

В.В. Важев<sup>1</sup>, М.А. Губенко<sup>2</sup>, Н.В. Важева<sup>2</sup>, Э.М. Ергалиева<sup>2</sup>, Е.А. Самсонюк<sup>3</sup>

<sup>1</sup>Костанайский социально-технический университет им. З. Алдамжара;

<sup>2</sup>Костанайский государственный педагогический институт;

<sup>3</sup>Костанайский государственный университет им. А.Байтурсынова  
(E-mail: vladimir.vazhev@gmail.com)

## Оценка токсичности органических соединений для *Tetrahymena pyriformis* методом QSAR

В статье рассмотрена возможность применения QSAR/QSTR метода для оценки показателя токсичности органических веществ  $-\lg(\text{IGC}_{50})$  по отношению к инфузории *Tetrahymena pyriformis* с использованием большого набора дескрипторов. Исследован набор веществ из 1093 соединений. Для всего набора получены показатели корреляции  $R = 0,968$ ,  $s = 0,26$ . Для тренировочной выборки  $R = 0,980$ ,  $s = 0,21$ , для контрольной  $R = 0,943$ ,  $s = 0,32$ .

**Ключевые слова:** органические соединения, токсичность, QSAR.

К настоящему времени синтезировано свыше пятидесяти миллионов химических соединений, десятки тысяч из них интенсивно вовлекаются в сферу практического использования. Определение различных свойств (физико-химических, разных видов биологической активности) этих веществ экспериментальным путем часто связано со значительными трудностями, возникающими, например, при получении достаточного количества вещества, его очистки, возможной нестойкости, токсичности и т.д. Кроме того, это не всегда представляется возможным. Такие исследования требуют значительных финансовых и временных затрат. В этой связи актуальным является создание надежных и достоверных расчетных методов оценки токсичности химических соединений. Поэтому все большее значение приобретают методы определения токсичности веществ путем расчетов. Наиболее перспективным из них считается QSAR/QSTR — метод, устанавливающий корреляционные соотношения структура–активность и структура–токсичность.

Достаточно удобным объектом для установления количественных соотношений структура – активность является инфузория *Tetrahymena pyriformis*. Она имеет высокую скорость роста, что позволило накопить для нее достаточно надежные и обширные банки данных о токсичности [1]. Показатели токсичности веществ, полученные на этих организмах, соответствуют таковым для теплокровных животных и человека. Токсичность веществ по отношению к *Tetrahymena pyriformis* часто выражают в единицах  $-\lg(\text{IGC}_{50})$ , где  $\text{IGC}_{50}$  — концентрация вещества в ммоль/л, вызывающая гибель половины особей популяции. Качество прогнозирования характеризуется коэффициентом корреляции  $R$  между прогнозируемыми и экспериментальными значениями параметра токсичности и стандартным отклонением  $s$ .

Вопросам моделирования и прогнозирования токсичности для *Tetrahymena pyriformis* посвящены десятки статей. Так, в работе [2] для прогнозирования токсичности 54 производных нитробензола в качестве дескрипторов использовались молекулярная масса, молярная рефракция, сродство к электрону и общая энергия. С их помощью было построено 8 моделей, которые также были проверены методом перекрестной проверки. Для первой модели использовался набор дескрипторов общей энергии, качество прогнозирования характеризовалось коэффициентом корреляции  $R^2 = 0,8424$ , стандартное отклонение  $s = 0,06$ , коэффициент перекрестной проверки  $R^2 = 0,8172$ . Вторая монопараметрическая модель QSTR, полученная при помощи дескрипторов сродства к электрону, характеризовалась следующими параметрами:  $R^2 = 0,74$ ;  $s = 0,08$ . В роли дескрипторов для третьей модели выступали дескрипторы общей энергии и сродства к электрону. Полученные результаты:  $R^2 = 0,89$ ;  $s = 0,05$ . Дескрипторы молярной рефракции и сродства к электрону использовались для построения четвертой модели, результаты  $R^2 = 0,88$ ;  $s = 0,05$ . Пятая модель была построена с помощью набора таких дескрипторов, как молярная рефракция, сродство к электрону, общая энергия. В результате были получены следующие данные:  $R^2 = 0,91$ ;  $s = 0,04$ . Результаты шестой модели:  $R^2 = 0,90$ ;  $s = 0,05$ . В качестве дескрипторов служили молекулярная масса, сродство к электрону, а также общая энергия. Далее использовался набор из четырех дескрипторов. Для седьмой модели были взяты дескрипторы молекулярной массы, молярной рефракции, сродства к электрону и общей энергии. Для данной модели

$R^2 = 0,91$ ;  $s = 0,04$ . И, наконец, в качестве дескрипторов для восьмой модели были использованы дескрипторы молекулярной массы, молярной рефракции, электроотрицательности и общей энергии. Результаты данной модели:  $R^2 = 0,91$ ;  $s = 0,04$ . Это лучшие результаты.

