ПРОТЕОМИКА, БИОИНФОРМАТИКА

УДК 541.69+519.25+518.5

РАСЧЕТ ОСТРОЙ ТОКСИЧНОСТИ ХИМИЧЕСКИХ СОЕДИНЕНИЙ ПРИ ИХ ВНУТРИВЕННОМ ВВЕДЕНИИ МЫШАМ НА ОСНОВЕ ЛОКАЛЬНЫХ РЕГРЕССИОННЫХ МОДЕЛЕЙ В ПЕРЕКРЫВАЮЩИХСЯ КЛАСТЕРАХ (ЛРМПК)

О.А. Раевский[1]*, В.Ю. Григорьев[1], Е.А. Липлавская[1], А.П. Вортс[2]

[1]Институт физиологически активных веществ РАН, Северный проезд, 1, 142432 Черноголовка, Московская область; факс: 8(49652)49-508; эл. почта: raevsky@ipac.ac.ru
[2]Институт здоровья и защиты потребителя, Европейская комиссия – Объединённый исследовательский центр, ул. Энрико Ферми 2749, 21027 Испра, Италия

Приведено компьютерное моделирование взаимосвязи физико-химических дескрипторов органических соединений и их острой токсичности при внутривенном введении мышам. Данный подход включает три стадии: отбор структурно-родственных соединений для каждого из рассматриваемых соединений указанной выборки (кластеризация), построение количественных соотношений структура-токсичность для каждого кластера (без включения рассматриваемого соединения), применение построенных КССА уравнений для оценки токсичности рассматриваемых соединений. Этот подход был использован для расчёта токсичности 10241 соединений при их внутривенном введении. Для 7759 соединений из указанного общего числа, имеющих структурных соседей с индексом сходства (индекс Танимото) на уровне 0,30 и выше, стандартное отклонение рассчитанных значений от экспериментальных составило 0,51 при ошибке экспериментального определения ±0,50 в величине $\log(1/LD_{50})$. Для оставшихся соединений (~24%) результаты расчетов не столь совершенны, что связано с отсутствием для них достаточного числа структурно-родственных аналогов. Предполагается, что описанная в данной статье КССА модель может быть полезной для предсказания биологической активности и токсичности больших массивов соединений.

Ключевые слова: КССА, токсичность, структурное сходство, HYBOT, DRAGON, кластеризация, регрессионные модели.

ВВЕДЕНИЕ. Традиционно для моделирования количественной взаимосвязи структура - активность (КССА) используются линейные регрессионные уравнения между различного рода дескрипторами и свойством (активность, токсичность). Такой подход основывается на предположении гладкого профиля свойства/активности. Однако в последнее время появились свидетельства, что указанный традиционный подход КССА моделирования неадекватно описывает свойство/активность разнообразных по структуре соединений [1].

* - адресат для переписки

Другими словами, традиционное регрессионное уравнение взаимосвязи свойства/активности с дескрипторами:

$$\log activity\ (property)_i = a_0 + a_1x_1 + a_2x_2 + \ldots\ldots a_nx_n \qquad (1),$$

не в состоянии описать сложный профиль свойства/активности для разнообразных по структуре химических соединений.

В 2001 году Раевский предложил регрессионную модель с учётом свойств ближайшего соседа/соседей [2]:

$$\log activity\ (property)_i = \log activity\ (property)_{nn} +$$
$$a_1(x_{1i}-x_{1nn}) + a_2(x_{2i}-x_{2nn})\ldots+ a_n(x_{ni}-x_{nnn}) \qquad (2),$$

где activity $(property)_i$ – активность/свойства рассматриваемого соединения, activity $(property)_{nn}$ – экспериментальное значение/свойство ближайшего структурного соседа.

Со статистической точки зрения, в таком подходе желательно включать в расчёт несколько структурных соседей и использовать в качестве рассчитанного значения среднее значение свойства/активности оцененное по уравнению (2) для каждого из этих соседей.

Уравнение (2) имеет важные преимущества по сравнению с уравнением (1):

1. Применение экспериментальных значений ближайших структурных соседей даёт возможность учесть особенности структуры рассматриваемого соединения в неявном виде.

2. Увеличивается вероятность обнаружения и расчета свойства/активности так называемых “утёсов” или “провалов” [1] в сложном ландшафте свойства/активности.

3. За счёт использования данных о ближайшем структурном соседе существенно уменьшается интервал экстраполяции данных.

Указанный подход был успешно применен для расчета липофильности [2], растворимости в воде [3, 4], адсорбции химических соединений в организме человека [5, 6], и токсичности по отношению к *Guppy* [7]. Опыт использования уравнения (2) показывает, что результаты расчета становятся лучше в случае, когда структурный сосед имеет близкие физико-химические свойства по отношению к рассматриваемому соединению. В то же время стало очевидным, что применение коэффициентов значений из уравнения (1) в уравнении (2) является удовлетворительным, но не идеальным приближением. Представляется, что лучший путь это развитие специфических локальных КССА моделей для кластеров структурно-родственных соединений. В литературе имеется много подобных подходов к использованию методологии дискретно-непрерывных моделей (например, SIMCA/PLS [8], кластерно-регрессионные модели [9, 10], DIREM [11]), в которых классификационно-регрессионные КССА успешно применены в расчете различных свойств/активности различных соединений. Однако во всех этих подходах изученные выборки разделялись на ограниченное число кластеров. И в случае обширных выборок такие кластеры содержат большое число разнообразных химических соединений. Очевидно, в таких случаях необходима разработка иных КССА моделей.

В качестве одного из решений этой проблемы можно рассматривать локальную неактивную регрессию (local lazy regression). В этом подходе, описанном в [12, 13], на основе метода *k*-ближайших соседей *(k*NN) производится кластеризация, и в каждом кластере строится собственная регрессионная модель.

