Биомедицинская химия, 2019 том 65, вып. 2, с. 123-132.

©Коллектив авторов

QSAR АНАЛИЗ ОСТРОЙ ТОКСИЧНОСТИ ОРГАНИЧЕСКИХ СОЕДИНЕНИЙ ПРИ ПЕРОРАЛЬНОМ ВВЕДЕНИИ МЫШАМ

О.В. Тиньков[1]*, В.Ю. Григорьев[2], П.Г. Полищук[3], А.В. Ярков[2], О.А. Раевский[2]

[1]Военный институт Министерства обороны,
3300, Молдова, Приднестровье, Тирасполь, Гоголя, 2"В"; *эл. почта: oleg.tinkov.chem@mail.ru;
[2]Институт физиологически активных веществ Российской академии наук,
142432, Московская область, Черноголовка, Северный пр., 1
[3]Институт молекулярной и трансляционной медицины, факультет медицины и стоматологии,
Университет Палацкого в г. Оломоуц, Чешская Республика, 77900, Оломоуц, Гневотинская, 1333/5

Зависимость между структурой органических соединений и острой токсичностью при пероральном введении мышам изучали с использованием 2D симплексного представления молекулярной структуры и метода “случайного леса” (Random Forest, RF). Были построены удовлетворительные (R^2_{test} = 0,61–0,62) модели количественной связи “структура-активность” (Quantitative Structure-Activity Relationship, QSAR). Проведена интерпретация полученных QSAR моделей. Рассчитаны вклады известных токсикофоров с установленными механизмами действия с целью оценить возможности их ранжирования относительно других структурных фрагментов. Для выявления новых потенциальных токсикофоров проанализировано влияние молекулярного окружения некоторых токсикофоров и вклады других, имеющих высокие ранги, фрагментов из списка общих функциональных групп и циклических систем. При проведении сравнительного QSAR анализа использованы онлайн версия экспертной системы “OCHEM” (https://ochem.eu) и алгоритм AMT (Arithmetic Mean Toxicity).

Ключевые слова: токсичность; QSAR; моделирование структура-свойство; симплексные дескрипторы; метод случайного леса; структурная интерпретация

DOI: 10.18097/PBMC20196502123

ВВЕДЕНИЕ

При разработке новых лекарственных средств необходима оценка различных видов токсичности, включая острую токсичность. При проведении доклинических исследований потенциальных лекарственных средств изучается острая токсичность при пероральном введении грызунам, в частности мышам. С целью повышения эффективности проводимых исследований широко применяют методы анализа QSAR (Quantitative Structure-Activity Relationship). Важнейшими преимуществами QSAR методов являются: а) выявление на ранних стадиях разработки потенциально опасных соединений; б) получение оценок механизмов действия физиологически активных соединений; в) осуществление целенаправленного молекулярного дизайна органических веществ с учётом влияния молекулярных фрагментов на изучаемое свойство.

В настоящее время разработано много математических моделей, описывающих количественно взаимосвязи структуры органических соединений с острой токсичностью. Детальный анализ существующих QSAR моделей для разных типов токсичности приведён в работе [1]. Обладающие достаточной предсказательной способностью классификационные QSAR модели острой токсичности (LD_{50}) для крыс с использованием представительных выборок соединений описаны в работе [2]. В работе [3] QSAR анализ острой токсичности осуществлён с использованием системы PASS, QNA дескрипторов и самосогласованной регрессии. В ходе анализа были построены QSAR модели острой токсичности (LD_{50}, крысы) для четырёх способов введения веществ (внутривенное, внутрибрюшинное, подкожное и пероральное). Выборка насчитывала 8972 соединения с изученной токсичностью при пероральном способе введения, 3549 соединений при внутривенном способе введения, 1314 соединений при внутрибрюшинном способе введения, а также 1084 соединения при подкожном способе введения. При этом, выборка из 7286 веществ (LD_{50}, ммоль/кг, крысы, перорально) была использована для сравнительного анализа, в ходе которого предлагаемый подход показал более высокие статистические характеристики. Полученные модели реализованы в виде экспертной системы GUSAR, размещённой в свободном доступе в сети Интернет [3]. Помимо этого, на основе системы PASS реализован свободнодоступный веб-сервис ROSC-Pred, позволяющий прогнозировать для изучаемых соединений проявление органоспецифической канцерогенности на грызунах [4]. Современный обзор Интернет-ресурсов, позволяющих вести поиск и разработку новых лекарственных препаратов, включая оценку различных видов токсичности, представлен в работе [5].

Существующие QSAR модели токсичности используют множество разных дескрипторов для описания молекулярной структуры и различные математические подходы для построения зависимостей. Во многих QSAR моделях отсутствует структурная интерпретация, что не позволяет проводить рациональный молекулярный дизайн

для получения соединений с низкой токсичностью. Детальный анализ методов интерпретации QSAR моделей представлен в публикации [6]. В обзоре [7] обсуждается использование QSAR подходов при изучении острой токсичности; при этом отмечается, что значительная часть существующих QSAR моделей острой токсичности по отношению к теплокровным животным имеет ограничения для широкого применения ввиду довольно скромных статистических характеристик этих моделей или незначительного числа соединений в рассмотренных выборках. В ходе QSAR моделирования острой токсичности органических соединений для млекопитающих достаточно часто используются представительные выборки, содержащие значения LD_{50} при пероральном введении крысам [2, 4, 8-10]. В большинстве известных экспертных систем, свободно доступных для пользования, например, "T.E.S.T.version 4.2" (https://www.epa.gov/chemical-research/users-guide-test-version-42-toxicity-estimation-software-tool-program-estimate), а также ряде коммерческих программных продуктов "TOPKAT" (http://accelrys.com), "ADMET Predictor" (www.simulationsplus.com) для прогнозирования острой токсичности соединений в отношении грызунов реализованы только QSAR модели, построенные на выборках, содержащих экспериментальные значения LD_{50}, при пероральном введении крысам. Необходимо отметить недавнюю публикацию [11], в которой описано построение QSAR моделей острой токсичности, в том числе LD_{50} при различных способах введения токсикантов крысам, мышам, кроликам, морским свинкам. В ходе проведённого исследования модели с более высокой прогнозирующей способностью были получены для кроликов и крыс.