При исследовании небольшого набора из 48 производных анилина, из которых 31 вещество вошло в тренировочную, а 17 веществ — в контрольную выборки, были получены следующие результаты:  $R^2 = 0,95$ ,  $s = 0,16$  [3].

В работе [4] представлены 7 моделей прогнозирования 50 нитробензолов. Для лучшей модели значение коэффициента корреляции  $R^2 = 0,87$ , стандартное отклонение  $s = 0,27$ .

Авторы [5] построили 8 моделей прогнозирования водной токсичности 250 фенолов по отношению к инфузории *Tetrahymena pyriformis*. Из них 187 веществ вошли в тренировочную выборку и 63 — в контрольную. То есть набор веществ контрольной выборки почти в три раза меньше набора веществ, составивших тренировочную выборку. В качестве дескрипторов использовались коэффициенты распределения в системе октанол–вода, а также дескрипторы программного комплекса Molconn-Z. Для лучшей модели коэффициент корреляции  $R^2 = 0,71$ . Однако для данной модели 9 веществ из тренировочной и 2 вещества из контрольной выборок были идентифицированы как выбросы, и, в конечном счете, объем тренировочной выборки составил 178 веществ, а контрольной — 61 вещество.

Моделирование токсичности 250 фенолов выполнили авторы работы [6]. Помимо контрольной (50 веществ) и тренировочной выборок (105 веществ), использовалась также калибровочная выборка (95 веществ). Лучшая модель получена с результатами  $R^2 = 0,77$ ,  $s = 0,41$ ;  $R^2 = 0,77$ ,  $s = 0,39$ ;  $R^2 = 0,77$ ,  $s = 0,40$  для тренировочной, калибровочной и контрольной выборок соответственно.

Для небольшого набора из 48 сложных эфиров использовались WHIM дескрипторы, генерируемые программой Dragon [7]. Авторы представили 6 моделей, которые очень близки по значениям. Коэффициент корреляции для наилучшей модели  $R^2 = 0,921$ , стандартное отклонение  $s = 0,240$ . Качество прогнозирования данных моделей было проверено методом скользящего контроля,  $R^2 = 0,911$ .

Авторы [8], изучая токсичность 250 фенолов, 200 веществ включили в тренировочную выборку и лишь 50 веществ — в проверочную. Ими использовались методы линейной регрессии (MLR) и нейронных сетей. Как для тренировочной, так и для контрольной выборок результаты модели нейронных сетей оказались немного лучше и характеризовались показателем  $R^2 = 0,71$ .

Прогнозирование токсичности 1093 органических соединений для инфузории *Tetrahymena pyriformis* выполнили в сотрудничестве авторы из шести академических групп, специализирующихся на химической информатике и компьютерной токсикологии [9]. Из них 644 соединения составили тренировочную выборку, 339 — первый проверочный набор и 110 — второй проверочный набор. Авторами было составлено в общей сложности 15 моделей. Для лучшей модели (consensus model)  $R^2_{\text{abs}} = 0,93$ , (тренировочная выборка);  $R^2_{\text{abs}} = 0,88$ , (проверочный набор I);  $R^2_{\text{abs}} = 0,77$ , (проверочный набор II).

В таблице 1 приведены литературные данные по качеству прогнозирования  $-\lg(\text{IGC}_{50})$  для *Tetrahymena pyriformis* за последние годы.