В настоящей работе в качестве дальнейшего развития концепции дискретно-непрерывных КССА моделей предлагается подход на основе регрессионных моделей в суперперекрывающихся кластерах. Указанный подход применен к расчёту острой токсичности при внутривенном введении 10241 органических соединений мышам. Следует отметить, что острая токсичность химических соединений по отношению к теплокровным является исключительно сложным явлением, связанным с взаимодействием соединений с большим числом биологических мишеней в организме. И как следствие указанной сложности явления и недостаточного понимания механизмов действия число публикаций, посвященных КССА острой токсичности по отношению к теплокровным невелико. Такие модели обсуждаются в обзорах [см. обзоры 14, 15]. При этом было выявлено, что подавляющее большинство из 150 опубликованных КССА моделей острой токсичности по отношению к теплокровным имеет весьма ограниченную пользу, что связано с довольно скромным статистическим качеством этих моделей и ограниченным числом соединений в рассмотренных выборках.

МЕТОДИКА. Выборка по острой токсичности химических соединений при их внутривенном введении мышам была отобрана из коммерческой базы данных SYMYX Toxicity Database [16]. Экспериментальные значения токсичности, приведенные 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 10241 нейтральное органическое соединение. Выбранные данные содержали значения LD_{50} для мышей различного пола, возраста, условий содержания. Влияние этих факторов на токсичность может быть весьма чувствительно и давать различия в значениях $\log(1/LD_{50})$ (ммоль/кг) по меньшей мере, ±0,50. Таким образом, удовлетворительная КССА модель должна иметь стандартное отклонение рассчитанных значений токсичности от экспериментальных на этом же уровне. Полный интервал $\log(1/LD_{50})$ для всех 10241 соединений составил 7,49 логарифмических единиц [от (-3,14) до 4,35]. Однако 95% всех соединений имеют значения $\log(1/LD_{50})$ в интервале (-1,50)–(2,50). Величина этого интервала и наблюдаемое нормальное распределение соединений внутри него (см. рис. 1) являются достаточными для построения КССА моделей.

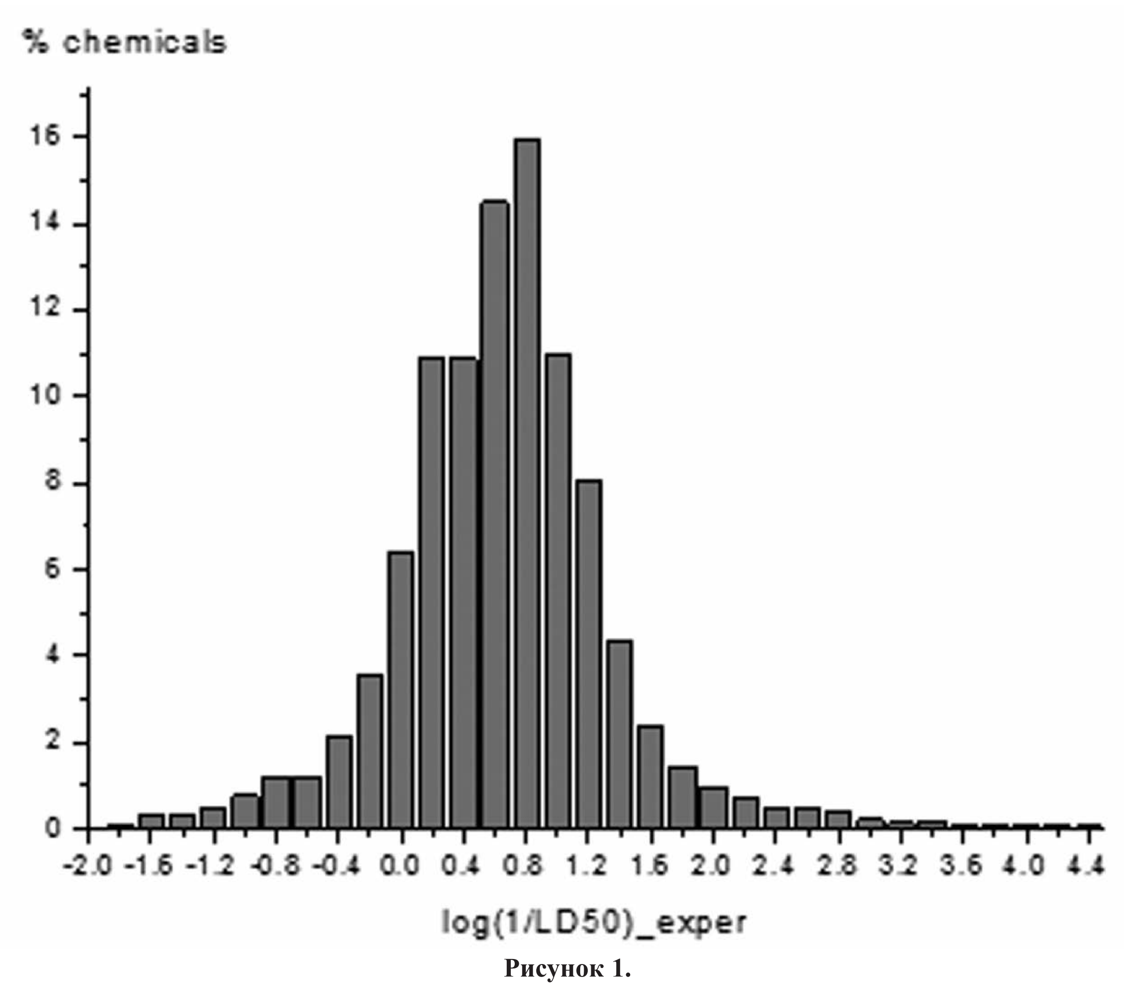


Рисунок 1.
Распределение соединений по величине $\log(1/LD_{50})_{\_exper}$.

В выборке представлены практически все функциональные химические группировки. С этой точки зрения возможные КССА модели для этого массива могут быть, в принципе, применены к самым разнообразным органическим производным. Тем не менее, мы установили, что 13% всей выборки не имеют структурных соседей с индексом Танимото (Tc) на уровне 0,50 и выше, около 10% имеют одного такого соседа, 8% имеют только двух структурных соседей, 6% имеют трёх соседей, и около 5% имеют четыре соседа с Tc≥0,50 (рис. 2). Это значит, что концепция структурного сходства может быть применена для оценки токсичности этих 42% соединений с большой осторожностью. И ошибка в расчётах $\log(1/LD_{50})$ (ммоль/кг) для них может быть больше ±0,50.