В связи с этим, в задачи данного исследования входило следующее: 1) построение QSAR моделей с целью прогноза острой токсичности органических соединений для мышей при пероральном способе введения; 2) структурная интерпретация построенных QSAR моделей с целью выявления фрагментов, в наибольшей степени влияющих на острую токсичность.

МЕТОДИКА

Выборка по острой токсичности химических соединений при их пероральном введении мышам была отобрана из коммерческой базы данных SYMYX Toxicity Database (htttp://www.symyx.com/products/databases/index.jsp). Экспериментальные значения токсичности, приведённые в мг/кг живого веса, были пересчитаны в $\log(1/LD_{50})$ (ммоль/кг) и использованы в качестве количественной меры острой токсичности. В данную выборку было включено 6226 органических соединений. Выбранные данные содержали значения LD_{50} для мышей различного пола, возраста, условий содержания. Влияние этих факторов на токсичность может быть весьма значительно и давать различия в значениях $\log(1/LD_{50})$ по меньшей мере ±0,50 [12]. С целью оценки предсказательной способности QSAR моделей исходный набор (рис. 1)

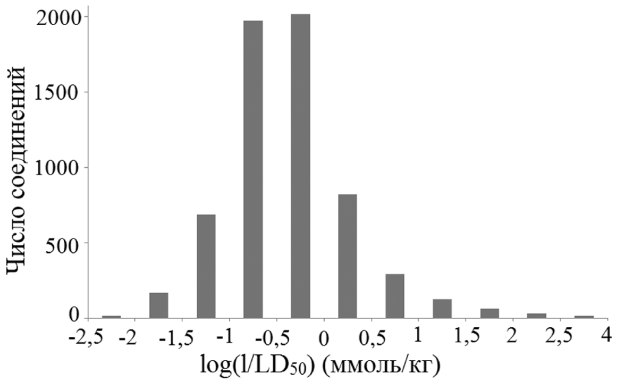


Рисунок 1. Распределение соединений по экспериментальной величине $\log(1/LD_{50})$.

разделяли на обучающую (4977 соединения) и тестовую (1249 соединения) выборки. Для этого с использованием компьютерной программы MolDivS [13] рассчитывали значение химического разнообразия (Diversity) каждого соединения выборки по отношению ко всем оставшимся соединениям на основе алгоритма [14] путём анализа парных величин разнообразия D(i,j). В качестве последних использовали величины, дополняющие оценки сходства (Similarity) S(i,j) этих соединений до единицы: D(i,j)=1-S(i,j). Расчет сходства проводили путём разбиения молекулы на фрагменты-сферы с центром в каждом атоме структуры и последующего определения числа совпадающих фрагментов. Затем проводили сортировку выборки по убыванию значения разнообразия. Далее, каждое пятое по порядку соединение помещали в тестовую выборку, остальные соединения относили к обучающей выборке.

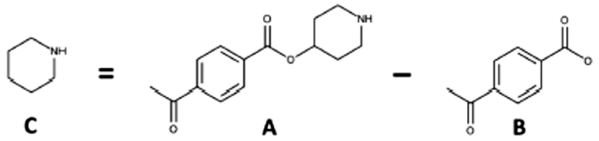
Структуры соединений и экспериментальные значения LD_{50} обучающей и тестовой выборок приведены, соответственно, в таблицах 1S и 2S дополнительных материалов.

Для стандартизации структур использовали "ChemAxon Standardizer" (https://chemaxon.com/products/chemical-structure-representation-toolkit).

Для описания молекулярной структуры были использованы 2D симплексные дескрипторы, рассчитываемые в рамках симплексного представления молекулярной структуры (Simplex Representation of Molecular Structure, SiRMS) (https://github.com/DrrDom/sirms).

Структурную интерпретацию при использовании симплексных дескрипторов (рис. 2) проводили в соответствии с подходом [15]; при этом вклад исследуемого фрагмента (**C**) в токсичность рассчитывали путём нахождения разницы между рассчитанными значениями активности/свойства для "родительской" структуры (**A**) и гипотетической структуры (**B**), полученной удалением исследуемого фрагмента (**C**) из "родительской" структуры (**A**).

При моделировании с использованием симплексных дескрипторов нами был применён метод "случайного леса" (RF, Random forest), реализованный в пакете "Scikit-learn" [16]. При построении моделей



W(**C**) = X(**A**) - X(**B**)

где, W(**C**) - вклад фрагмента (**C**);
X(**A**) - предсказанное значение активности / свойства для родительской структуры (**A**);
X(**B**) - предсказанное значение активности / свойства для гипотетической структуры (**B**).

Рисунок 2. Использованный принцип структурной интерпретации.