Т а б л и ц а 1

**Результаты прогнозирования  $-\lg(\text{IGC}_{50})$  различных классов веществ для *Tetrahymena pyriformis* по литературным данным за 2008–2015 гг.**

Лит.	Вещества	Общее количество веществ	$R^2$	$s$
[2]	Производные нитробензола	54	0,91	0,04
[3]	Производные анилина	48	0,95	0,16
[4]	Производные нитробензола	50	0,87	0,27
[5]	Производные фенола	250	0,71	–
[6]	Производные фенола	250	0,77	0,40
[7]	Сложные эфиры	48	0,92	0,24
[8]	Производные фенола	250	0,71	–
[9]	Органические соединения	1093	0,93	–

Приведенные данные в достаточной мере характеризуют современное состояние прогнозирования  $-\lg(\text{IGC}_{50})$  для *Tetrahymena pyriformis*. Качество лучших моделей определено значениями коэффициентов корреляции в пределах от 0,71 до 0,95 и значениями стандартных отклонений  $s$  до 0,40. Очевидно, что дальнейший прогресс в этом направлении связан с поиском более информативных дескрипторов, используемых при кодировании структуры молекул, поскольку биологический отклик является функцией большого и неопределенного числа факторов.

В данной работе исследована возможность использования большого набора дескрипторов, генерируемых программой Dragon. Показатели токсичности  $-\lg(\text{IGC}_{50})$  для *Tetrahymena pyriformis* получены из приложения к статье [9]. Тренировочная выборка, так же как и в [9], состояла из 644 соединений, контрольная — из 493.

Расчеты выполнены с помощью разработанной нами [10] компьютерной программы PROGROC (PROGgram ROBustness Calculation). Важной и уникальной особенностью этой программы является возможность привлечения без предварительного отбора такого числа дескрипторов, которое может превышать число веществ. Так, для данного набора веществ всего было задействовано 1017 дескрипторов различной природы. Это позволило получить модель более высокого качества, которая характеризовалась  $R = 0,980$ ,  $s = 0,22$  для тренировочной выборки того же размера и  $R = 0,943$ ,  $s = 0,32$  — для контрольной. Также была получена модель высокого качества при уменьшении размера тренировочной выборки до 600 соединений (493 — контрольная). Результаты прогнозирования  $-\lg(\text{IGC}_{50})$  для этого случая приведены рисунках 1, 2.

На рисунке 1 показана корреляция между экспериментальными и вычисленными по молекулярным дескрипторам значениями  $-\lg(\text{IGC}_{50})$  для тренировочной и контрольной выборок.

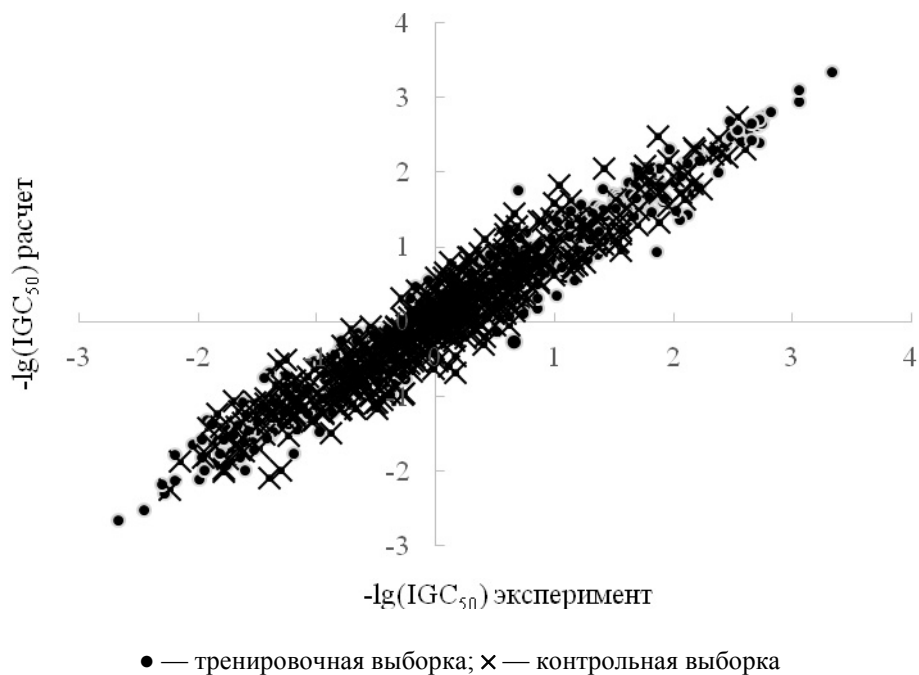


Рисунок 1. Корреляция между экспериментальными значениями  $-\lg(\text{IGC}_{50})$  и вычисленными по молекулярным дескрипторам для тренировочной и контрольной выборок

Частота распределения ошибок прогнозирования  $-\lg(\text{IGC}_{50})$  с помощью молекулярных дескрипторов показана на гистограмме (рис. 2).