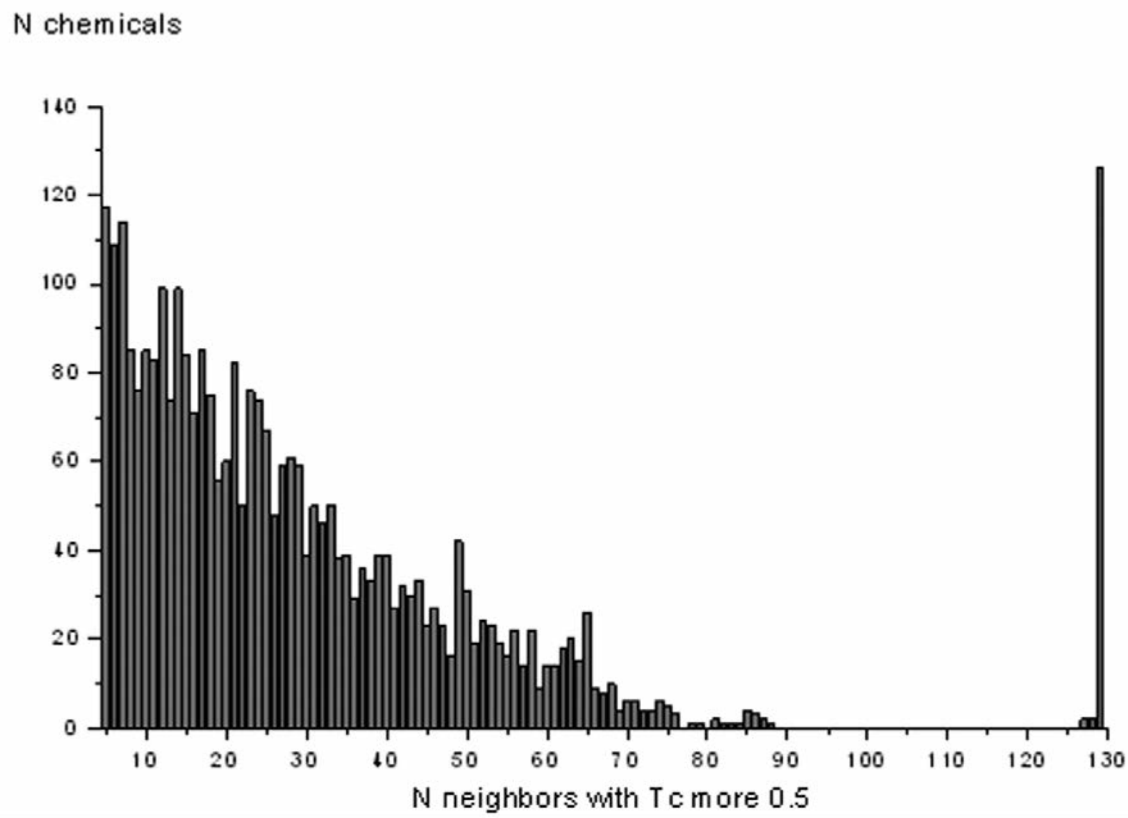


Рисунок 2.
Распределение соединений по размеру кластера (количеству структурных соседей с Tc≥0,5).

Для расчёта дескрипторов использовались программы HYBOT [17] и DRAGON [18].

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ.

Алгоритм построения локальных регрессионных моделей в перекрывающихся кластерах (ЛРМСК) может быть представлен следующим образом:

Шаг 1. Для каждого рассматриваемого (i) соединения из выборки, содержащей n соединений, все остальные n-1 соединений ранжируются в соответствии с их структурным сходством по отношению к рассматриваемому. Данная процедура выполняется программой MOLDIVS [19] на основе значений индекса Танимото. В результате получается n отдельных рядов соединений.

Шаг 2. Фиксируется минимальный порог структурного сходства с помощью индекса Танимото. И в каждом из n рядов отбираются структурно-родственные соединения, имеющие индекс структурного сходства, выраженный индексом Танимото, выше фиксированного уровня. Таким образом, формируются n суперперекрывающихся кластеров (рис. 3).