симплексными дескрипторами для внутренней валидации использовали процедуру пятикратной (5-fold) кросс-валидации. Область применимости (Applicability Domain, AD) определяли следующим образом: если значение хотя бы одного дескриптора выходило за пределы его минимального или максимального значения для обучающей выборки, то соединение тестовой выборки, содержащее данный дескриптор, не входило в область применимости (Bounding Box). Вышеописанный подход для QSAR моделирования с использованием симплексных дескрипторов реализован в виде программного обеспечения SPCI, которое находится в свободном доступе по адресу http://qsar4u.com/pages/sirms_qsar.php.

Для сравнительного анализа предсказательной способности QSAR моделей, построенных с помощью симплексных дескрипторов, при моделировании мы также использовали физико-химические дескрипторы, рассчитанные программой HYBOT, в основе которой лежит количественная оценка способности соединений к образованию водородной связи [17] и ряд дескрипторов, рассчитанных программой DRAGON Plus 5.5 (http://www.talete.mi.it/dragon.htm). При использовании данных дескрипторов для расчёта токсичности соединений использовали алгоритм AMT (Arithmetic Mean Toxicity, арифметическая средняя токсичность) [18].

Для сравнительного QSAR анализа нами была использована экспертная система OCHEM (https://ochem.eu). При этом для моделирования одновременно были выбраны дескрипторы ALogPS, StructuralAlerts, Fragmentor (Length 2-4), GSFrag (GSFrag GSFragL), ChemaxonScaffold (Scaffolds frameworks) и метод “ассоциативных нейронных сетей” (ASsociative Neural Networks, ASNN), подробно описанные в руководстве пользователя (http://docs.ochem.eu/display/MAN/OCHEM+Introduction).

При расчёте статистических критериев для тестовых выборок при моделировании вышеописанными методами учитывали попадание соединений в соответствующие области применимости.

Качество прогнозов на основе QSAR моделей оценивали по величинам *коэффициента детерминации (Q^2):*

$$Q^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - y_{mean})^2} \quad (1),$$

а также *среднеквадратичной ошибки (RMSE):*

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

где Q^2 – коэффициент детерминации, y_i – наблюдаемое значение активности для i-го соединения, $\hat{y}_i$ – предсказанное значение активности для i-го соединения, y_{mean} – среднее значение наблюдаемой активности, m – количество соединений в выборке.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Результаты QSAR моделирования приведены в таблице 1, из которой видно, что в ходе исследования при применении HYBOT, DRAGON дескрипторов, метода AMT, а также при использовании симплексных дескрипторов, и метода RF получены адекватные QSAR модели, обладающие сопоставимой друг с другом предсказательной способностью. Рассчитанные значения токсичности приведены в дополнительных материалах (таблица 1S – для обучающей выборки, таблица 2S – для тестовой выборки).

QSAR модель, построенная с использованием дескрипторов ALogPS, StructuralAlerts, Fragmentor (Length 2-4), GSFrag (GSFrag GSFragL), ChemaxonScaffold (Scaffolds Frameworks) и применением метода ассоциативных нейронных сетей, доступна для прогнозирования токсичности в рамках Интернет-ресурса OCHEM (https://ochem.eu/model/752).

Таблица 1. Статистические характеристики QSAR моделей.

Дескрипторы	Метод построения QSAR модели	Q^2	RMSE	Q^2_{test}	$RMSE_{test}$	Часть соединений, вошедших в AD, %
HYBOT, DRAGON	AMT	0,59	0,43	0,62	0,40	100
SiRMS	RF	0,57	0,45	0,61	0,40	86
«ALogPS», «StructuralAlerts», «Fragmentor (Length 2-4)», «GSFrag (GSFrag GSFragL)», «ChemaxonScaffold (scaffolds frameworks)» *	ASNN	0,48	0,48	0,55	0,44	100

Примечание. При построении данной модели интернет-ресурсом “OCHEM” не были рассчитаны дескрипторы для молекул, содержащих более 100 атомов, в результате чего 57 соединений обучающей выборки и 13 соединений тестовой выборки были проигнорированы при моделировании.

Одним из OECD (Organization for Economic Cooperation and Development) принципов проведения QSAR исследований является возможность интерпретации полученной модели. Однако, многие из общеизвестных дескрипторов, например, сгенерированных программой DRAGON, не позволяют осуществлять структурную интерпретацию построенных зависимостей. В связи с этим, использование симплексных дескрипторов, на наш взгляд, является определенным преимуществом. Необходимо отметить, что в рамках Интернет-ресурса OCHEM реализован метод молекулярных пар, который также позволяет проводить интерпретацию [19]. Однако, для интерпретации нами была выбрана модель, построенная с помощью симплексных дескрипторов, поскольку: (а) её статистические характеристики несколько лучше; (б) в ней отсутствуют ограничения количества атомов в молекуле при расчёте дескрипторов, имеющиеся в экспертной системе OCHEM. Кроме того, описанный выше метод интерпретации с использованием симплексных дескрипторов по сравнению с методом молекулярных пар больше подходит для анализа структурно разнообразных выборок. Анализ сравнительного использования двух вышеуказанных подходов при интерпретации QSAR моделей представлен в публикации [20].