Гистограмма распределения частот ошибок прогнозирования  $\lg(\text{IGC}_{50}-1)$  несколько отличается от нормального распределения, что, вероятно, вызвано низкой точностью использованных экспериментальных данных.

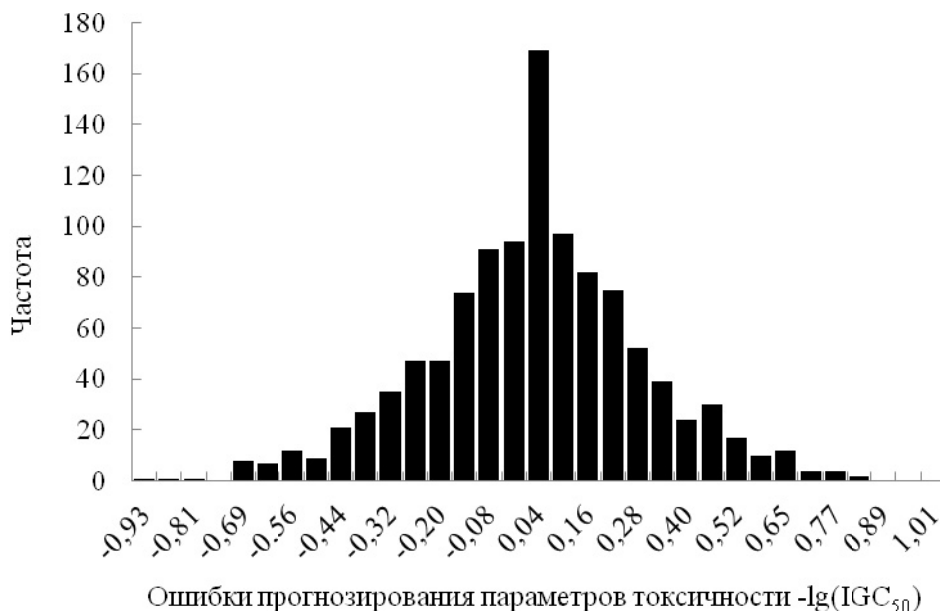


Рисунок 2. Гистограмма распределения частот ошибок прогнозирования  $-\lg(\text{IGC}_{50})$  по молекулярным дескрипторам

Статистические параметры модели приведены в таблице 2.

Таблица 2

**Показатели корреляции между экспериментальными и вычисленными значениями  $-\lg(\text{IGC}_{50})$**

Показатели корреляции	Весь набор	Тренировочная выборка	Контрольная выборка
$R$	0,964	0,979	0,938
$s$	0,28	0,22	0,33
Количество веществ	1093	600	493

Таблица 3 относится к части контрольной выборки и иллюстрирует порядки значений  $-\lg(\text{IGC}_{50})$ , точность прогнозирования и разнообразие классов химических соединений.

Таблица 3

**Экспериментальные [9] и рассчитанные значения  $-\lg(\text{IGC}_{50})$  для *Tetrahymena pyriformis***

№	Соединение	Эксп.	Расч.	Разн.
1	2	3	4	5
1	2-Этилмасляная кислота	-0,15	-0,30	0,15
2	2-Метил-4-хлоранилин	0,35	0,47	-0,12
3	Толуол	0,25	-0,38	0,63
4	Диэтил-2-этилмалонат	-0,24	-0,09	-0,15
5	3-Амино-1,2-пропандиол	-1,81	-1,51	-0,30
6	4-Хлор-1-бутанол	-0,76	-0,83	0,07
7	2-Нитрорезорцинол	0,66	0,28	0,38
8	4-Бифенилметанол	0,92	0,80	0,12
9	3-Гексин-3,5-диол	-0,46	-0,94	0,48
10	3'-Нитроацетофенон	0,32	0,43	-0,11
11	1,2,3-Трихлор-5-нитробензол	1,55	1,47	0,08
12	2-Хлорбензальдегид	0,49	0,41	0,08
13	3-Хлор-2-бутанон	0,27	0,28	-0,01
14	Этенилэтаноат	-0,86	-0,69	-0,17
15	3-Нитро-4-хлорфенол	1,27	0,91	0,36
16	Диэтиловый эфир пропилмалоновой кислоты	0,13	-0,11	0,24
17	2,6-Дихлорфенол	0,73	0,77	-0,04