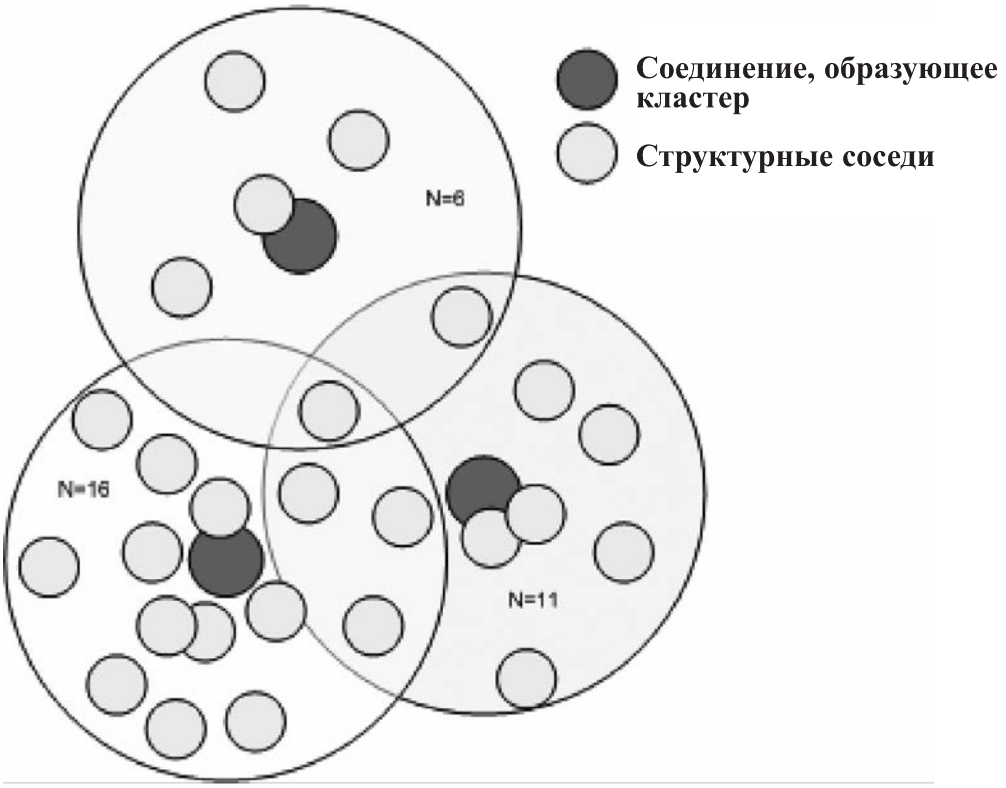


Рисунок 3.
Графическое представление метода ЛРМСК.

Шаг 3. Строятся отдельные локальные регрессионные КССА модели для каждого кластера на основе какого-либо дескриптора(ов). В настоящей работе были использованы только одно-, двух- или трёхпараметровые линейные регрессионные модели при наличии в кластере соответственно 6 и более, 11 и более, 16 и более структурных соседей. Это обеспечивает наличие минимум 5 соединений на один дескриптор в процессе перекрестного контроля с исключением по одному. Подобные нелинейные модели будут обсуждаться в отдельной публикации.

Очень важно отметить, что рассматриваемое соединение никогда не включается в построение этих регрессионных уравнений.

Шаг 4. Полученные в шаге 3 КССА уравнения используются для расчёта активности рассматриваемых соединений с применением уравнения (1).

Таким образом, в рамках данного подхода каждое рассматриваемое соединение является тестовым соединением. А качество финальных результатов использования описанных моделей определяется размером и качеством выборки, наличием достаточного количества близких структурных соседей в каждом кластере, стабильностью построенных локальных КССА моделей.

Фактически представленный выше алгоритм модели ЛРМСК является логическим развитием регрессионной модели с учетом свойств ближайшего соседа/соседей, описанной в публикации [2]. Различие этих двух моделей связано только со способом использования информации о ближайших структурных соседях. Если в предыдущей модели свойства ближайших соседей использовались по отдельности для расчёта свойства рассматриваемого соединения на основе общих коэффициентов регрессионного уравнения, то в настоящем подходе ближайшие структурные соседи используются для построения собственной локальной регрессионной модели кластера.

КССА для острой внутривенной токсичности на основе ЛРМСК подхода.

Таблица 1 содержит результаты расчётов острой внутривенной токсичности по отношению к мышам с помощью ЛРМСК подхода в случае использования дескрипторов программы HYBOT.

Таблица 1. Результаты расчётов острой токсичности при внутривенном введении по отношению к мышам с помощью ЛРМСК подхода (HYBOT дескрипторы, между дескрипторами R≤0,4).

Tc	a_0	er_a_0	a_1	er_a_1	R^2	SD	Q^2	Sd_{cv}	F	N	1-парам. ур.	2-парам. ур.	3-парам. ур.
0,3	0,01	0,01	0,99	0,01	0,434	0,51	0,434	0,51	5957,8	7759	6425	1176	158
0,4	0,06	0,01	0,91	0,01	0,53	0,46	0,529	0,46	5646,3	5011	4074	810	127
0,5	0,06	0,01	0,9	0,01	0,621	0,40	0,62	0,40	5169,5	3157	2514	540	103
0,6	0,09	0,01	0,87	0,02	0,621	0,40	0,618	0,40	2861,8	1748	1457	260	31
0,7	0,15	0,02	0,76	0,02	0,618	0,38	0,613	0,38	1489,6	924	814	99	11
0,8	0,10	0,03	0,88	0,03	0,659	0,37	0,643	0,38	665,7	346	272	74	0

В начале работы был выделен порог структурного сходства на уровне Tc≥0,30, и отобраны соединения, имеющие выше этого порога не менее шести соседей. Последнее условие обеспечивает минимальное число точек для построения однопараметрового уравнения с возможностью использования процедуры кросс-валидации. Это уменьшило общее число рассматриваемых соединений до 7759. Что касается двух- и трёхпараметровых уравнений отбирались только те дескрипторы, которые имели коэффициент взаимной корреляции на уровне r≤0,4. Распределение соединений в полученных кластерах представлено на рисунке 4.