Для QSAR модели, полученной с помощью симплексных дескрипторов и метода RF (табл. 1), была проведена структурная интерпретация. В ходе интерпретации для набора молекулярных фрагментов были определены вклады в изучаемую величину острой токсичности. Данный набор молекулярных фрагментов был сформирован из стандартных функциональных групп (51 фрагмент), набора известных токсикофоров (23 фрагмента, часть из которых приведена в таблице 2, столбец 1), приведённых в наших предыдущих работах [20], а также молекулярных фрагментов (228763 шт.), полученных в ходе автоматической фрагментации соединений обучающей выборки с помощью программы SPCI при использовании SMART-шаблона [#6+0;!$(*=,#[!#6])]!@!=!#[*], кодирующего разрываемые связи [20]. Структуры использованных фрагментов приведены в дополнительных материалах (табл. 3S).

Анализу при интерпретации подверглись только те молекулярные фрагменты, которые были обнаружены в 20-ти и более соединениях обучающей выборки, что, по нашему мнению, позволило сконцентрировать внимание на фрагментах, которые устойчиво влияют на изучаемый вид токсичности, и избежать, в некоторой степени, влияния случайных факторов, например, ошибок в экспериментальных данных или предсказанных значениях токсичности и вкладах фрагментов. Для автоматически сгенерированных фрагментов из-за их значительного количества использовали дополнительный фильтр, выделяющий только те молекулярные фрагменты, которые обладают относительно высоким средним рассчитанным вкладом в величину острой токсичности (превосходящим 1). Результаты интерпретации приведены на рисунке 3, а также в таблицах 2 и 3.

Проведённая интерпретация позволила количественно описать и ранжировать известные токсикофоры (табл. 2, столбец 1) по величине вклада в острую токсичность при пероральном введении токсикантов мышам, детализировать молекулярное окружение известных токсикофоров, выделив структурные производные данных фрагментов, которые вносят больший или меньший вклады в токсичность. Например, при детальном анализе вклада карбаматной группы выделены три фрагмента (2,2-диметил-2,3-дигидробензофуран-7-ил, N-(аминосульфанил)-N-метилкарбамат, метил(пропоксисульфанил)карбамат, фенилкарбамат), которые обладают значительно большим рассчитанным вкладом в величину острой токсичности.

Фрагменты, детализирующие молекулярное окружение известных токсикофоров (табл. 2, столбец 2), были определены в ходе автоматической генерации.

Полученные результаты позволили не только подтвердить высокий вклад в токсичность известных токсикофоров, которые описаны в литературе и рассмотрены в QSAR исследованиях других авторов [2, 25], но и предложить ряд новых потенциальных токсикофоров (табл. 3). В частности, при анализе было обращено внимание на то, что высоким расчётным значением вклада в токсичность обладает такая функциональная группа, как пиперазин, которая часто встречается в относительно нетоксичных лекарственных средствах. Анализ окружения данного фрагмента выявил такие фрагменты, как 1-карбоксимидоилпиперазин, фенилпиперазин, которые обладают более высоким вкладом в токсичность и, согласно литературным данным [26], влияют на важнейшие дофаминовые, серотониновые и адренорецепторы.

Выявленные в ходе исследования молекулярные фрагменты, обладающие высоким вкладом в острую токсичность (табл. 2 и 3), интегрированы в качестве “structural alerts” (структурных сигналов, токсикофоров) в экспертную систему OCHEM (Endpoint - LD50_mo_oral) и могут быть использованы при виртуальном скрининге потенциально опасных органических соединений.

В настоящее время структурные фрагменты, наличие которых указывает на потенциально высокую токсичность соединения, привлекают все большее внимание различных специалистов. Разработано достаточно много программных продуктов, позволяющих проводить виртуальный скрининг соединений для выявления токсикофоров, например, OECD QSAR Toolbox (http://www.qsartoolbox.org), “Toxtree” (http://toxtree.sourceforge.net/). Детальный анализ преимуществ и недостатков использования токсикофоров при рациональном дизайне и разработке физиологически активных веществ подробно проведён в работе [28].

Предложенные нами фрагменты не стоит рассматривать как “структурные фильтры” для виртуального скрининга. Очевидно, что токсичность соединения является интегральной характеристикой, существенным образом зависящей от окружения

