens (–F, –Cl) and such fragments as different alkyls and –N(C<sub>2</sub>H<sub>5</sub>)<sub>2</sub> have a positive effect on the penetration. The developed QSAR models were included in the expert system to predict the penetration of substances through the BBB.

**Keywords:** Blood-brain-barrier (BBB), QSAR, simplex approach, RF, PLS.

## REFERENCE

1. Pardridge W.M. *Blood-Brain Barrier delivery*. Drug Discovery Today, 2007, vol. 12, no 1–2, pp. 54–61. <http://doi.org/10.1016/j.drudis.2006.10.013>
2. Dagenais C., Avdeef A., Tsinman O., Dudley A. *P-Glycoprotein Deficient Mouse in situ Blood-Brain Barrier Permeability and its Prediction using an in combo PAMPA Model*. Eur. J. Pharm. Sci., 2009, vol. 38, no 2, pp. 121–137. <http://doi.org/10.1016/j.ejps.2009.06.009>
3. Murakami H., Takanaga H., Matsuo H., Ohtani H., Sawada Y. *Comparison of blood-brain barrier permeability in mice and rats using in situ brain perfusion technique*. Am. J. Phys. Heart Circ. Physiol., 2000, pp. 1022–1029. <http://doi.org/10.1152/ajpheart.2000.279.3.H1022>
4. Lanevskij K., Japertas P., Didziapetris R., Petrauskas A. *Ionization-Specific prediction of Blood-Brain Permeability*. J. Pharm. Sci., 2009, vol. 98, no 1, pp. 122–134. <http://doi.org/10.1002/jps.21405>
5. Suenderhauf C., Hammann F., Huwyler J. *Computational Prediction of Blood-Brain Barrier Permeability Using Decision Tree Induction Huwyler*. Mol., 2012, vol. 17, no 9, pp. 10429–10445. <http://doi.org/10.3390/molecules170910429>
6. Kuz'min V.E., Artemenko A.G., Muratov E.N., Polischuk P.G., Ognichenko L.N., Liahovsky A.V., Hromov A.I., Varlamova E.V. *Virtual screening and molecular design based on hierarchical QSAR technology*. In: Recent Advances in QSAR Studies. Eds. Puzyn T, Leszczynski J, Cronin M., London, Springer, 2010, pp. 127–176. [http://doi.org/10.1007/978-1-4020-9783-6\\_5](http://doi.org/10.1007/978-1-4020-9783-6_5)
7. Kuz'min V., Artemenko A., Ognichenko L., Hromov A., Kosinskaya A., Stelmakh S., Sessions Z., Muratov E. *Simplex representation of molecular structure as universal QSAR/QSPR tool*. J. Struct. Chem., 2021, vol. 32, pp. 1365–1392. <http://doi.org/10.1007/s11224-021-01793-z>
8. Rannar S., Lindgren F., Geladi P., Wold S. *A PLS kernel algorithm for data sets with many variables and fewer objects. Part I: Theory and algorithm*. J. Chemom., 1994, vol. 8, no 2, pp. 111–125. <http://doi.org/10.1002/cem.1180090604>
9. Breiman L. *Random Forests*. Machine Learning, 2001, vol. 45, pp. 5–32. <http://doi.org/10.1023/A:1010933404324>
10. Gramatica P. *Principles of QSAR models validation: internal and external*. QSAR Comb. Sci., 2007, vol. 26, no 5, pp. 694–701. <http://doi.org/10.1002/qsar.200610151>
11. Meloun M., Militku J., Hill M. *Crucial problems in regression modelling and their solutions*. Analyst, 2002, vol. 127, pp. 433–450. <http://doi.org/10.1039/b110779h>
12. Polischuk P.G., Muratov E.N., Artemenko A.G., Kolumbin O.G., Muratov N.N., Kuz'min V.E. *Application of Random Forest method to QSAR prediction of aquatic toxicity*. J. Chem. Inf. Mod., 2009, vol. 49, no 11, pp. 2481–2488. <http://doi.org/10.1021/ci900203n>