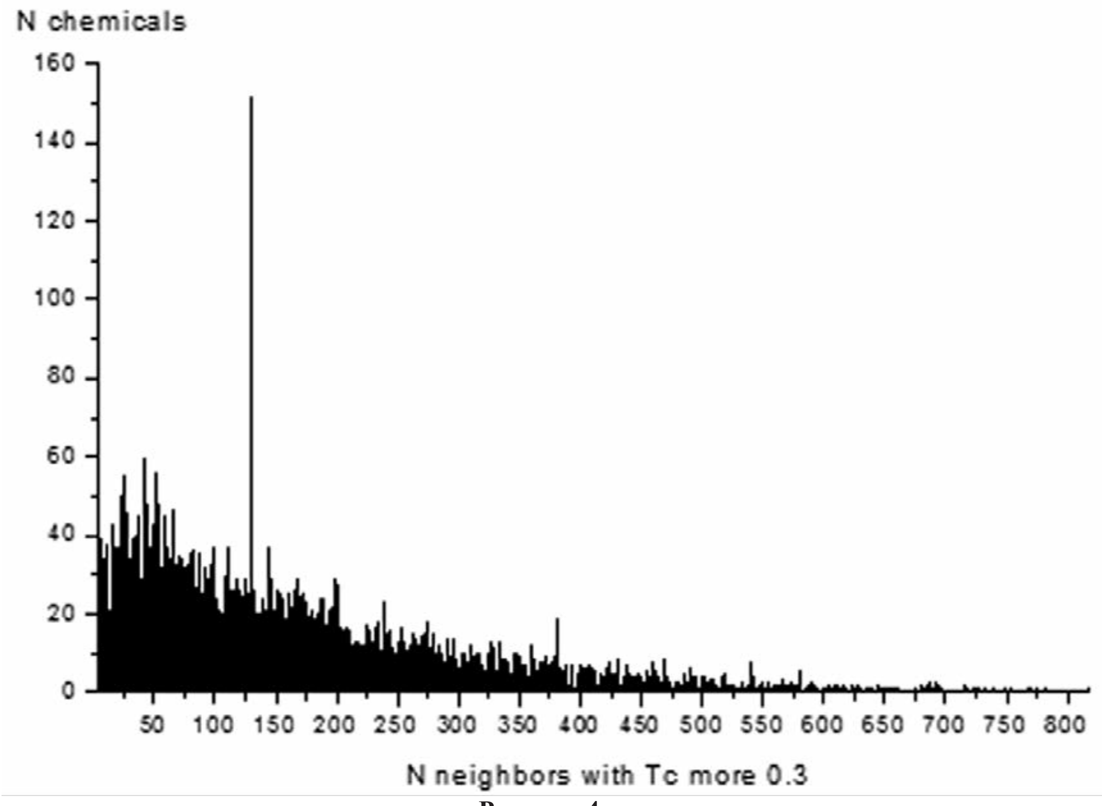


Рисунок 4.
Распределение соединений по размеру кластера (количеству структурных соседей с Tc≥0,3).

Как видно из этого рисунка 153 соединения имеют по 129 структурных соседей. Эти кластеры содержат производные рифомицина. 60 соединений имеют кластеры с 50 структурными соседями. А одно соединение имеет даже 775 соседей на уровне Tc≥0,30. Все другие соединения имеют меньшее число структурно-родственных соседей в кластерах, но в достаточном их количестве для построения одно-, двух- или трёхпараметровых регрессионных моделей.

В результате было получено 7759 кластеров, которые позволили по имеющимся в каждом из них структурным соседям построить 7759 соответствующих регрессионных уравнений (6425 - однопараметровых, 1176 – двухпараметровых и 158 – трехпараметровых) для расчёта токсичности соответствующих 7759 соединений. Для оставшихся 2482 соединений не было обнаружено достаточного числа структурных соседей, необходимого для оценки токсичности с помощью описанного метода. Можно полагать, что последующее расширение базы данных по острой внутривенной токсичности по отношению к мышам в дальнейшем может привести к расчёту токсичности и этих соединений при условии появления среди новых данных соответствующих структурно-родственных соединений.

Три дескриптора HYBOT (композитный дескриптор, характеризующий свободноэнергетический протонодонорный фактор на единицу молекулярной поляризуемости (ΣCd/α); композитный дескриптор, характеризующий энтальпийный протонодонорный фактор на единицу молекулярной поляризуемости ((ΣEd/α) и молекулярная поляризуемость (α)) оказались наиболее используемыми в однопараметровых уравнениях (20%, 18% и 9% всех однопараметровых уравнений соответственно (рис. 5)).

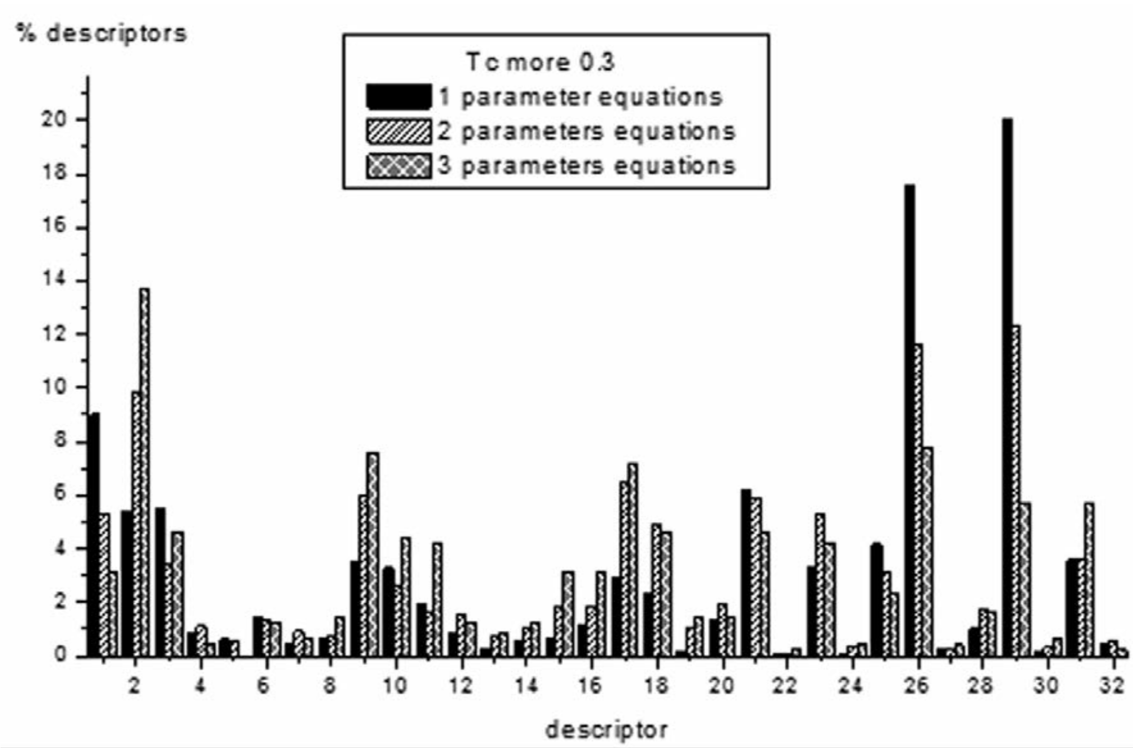


Рисунок 5.