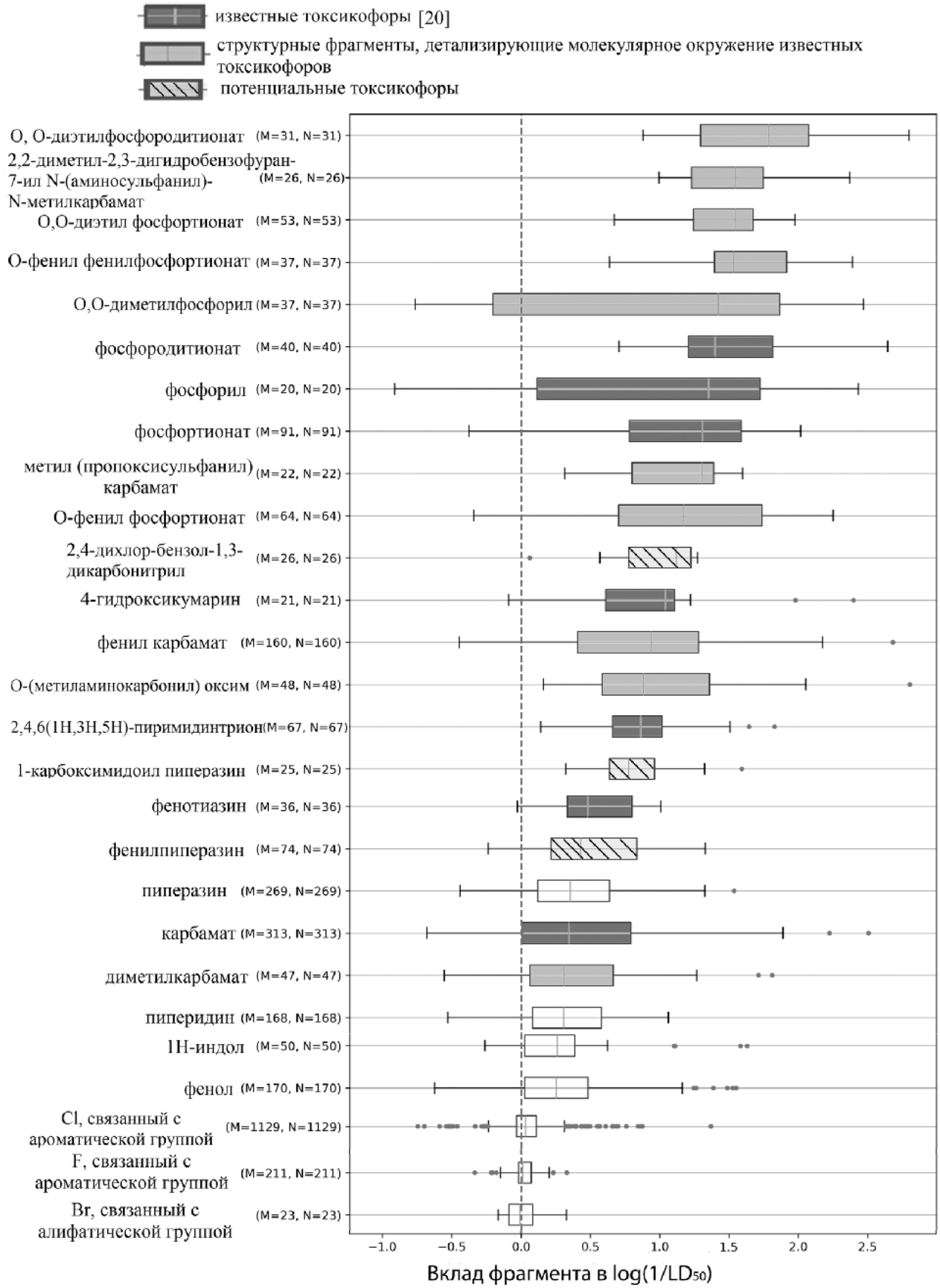
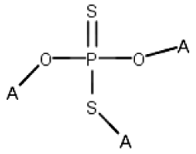
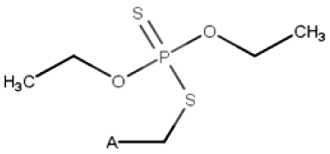
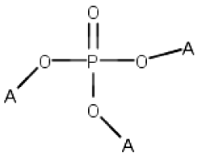
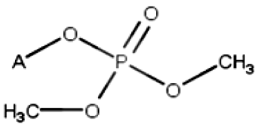
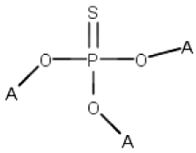
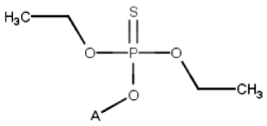
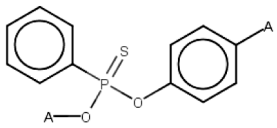
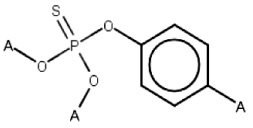
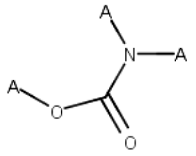
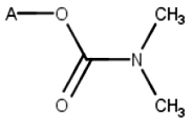
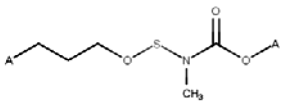
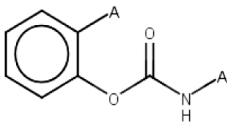
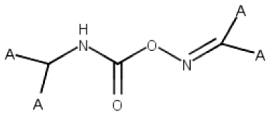
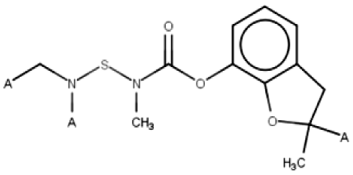
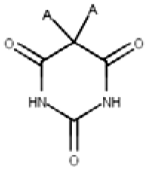
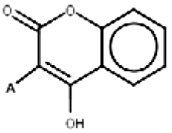
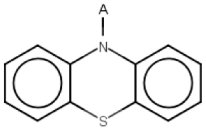


Рисунок 3. Вклады некоторых молекулярных фрагментов в острую токсичность по отношению к мышам при пероральном введении токсикантов. M – количество соединений, содержащих фрагмент. N – количество обнаружений соответствующего фрагмента в выборке. Структуры фрагментов, а также график с вкладами для всех молекулярных фрагментов приведены в дополнительных материалах (рис. 1S).