1	2	3	4	5
18	4-Нитро-2-хлоранилин	0,75	0,50	0,25
19	5-Нитро-2-хлорбензальдегид	0,53	1,28	-0,75
20	3,5-Диметил-1-гексин-3-ол	-0,55	-0,13	-0,42
21	4,5-Диметил-2-хлорфенол	0,69	0,89	-0,20
22	4-Фенил-1-бутанол	0,12	0,09	0,03
23	1,3-Динитро-2,4,5-трихлорбензол	2,60	2,37	0,23
24	2-Гидрокси-4,5-диметилацетофенон	0,71	1,01	-0,30
25	3,5-Дихлорнитробензол	1,13	1,61	-0,48
26	2-Нитро-4-хлортолуол	0,82	1,03	-0,21
27	<i>втор</i> -Бутиловый эфир уксусной кислоты	-0,68	-1,10	0,42
28	Метилнонилат	1,04	1,16	-0,12
29	3,5-Гептадион	-0,38	-0,11	-0,27
30	2-Метил-5-хлоранилин	0,50	0,48	0,02
31	1-Ноиламин	1,70	1,17	0,53
32	4-Нитробензальдегида	0,20	0,56	-0,36
33	1-Октил цианид	0,62	0,34	0,28
34	2,4-Дихлоранилин	0,56	0,72	-0,16
35	Терефталдегид	-0,09	0,18	-0,27
36	3-Формилбензонитрил	-0,02	-0,33	0,31
37	<i>трет</i> -Пентиламин	-0,70	-1,40	0,70
38	4-( <i>втор</i> -Бутил)анилин	0,61	0,95	-0,34
39	2-Тридецин-1-ол	2,37	2,15	0,22
40	Гептанофенон	1,56	1,13	0,43

Результаты нашего исследования показывают, что применение полного набора дескрипторов, вычисляемых программой Dragon в сочетании с разработанным нами алгоритмом прогнозирования, позволяет получать модели параметра токсичности IGC<sub>50</sub> в отношении к инфузории *Tetrahymena pyriformis* с высоким качеством, превосходящим соответствующие модели, описанные в литературе.

### Список литературы

- 1 The University of Tennessee: College of veterinary medicine // [ER]. Access mode: <http://www.vet.utk.edu/research/chemical.shtml>
- 2 Kumar M.S., Vibhanjali M., Tripathi P.N., Adil K.M. Quantum Chemical Descriptors Based QSTR Study of Nitrobenzene Derivatives against Tetrahymena Pyriformis // Research Journal of Chemical Sciences. — 2014. — Vol. 4, No. 2. — P. 29–37.
- 3 Bouaoune A., Lourici L., Haddag H., Messadi D. Inhibition of Microbial Growth by Anilines: A QSAR Study // Journal of Environmental Science and Engineering. — 2012. — A 1. — P. 663–671.
- 4 Bellifa K., Mekelleche S.M. QSAR study of the toxicity of nitrobenzenes to Tetrahymena pyriformis using quantum chemical descriptors // Arabian Journal of Chemistry. — 2012 // [ER]. Access mode: <http://dx.doi.org/10.1016/j.arabjc.2012.04.031>.
- 5 Jiang D.X., Li Y., Li J., Wang G.X. Prediction of the Aquatic Toxicity of Phenols to Tetrahymena Pyriformis from Molecular Descriptors // Int. J. Environ. Res. — 2011. — Vol. 5, No. 4. — P. 923–938.
- 6 Toropov A.A., Toropova A.P., Benfenati E., Manganaro A. QSAR modelling of the toxicity to Tetrahymena pyriformis by balance of correlations // Mol. Divers. — 2010. — Vol. 14. — P. 821–827.
- 7 Vlaia V., Olariu T., Vlaia L., Butur M., Ciubotariu C., Medeleanu M., Ciubotariu D. Quantitative structure-activity relationship (QSAR) Analysis of the toxicity of aliphatic esters by means of weighted holistic invariant molecular (whim) descriptors // Farmacia. — 2009. — Vol. 57, No. 4. — P. 511–522.
- 8 Enoch S.J., Cronin M.T.D., Schultz T.W., Madden J.C. An evaluation of global QSAR models for the prediction of the toxicity of phenols to Tetrahymena pyriformis // Chemosphere. — 2008. — Vol. 71, No. 7. — P. 1225–1232.
- 9 Zhu H., Tropsha A., Fourches D., Varnek A., Papa E., Gramatica P., Öberg T., Dao P., Cherkasov A., Tetko I. Combinatorial QSAR Modeling of Chemical Toxicants Tested against Tetrahymena pyriformis // J. Chem. Inf. Model. — 2008. — Vol. 48. — P. 766–784.
- 10 Важев В.В., Кухарева А.Ю., Цебрук А.В. Компьютерное прогнозирование и экспертиза значений температуры кипения спиртов по ИК-спектрам // Вестн. Караганд. ун-та. Сер. Химия. — 2010. — № 2 (58). — С. 5–10.