Распределение дескрипторов программы HYBOT в одно-, двух- и трёхпараметровых уравнениях при Tc≥0,3 и взаимной корреляции дескрипторов на уровне r≤0,4

(1 - α, 2 - $maxQ^+$, 3 - $maxQ^-$, 4 - ΣQ^+, 5 - ΣQ^-, 6 - Σ|Q|, 7 - $\Sigma Q^+/\alpha$, 8 - $\Sigma Q^-/\alpha$, 9 - maxEa, 10 - maxCa, 11 - maxCa(o), 12 - maxEd, 13 - maxCd, 14 - maxEa* maxEd, 15 - maxCa* maxCd, 16 - maxCa(o)* maxCd(o), 17 - ΣEa, 18 - ΣEd, 19 - ΣEad, 20 - ΣCa, 21 - ΣCd, 22 - ΣCad, 23 - ΣCa(o), 24 - ΣCad(o), 25 - ΣEa/α, 26 - ΣEd/α, 27 - ΣEad/α, 28 - ΣCa/α, 29 - ΣCd/α, 30 - ΣCad/α, 31 - ΣCa(o)/α, 32 - ΣCad(o)/α).

Парная комбинация данных дескрипторов и свободноэнергетических или энтальпийных протонодонорных факторов водородной связи (ΣCa, ΣEa) довольно распространена в двухпараметровых уравнениях. Необходимо заметить, что эти дескрипторы HYBOT напрямую связаны с транспортными свойствами (растворимость в воде, липофильность и т.д.) химических соединений и лекарств. Таким образом, можно предположить, что транспортные свойства химических соединений и лекарств существенно влияют на их острую внутривенную токсичность.

Статистические параметры корреляции между экспериментальными и рассчитанными на основе соответствующих локальных регрессионных уравнений значениями токсичности для 7759 соединений приведено в таблице 1 (строка 1). В указанном окончательном уравнении свободный член и наклон оказались идеальными (0,01±0,01 и 0,99±0,01 соответственно), а стандартное отклонение в процедуре перекрестного контроля на уровне ошибки экспериментального измерения этого свойства для мышей разного возраста, пола, условий их содержания и условий измерений (±0,50).

Использование порога структурного сходства на уровне Tc≥0,40 и взаимной корреляции дескрипторов на уровне r≤0,4 существенно уменьшило размер выборки (с 7759 до 5011 соединений) и количество структурных соседей в кластере. Параметры и статистические критерии корреляции между экспериментальными и рассчитанными значениями $\log(1/LD_{50})$ стали несколько лучше (табл. 1, строка 2). Эта тенденция была обнаружена и в процессе дальнейшего увеличения порога структурного сходства при взаимной корреляции между дескрипторами r≤0,4: при Tc≥0,50 – 3157 уравнений; при Tc≥0,60 – 1748 уравнений; при Tc≥0,70 – 924 уравнений; при Tc≥0,80 – 346 уравнений.

Такие же расчеты были проведены и с дескрипторами программы DRAGON. Зависимости значений коэффициента корреляции и стандартного отклонения от величины порога структурного сходства в уравнениях сравнения экспериментальных и рассчитанных значений $\log(1/LD_{50})$ представлены на рисунке 6.

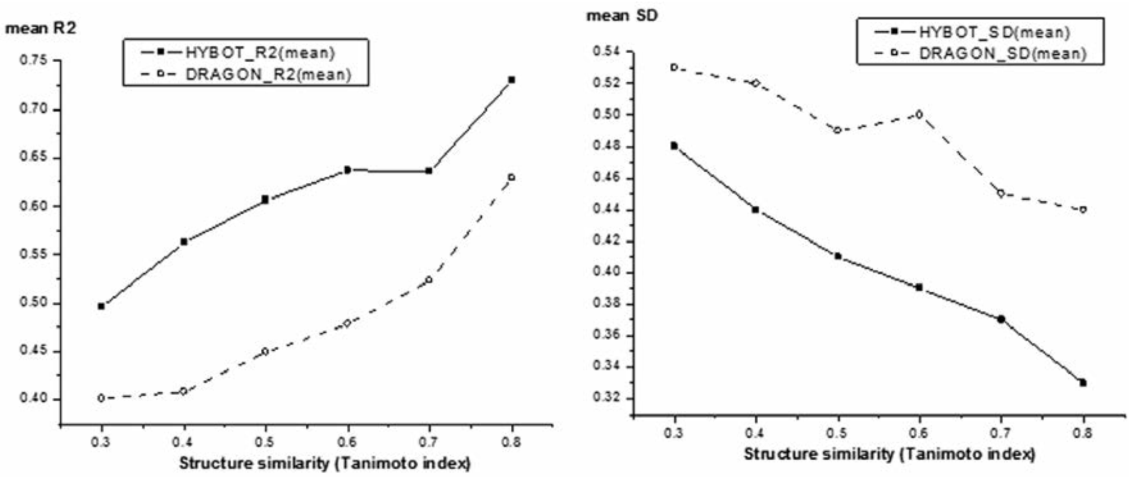
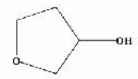
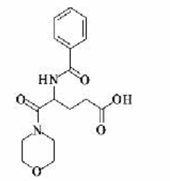
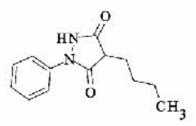
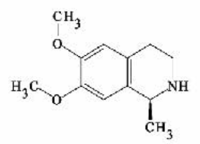
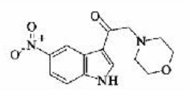
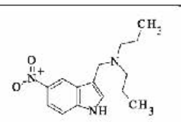
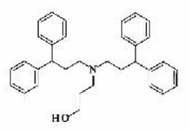
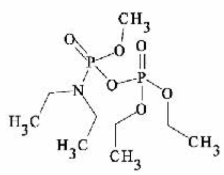
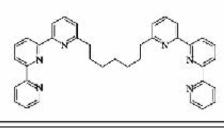


Рисунок 6.
Зависимости значений коэффициента корреляции и стандартного отклонения от величины порога структурного сходства в уравнениях сравнения экспериментальных и рассчитанных значений log1/LD50 (HYBOT и DRAGON дескрипторы).