Таблица 2. Структуры известных токсикофоров, их молекулярное окружение

Известные токсикофоры [20]	Структурные фрагменты, детализирующие молекулярное окружение известных токсикофоров			Механизм токсического действия
 фосфородитионат	 O,O-диэтилфосфородитионат			Ингибирование холинэстеразы [21]
 фосфорил	 O,O-диметилфосфорил			
 фосфортионат	 O,O-диэтилфосфортионат	 O-фенил-фенилфосфортионат	 O-фенил-фосфортионат	
 карбамат	 диметилкарбамат	 метил(пропоксисульфанил) карбамат	 фенилкарбамат	
	 O-(метиламинокарбонил) оксим	 2,2-диметил-2,3-дигидро-1-бензофуран-7-ил N-(аминосульфанил)-N-метилкарбамат		
 2,4,6(1H,3H,5H)-пиримидинтрион				Взаимодействие с ГАМК-рецепторами [22]
 4-гидроксикумарин				Антикоагулирующие свойства (антагонисты витамина К) [23]
 фенотиазин				Ингибирование дофаминовых, серотониновых, гистаминовых рецепторов [24]

Примечание. A – место присоединения фрагмента к остальной части молекулы.

Таблица 3. Потенциальные токсикофоры

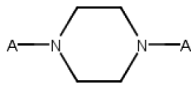
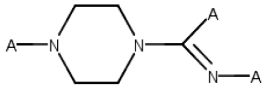
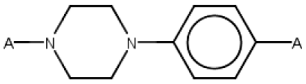
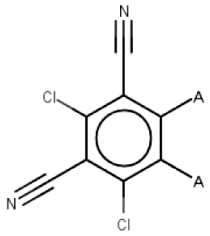
Функциональная группа, при детализации молекулярного окружения которой выявлены потенциальные токсикофоры	Потенциальные токсикофоры		Вероятный механизм токсического действия
 пиперазин	 1-карбоксимидоилпиперазин	 фенилпиперазин	Взаимодействие с дофаминовыми и серотониновыми, адренорецепторами [26]
Cl, связанный с ароматической группой	 2,4-дихлор-бензол-1,3-дикарбонитрил		Снижение в клетках уровня глутатиона, который участвует в поддержании окислительно-восстановительного потенциала, в процессах детоксикации ксенобиотиков [27]

Таблица 4. Классы опасности химических веществ

Показатель	Норма для класса опасности			
	1-го	2-го	3-го	4-го
острая токсичность, LD_{50}, мг/кг	≤ 50	> 50–500	> 500– 5000	> 5000

структурного фрагмента. Достаточно много примеров, когда при незначительном изменении структуры, например замены 1-2 атомов, биологическая активность молекулы кардинально меняется [29]. Основная цель выделения рассмотренных нами фрагментов – идентифицировать структурные участки для рационального дизайна, в частности, для модификации уже существующих высокотоксичных соединений с целью снижения их острой токсичности.

Так, например, из соединений тестовой выборки нами были отобраны вещества, имеющие в своей структуре O,O-диэтилфосфородитионат (молекулярный фрагмент с наибольшим рассчитанным вкладом в токсичность, см. рис. 3, табл. 2) и условно относящиеся по классификации Агентства по охране окружающей среды США (United States Environmental Protection Agency, U.S. EPA) к первому, наиболее опасному классу веществ. Один из показателей (острая токсичность), учитываемый при классификации веществ по категориям опасности, приведён в таблице 4, более детальная информация о классификации веществ, разработанная U.S. EPA, доступна по ссылке https://www.epa.gov/sites/production/files/2018-04/documents/lrm-complete-mar-2018.pdf.

Далее в отобранных соединениях молекулярный фрагмент O,O-диэтилфосфородитионат был заменен на фрагмент O-фенилфосфортионат, который обладает наименьшим вкладом в токсичность среди рассмотренных фосфорорганических фрагментов. Для полученных таким образом гипотетических соединений на следующем этапе был проведён прогноз острой токсичности с использованием моделей, разработанных с помощью экспертной системы OCHEM, а также симплексных дескрипторов и метода RF, реализованных в программе SPCI (табл. 5).

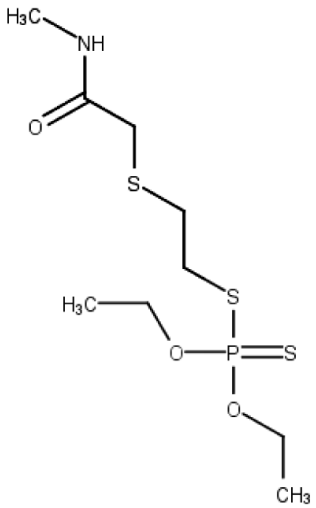
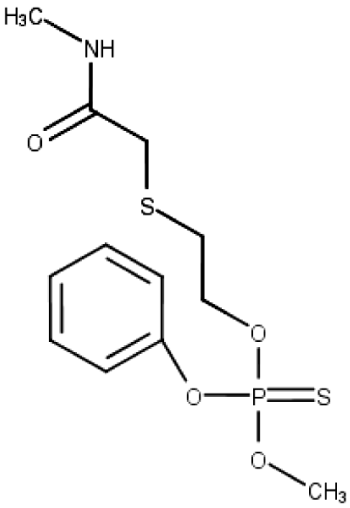
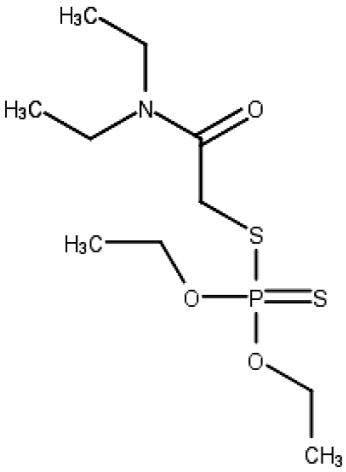
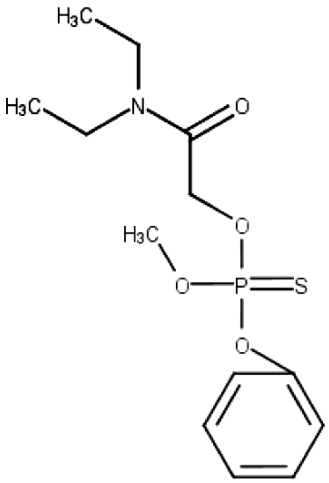
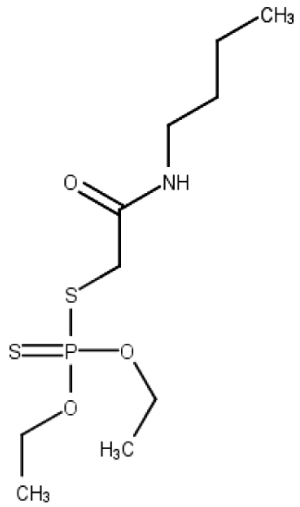
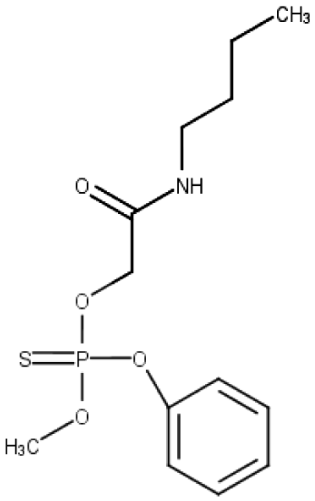
Для созданных гипотетических соединений с помощью программы Marvin Sketch v. 18.8.0 (https://chemaxon.com/products/marvin) была проведена проверка наличия структур в базе данных PubChem (https://pubchem.ncbi.nlm.nih.gov/), в ходе которой указанные вещества не были обнаружены.