В.В. Важев, М.А. Губенко, Н.В. Важева, Э.М. Ергалиева, Е.А. Самсонык

### Органикалық қосылыстардың *Tetrahymena pyriformis* үшін улылығын QSAR әдісімен бағалау

Мақалада органикалық қосылыстардың *Tetrahymena pyriformis* инфузориясына қатысты  $-\lg(\text{IGC}_{50})$  улылық көрсеткішін бағалау үшін дескрипторлардың үлкен жиынтығын пайдаланып, QSAR/QSTR әдісін қолдану мүмкіндігі зерттелген. Барлығы 1093 қосылыстан тұратын заттар жиынтығы қарастырылған. Заттардың барлық жиынтығы үшін  $R = 0,968$ ,  $s = 0,26$  корреляция көрсеткіштері алынған. Заттардың жаттығулық іріктелген жиынтығы үшін  $R = 0,980$ ,  $s = 0,21$ , бақылау жиынтығы үшін  $R = 0,943$ ,  $s = 0,32$ .

V.V. Vazhev, M.A. Gubenko, N.V. Vazheva, E.M. Ergalieva, E.A. Samsonyuk

### Estimation of the toxicity of organic compounds to *Tetrahymena pyriformis* by QSAR method

The possibility of the use of QSAR / QSPR method is investigated for estimating the toxicity of organic compounds  $-\lg(\text{IGC}_{50})$  to *Tetrahymena pyriformis* using a large set of descriptors. The set of substances from 1093 compounds is investigated. For all set correlation indicators are received  $R = 0,968$ ,  $s = 0,26$ . For training set there are  $R = 0,980$ ,  $s = 0,21$ , for test —  $R = 0,943$ ,  $s = 0,32$ .

#### References

- 1 *The University of Tennessee: College of veterinary medicine*, <http://www.vet.utk.edu/research/chemical.shtml>.
- 2 Kumar M.S., Vibhanjali M., Tripathi P.N., Adil K.M. *Research Journal of Chemical Sciences*, 2014, 4, 2, p. 29–37.
- 3 Bouaoune A., Lourici L., Haddag H., Messadi D. *Journal of Environmental Science and Engineering*, 2012, A 1, p. 663–671.
- 4 Bellifa K., Mekelleche S.M. *Arabian Journal of Chemistry*, 2012, <http://dx.doi.org/10.1016/j.arabjc.2012.04.031>.
- 5 Jiang D.X., Li Y., Li J., Wang G.X. *Int. J. Environ. Res.*, 2011, 5, 4, p. 923–938.
- 6 Toropov A.A., Toropova A.P., Benfenati E., Manganaro A. *Mol. Divers*, 2010, 14, p. 821–827.
- 7 Vlaia V., Olariu T., Vlaia L., Butur M., Ciubotariu C., Medeleanu M., Ciubotariu D. *Farmacia*, 2009, 57, 4, p. 511–522.
- 8 Enoch S.J., Cronin M.T.D., Schultz T.W., Madden J.C. *Chemosphere*, 2008, 71, 7, p. 1225–1232.
- 9 Zhu H., Tropsha A., Fourches D., Varnek A., Papa E., Gramatica P., Öberg T., Dao P., Cherkasov A., Tetko I. *J. Chem. Inf. Model.*, 2008, 48, p. 766–784.
- 10 Vazhev V.V., Kuhareva A.Yu., Tsebruk A.V. *Bull. of Karaganda University. Chemistry Ser.*, 2010, 2(58), p. 5–10.