Очевидно, что порог структурного сходства имеет существенное влияние на точность расчётов. Статистические критерии таких уравнений становятся лучше при большем значении данного порога. Но одновременно количество получаемых уравнений, а, следовательно, и число соединений с рассчитанным значением токсичности уменьшается. Это явно демонстрирует эффективность использования концепции структурного сходства для оценки свойств/активности соединений. Рисунок 6 также демонстрирует, что лучшие результаты расчётов получаются при использовании дескрипторов программы HYBOT.

Эффективность описанного в данной работе подхода ЛРМСК была протестирована сравнением результатов расчета этим методом и традиционным методом с использованием 6 ближайших структурных соседей. Результаты этого сопоставления представлены в таблице 2.

Таблица 2. Результаты расчёта токсичности ряда химических соединений методом ЛРМСК и 6 ближайшими структурными соседями.

№	CAS номер	Структура	$\log(1/LD_{50})$ 6NN	$\log(1/LD_{50})$ ЛРМСК	$\log(1/LD_{50})$ exper
1	453-20-3		-1.12	-1.51	-1.64
2	6460-75-9		-0.66	-1.26	-1.10
3	2210-63-1		0.01	-0.27	-0.41
4	493-48-1		0.61	0.33	0.09
5	101832-08-0		1.17	0.40	0.51
6	506-12-7		0.11	0.88	0.88
7	3414-66-2		2.13	1.30	1.14
8	52017-07-9		0.39	1.20	1.19
9	625-13-8		1.85	2.63	2.58
10	57154-77-5		1.24	3.07	3.21

Ниже представлены в качестве примера результаты расчётов токсичности для 2-(3-(N-бензилацетамидо)-2,4,6-трийодфеноксигексановой кислоты ($\log(1/LD_{50})\_{exper}$ = 0,70) с помощью данного подхода при различных уровнях структурного сходства:

Tc≥0,30
Уравнения для структурных соседей:
$\log(1/LD_{50})\_{exper} = 1{,}50(\pm 0{,}10) - 5{,}97(\pm 0{,}37)^{*}\ \Sigma Cad/\alpha$ (3)
n=209, R^2=0,551, sd=0,43, Q^2=0,541, sd_{cv}=0,43, F=254,1
$\log(1/LD_{50})\_{calc}$ =0,40

Tc≥0,40
$\log(1/LD_{50})\_{exper} = 1{,}30(\pm 0{,}19) - 5{,}00(\pm 0{,}85)^{*}\ \Sigma Ca(o)d/\alpha$ (4)
n=85, R^2=0,293, sd=0,29, Q^2=0,229, sd_{cv}=0,31, F=34,4
$\log(1/LD_{50})\_{calc}$ =0,38

Tc≥0,50
$\log(1/LD_{50})\_{exper} = 1{,}30(\pm 0{,}19) - 5{,}00(\pm 0{,}85)^{*}\ \Sigma Cad/\alpha$ (5)
n=49, R^2=0,509, sd=0,25, Q^2=0,400, sd_{cv}=0,27, F=48,6
$\log(1/LD_{50})\_{calc}$ =0,39

Tc≥0,60
$\log(1/LD_{50})\_{exper} = 2{,}45(\pm 0{,}43) - 11{,}65(\pm 2{,}23)^{*}\ \Sigma \mathbf{E}ad/\alpha$ (6)
n=27, R^2=0,523, sd=0,12, Q^2=0,465, sd_{cv}=0,13, F=27,4
$\log(1/LD_{50})\_{calc}$ =0,43

Tc≥0,70
$\log(1/LD_{50})\_{exper} = 3{,}93(\pm 0{,}58) - 19{,}23(\pm 3{,}02)^{*}\ \Sigma \mathbf{E}ad/\alpha$ (7)
n=6, R^2=0,910, sd=0,07, Q^2=0,845, sd_{cv}=0,09, F=40,6
$\log(1/LD_{50})\_{calc}$ =0,59

В приведенном выше примере наблюдается изменение дескриптора, дающего наилучшую корреляцию в однопараметровых уравнениях, с изменением порога структурного сходства и числа соединений, участвующих в корреляционном уравнении. Однако, все указанные КССА модели сходны по физико-химической интерпретации, поскольку все эти дескрипторы ($\Sigma Cad/\alpha$, $\Sigma Ca(o)d/\alpha$, $\Sigma \mathbf{E}ad/\alpha$) отражают одно и тоже физико-химическое свойство: протоноакцепторную и протонодонорную способность единицы молекулярной поляризуемости.

Как уже отмечалось, токсичность рассматриваемого соединения никогда не включается в расчёты с помощью ЛРМСК модели. Соответственно любое соединение в исследуемой выборке или другой внешней базе данных может быть рассмотрено как независимое тестовое. Тем не менее, в настоящей работе были сформированы обучающая и тестовая выборки для проверки устойчивости полученных локальных КССА моделей. Для этого все соединения были ранжированы в порядке их структурного различия, используя следующий алгоритм программы MOLDIVS: a) выбор первого соединения, которое наиболее структурно отличается от остальных; b) выбор второго соединения, которое наиболее структурно отличается от первого; c) выбор третьего соединения, которое наиболее структурно отличается от первых двух и т.д. Затем в тестовую выборку было отобрано каждое пятое соединение из полученного ряда. Оставшиеся соединения были включены в обучающую выборку. Результаты применения ЛРМСК модели для обучающей и тестовой выборок представлены в таблице 3.

Таблица 3. Результаты расчётов острой токсичности при внутривенном введении мышам с помощью ЛРМСК подхода (HYBOT дескрипторы) для обучающей и тестовой выборки.