Таким образом, мы продемонстрировали, что модификация структуры потенциально позволяет снизить токсичность опасных соединений, уменьшив класс опасности с первого на второй.

ЗАКЛЮЧЕНИЕ

Проведённые исследования позволили получить ряд адекватных регрессионных QSAR моделей, обладающих удовлетворительными точностью и предсказательной способностью. На основе полученных QSAR моделей определены структурные фрагменты, устойчиво повышающие острую токсичность при пероральном введении мышам, что может быть полезным при проведении молекулярного дизайна новых соединений с невысокой, приемлемой степенью токсичности.

Таблица 5. Результаты молекулярного дизайна

Соединения тестовой выборки	Эксп. знач. LD_{50}, мг/кг/день	Гипотетические соединения	Предсказ. знач. LD_{50}, мг/кг/день			
			OCHEM	AD	SPCI	AD
	14,0		76,8	-	64,14	+
	38,7		64,8	+	86,11	+
	51,0*		382,0	+	81,84	+

Примечание. Эксп. знач. – экспериментальные значения; Предсказ. знач. – предсказанные значения. * - Данное соединение нами условно было отнесено к первому классу опасности, поскольку оно незначительно превосходит порог указанного класса, и принятое упрощение позволило расширить набор исследуемых соединений. Знаки «+» или «-» соответственно означают вхождение соединения в область применимости модели (AD).

Дальнейшее развитие данных исследований планируется в направлении построения классификационно-регрессионных QSAR моделей органических соединений с целью выявления важнейших токсикофоров, а также исследования различных видов токсичности при других способах введения токсикантов (ингаляционно, внутривенно, накожно).

ДОПОЛНИТЕЛЬНЫЕ МАТЕРИАЛЫ

Дополнительные материалы доступны в электронной версии статьи на сайте журнала (http://pbmc.ibmc.msk.ru).

1. Соединения обучающей выборки, а также наблюдаемые и предсказанные значения токсичностей, полученные моделями AMT, SPCI, OCHEM.

2. Структуры соединений тестовой выборки в виде SMILE, а также наблюдаемые и предсказанные моделями AMT, SPCI, OCHEM значения токсичностей.

3. SMARTS шаблоны для анализа всех фрагментов.

4. Вклады всех молекулярных фрагментов в величину острой токсичности.

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

Настоящая статья не содержит каких-либо исследований с участием людей или с использованием животных в качестве объектов.

ЛИТЕРАТУРА

1. *Ekins S.* (2018) Computational Toxicology: Risk Assessment for Pharmaceutical and Environmental Chemicals, John Wiley & Sons, Hoboken, NJ.
2. *Xiao L., Lei C., Feixiong C., Zengrui W., Hanping B., Congying X., Weihua L., Guixia L., Xu S., Yun T.* (2014) J. Chem. Inf. Model., **54**, 1061-1069.
3. *Lagunin A., Zakharov A., Filimonov D., Poroikov V.* (2011) J. Molecular Informatics, **30**(2-3), 241-250.
4. *Lagunin A., Rudik A., Druzhilovsky D., Filimonov D., Poroikov V., Wren J.* (2018), Bioinformatics, **34**(4), 710-712.
5. *Беженцев В.М., Дружиловский Д.С., Иванов С.М., Филимонов Д.А., Sastry N.G., Поройков В.В.* (2017) Хим.-фарм. журнал, **51**(2), 3-11.
6. *Polishchuk P.* (2017) J. Chem. Inf. Model., **57**(11), 2618-2639.
7. *Devillers J., Devillers H.* (2009) SAR QSAR Environ. Res., **20**(5/6), 467-500.
8. *Hao Z., Martin T.M., Ye L., Sedykh A., Young D.M., Tropsha A.* (2009) Chem. Res. Toxicol., **22**, 1913-1921.
9. *Ballabio D., Grisoni F., Consonni V., Todeschini R.* (2018) Mol. Inf., **37**, 1800124. DOI: 10.1002/minf.201800124
10. *Gonella D.R., Manganelli S., Esposito A., Roncaglioni A., Manganaro A. Benfenati E.* (2015) SAR QSAR Environ. Res., **26**(1), 1-27.
11. *Sosnin S., Karlov D., Tetko I.V., Fedorov M.V.* (2018) J. Chem. Inf. Model. (Published: December 27, 2018 online). DOI: 10.1021/acs.jcim.8b00685
12. *Раевский О.А.* (2013) Свойства химических соединений и лекарств как функции их структуры. КДУ, Добросвет, М.
13. *Раевский О.А., Герасименко В.А., Трепалин С.В.* (1999) Свидетельство об официальной регистрации программы для ЭВМ MOLDIVS (Molecular Diversity & Similarity) № 990093 от 26 февраля 1999 г., Москва, Федеральная служба по интеллектуальной собственности, патентами товарным знакам.
14. *Trepalin S.V., Gerasimenko V.A., Kozyukov A.V., Savchuk N.Ph., Ivaschenko A.A.* (2002) J. Chem. Inf. Comput. Sci., **42**, 249-258.
15. *Polishchuk P.G., Kuz'min V.E., Artemenko A.G. Muratov E.N.* (2013) Molecular Informatics, **32**, 843-853.
16. *Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V. et al.* (2011) J. Mach. Learn. Res., **12**, 2825-2830.
17. *Raevsky O., Grigor'ev V., Kireev D., Zefirov N.S.* (1992) Quant. Struct.-Act. Relat., **11**, 49-63.
18. *Raevsky O., Grigor'ev V., Modina E., Worth A.P.* (2010) SAR QSAR Environ. Res., **21**, 265-275.
19. *Sushko Y., Novotarskyi S., Körner R., Vogt J., Abdelaziz A., Tetko I.V.* (2014) J. Cheminform., **6**(1), 48.
20. *Polishchuk P., Tinkov O., Khristova T., Ognichenko L., Kosinskaya A., Varnek A., Kuz'min V.*. (2016), J. Chem. Inf. Model., **56**(8), 1455-1469.
21. *Colovic M., Krstic D., Lazarevic-Pasti T., Bondžić A.M., Vasić V.M.* (2013) Curr. Neuropharmacol., **11**(3), 315-335.
22. *Löscher W., Rogawski M.*A. (2012) Epilepsia, **53**, 12-25.
23. *Littin K., O'Connor C., Eason C.* (2000) New Zealand Plant Protection, **53**, 310-315.
24. *Lester P., Myers D.* (2012) in: The Laboratory Rabbit, Guinea Pig, Hamster, and Other Rodents (Suckow M., Stevens K., Wilson R., eds.), Academic Press, pp. 33-56.
25. *Sazonovas A., Japertas P., Didziapetris R.* (2010) SAR QSAR Environ. Res., **21**, 127-148.
26. *Bürki H.R., Ruch W., Asper H.* (1975) Psychopharmacologia, **41**(1), 27-33.
27. *Gallagher E., Canada A., Giulio R.* (1992) Aquatic Toxicol., **23**, 155-168.
28. *Alves V., Muratov E., Capuzzi S., Politi R., Low Y., Braga R., Zakharov A.V., Sedykh A., Mokshyna E., Farag S. et al.* (2016) Green Chem., **18**, 4348-4360.
29. *Politi R., Rusyn I., Tropsha A.* (2014) Toxicol. Appl. Pharmacol., **280**(1), 177-189.

Поступила в редакцию: 30. 12. 2018.
После доработки: 30. 12. 2018.
Принята к печати: 04. 03. 2019.

QSAR INVESTIGATION OF ACUTE TOXICITY OF ORGANIC COMPOUNDS DURING ORAL ADMINISTRATION TO MICE

O.V. Tinkov[1]*, V.Yu. Grigorev[2], P.G. Polishchuk[3], A.V. Yarkov[2], O.A. Raevsky[2]

[1]Military Institute of the Ministry of Defense,
2"B" Gogol str., Transdniestria, Tiraspol, 3300 Moldova, ; *e-mail: oleg.tinkov.chem@mail.ru;
[2]Institute of Physiologically Active Compounds, Russian Academy of Sciences,
1 Severniy proezd, Chernogolovka, Moscow region, 142432 Russia,
[3]Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry,
Palacký University, 1333/5 Hněvotínská str., Olomouc, 77900 Czech Republic

The effect of the structure of organic compounds on the acute toxicity upon oral injection in mice was studied using 2D simplex representation of the molecular structure and Random forest (RF) methods. Satisfactory quantitative structure-activity relationship (QSAR) models were constructed (R^2_{test} = 0,61–0,62). The interpretation of the obtained QSAR models was carried out. The contributions of known toxicophores with established mechanisms of action were calculated in order to confirm the ability of the interpretation approach to correctly rank them relative to other structural fragments. The influence of the molecular surroundings of some toxicophores was analyzed. We analyzed the contributions of other highly ranked fragments from the list of common functional groups and ring systems in order to find new potential toxicophores. The on-line version of the expert system “OCHEM” (https://ochem.eu) and Arithmetic Mean Toxicity (AMT) approach were used for a comparative QSAR study.

Key words: toxicity; QSAR; structure-property modeling; simplex descriptors; random forest; structural interpretation