Tc	R между дескр-ми	ряд	a_0	er_a_0	a_1	er_a_1	R^2	SD	Q^2	SD_{cv}	F	N
		train	0,01	0,01	0,98	0,01	0,428	0,52	0,428	0,52	4403,4	5876
0,3	0,4	test	0,05	0,02	0,92	0,03	0,398	0,53	0,394	0,53	971,8	1469
		train	0,04	0,01	0,93	0,01	0,553	0,45	0,552	0,45	4455,4	3604
0,4	0,4	test	0,07	0,02	0,89	0,03	0,538	0,46	0,535	0,46	1055,5	908
		train	0,09	0,02	0,84	0,02	0,576	0,42	0,574	0,42	3011,3	2222
0,5	0,4	test	0,09	0,02	0,83	0,03	0,59	0,43	0,577	0,44	803,2	560

Из представленных данных очевидна стабильность полученных ЛРМСК моделей и эффективность использования концепции структурного сходства для расчета активности химических соединений.

ЗАКЛЮЧЕНИЕ. На основе концепции структурного сходства и локальных регрессионных моделей предложен КССА подход, позволяющий рассчитывать активность (токсичность) больших массивов органических соединений разнообразных химических классов. Указанный подход успешно применён к расчёту токсичности 7759 соединений, каждое из которых имеет в рассматриваемой выборке достаточное количество структурно-родственных соединений для построения собственной локальной регрессионной модели соответствующего кластера.

Работа выполнена при частичной финансовой поддержке МНТЦ (проект №3777).

ЛИТЕРАТУРА

1. *Maggiora G.M.* (2006) J. Chem. Inf. Model., **46**(4), 1535.
2. *Raevsky O.A.* (2001) SAR QSAR in Environ. Res., **12**(4), 367-381.
3. *Schaper K.-J., Kunz B., Raevsky O.A.* (2003) QSAR & Comb.Sci., **22**, 943-958.
4. *Raevsky O.A., Raevskaja O.E., Schaper K.-J.* (2004) QSAR & Comb.Sci. **23**, 327-343.
5. *Raevsky O.A., Trepalin S.V., Trepalina E.P., Gerasimenko V.A., Raevskaja O.E.* (2002) J. Chem. Inf. Comput. Sci., **42**(3), 540-549.
6. *Raevsky O.A., Schaper K.-J., Artursson P., McFarland J.W.* (2001) Quant. Struct.-Act .Relat., **20**, 402-413.
7. *Raevsky O.A., Dearden J.C.* (2004) SAR QSAR in Environ. Res., **15**(5/6), 433-448.
8. *Wold S., Sjostrom M.* (1977) in: Chemometrics: Theory and Application (Kowalski B.R., ed.) ACS Symp. Series №52, ACS, Washington, pp. 243-282.
9. *Раевский О.А., Чистяков В.В., Агабекян Р.С., Сапегин А.М., Зефиров Н.С.* (1990) Биоорг. химия, **16**, 509-522.
10. *Yuan H., Wang Y., Chang Y.* (2007) J. Chem. Inf. Model., **47**(1), 159-169.
11. *Raevsky O.A., Sapegin A.M., Zefirov N.S.* (1994) Quant. Struct.-Act. Relat., **13**, 412-418.
12. *Guha R., Dutta D., Jurs P.C., Chen T.* (2006) J. Chem. Inf. Model., **46**(4), 1836-1847.
13. *Gunturi S.B., Archana K., Khandelwal A., Narayanan R.* (2008) QSAR Comb. Sci., **27**, 1305-1317.

14. *Tsakovska I., Lessigiarska I., Netzeva T., Worth A.* (2008) QSAR & Combi. Sci., **27**, 41-48.
15. *Devillers J., Devillers H.* (2009) SAR QSAR Environ. Res., **20**(5/6), 467-500.
16. SYMYX Toxicity Database. Available at htttp://www.symyx.com/products/databases/index.jsp
17. *Raevsky O.A.* (1997) in: Computer-Assisted Lead Finding and Optimization (Waterbeemd H., Testa B., Folkers G., eds.), Verlag, Basel, pp. 367-378.
18. DRAGON, Talete srl, Italy. Available at http://www.talete.mi.it/dragon.htm
19. *Gerasimenko V.A., Trepalin S.V., Raevsky O.A.* (2000) in: Molecular Modeling and Prediction of Bioactivity (Gundertofte K., Jorgensen F.S., eds.), Kluwer Academic/Plenum Publishers, New York, pp. 423-424.

Поступила: 14. 09. 2010.

ACUTE INTRAVENOUS TOXICITY TO MICE CALCULATIONS ON THE BASIS LOCAL REGRESSION MODELS IN SUPEROVERLAPPING CLUSTERS (LRMSC)

***O.A. Raevsky*[1]*, V.Ju. Grigor'ev*[1]*, E.A. Liplavskaya*[1]*, A.P. Worth*[2]**

[1]Department of Computer-Aided Molecular Design, Institute of Physiologically Active Compounds of Russian Academy of Sciences, 142432, Chernogolovka, Moscow region, Russia; fax: 8(49652)49-508; e-mail: raevsky@ipac.ac.ru
[2]Institute for Health and Consumer Protection, European Commission - Joint Research Centre, Via Enrico Fermi 2749, 21027 Ispra (Va), Italy

Modeling of quantitative structure – activity relationships between physicochemical descriptors of organic chemicals and their acute intravenous toxicity in mice have been presented. This approach includes three steps: structure-similarity chemicals selection for every chemical-of-interest (clusterization); construction of quantitative structure – toxicity models for every cluster (without including of chemical-of-interest); application of obtained QSAR equations for chemical-of-interest toxicity estimation. This approach has been applied for acute intravenous toxicity calculations of 10241 organic chemicals. For 7759 chemicals which has enough quantity of structural neighbours with the Tanimoto index (Tc) on the level 0.30 and over, a standard deviation of calculation vs. experimental log(1/LD_{50}) values is equal to 0.51 at the estimation of experimental determination on the level 0.50. The results of calculations isn't so good for remain chemicals (~24%). It is connect with absence of sufficient number of structure similarity neighbours. It's assumed this QSAR approach can be useful for activity and toxicity prediction of chemicals large sets.

Key words: QSAR, toxicity, structural similarity, HYBOT, DRAGON, clusterization, regression models.