

На правах рукописи

ГРИГОРЬЕВ Вениамин Юрьевич

**КОЛИЧЕСТВЕННЫЕ МОДЕЛИ «СТРУКТУРА–СВОЙСТВО»
ОРГАНИЧЕСКИХ СОЕДИНЕНИЙ**

02.00.03 – органическая химия

02.00.04 – физическая химия

Автореферат
диссертации на соискание ученой степени
доктора химических наук

Черноголовка – 2013

Работа выполнена в Федеральном государственном бюджетном учреждении науки Институте физиологически активных веществ РАН

Официальные
оппоненты:

Балакин Константин Валерьевич, доктор химических наук, Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования Московский физико-технический институт, г. Москва, заведующий кафедрой общей химии

Дьячков Павел Николаевич, доктор химических наук, профессор, Федеральное государственное бюджетное учреждение науки Институт общей и неорганической химии им. Н.С. Курнакова РАН, г. Москва, ведущий научный сотрудник лаборатории квантовой химии

Клюев Михаил Васильевич, доктор химических наук, профессор, Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования Ивановский государственный университет, г. Иваново, заведующий кафедрой органической и физической химии

Ведущая организация: Федеральное государственное бюджетное учреждение науки Институт биохимической физики им. Н.М. Эмануэля РАН, г. Москва

Защита состоится «10» декабря 2013 г. в 14:00 на заседании диссертационного совета Д 002.102.01 при Институте физиологически активных веществ РАН по адресу: 142432, Московская обл., г. Черноголовка, Северный проезд, д. 1.

С диссертацией можно ознакомиться в библиотеке ИФАВ РАН.

Автореферат разослан « » _____ 2013 г.

Ученый секретарь диссертационного совета,
кандидат химических наук

Афанасьева С.В.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. Одной из фундаментальных задач химии является установление количественной связи между структурой вещества и его свойствами. К настоящему времени в мире синтезировано несколько десятков миллионов химических соединений, большинство из которых относится к органическим веществам. Эти соединения могут быть охарактеризованы большим количеством свойств, общее число которых, по некоторым оценкам, приближается к 10000.

Для многих соединений отсутствуют данные об их физико-химических и биологических свойствах, например, о способности к образованию водородной связи или об острой токсичности. Но распределение Н-донорных и Н-акцепторных центров и величины их активности нужны при конструировании новых биологически активных соединений, а знание количественных характеристик токсичности необходимо для оценки опасности промежуточных и конечных веществ при планировании и проведении многостадийного органического синтеза.

Ведущей задачей, которую решает органическая химия, является синтез соединений с полезными свойствами. Одним из рациональных путей достижения этой цели является компьютерный молекулярный дизайн на основе математических моделей «количественная связь структура – свойство» (QSPR)/«количественная связь структура – активность» (QSAR). Компьютерное конструирование QSPR/QSAR моделей требует гораздо меньше времени и материальных ресурсов, чем экспериментальные исследования, а созданные модели позволяют оценивать свойства неисследованных химических соединений и предсказывать новые структуры с заданными свойствами.

Работа выполнена в соответствии с планами научно-исследовательских работ Института физиологически активных веществ РАН «Компьютерный молекулярный дизайн, включая количественную связь структура-свойство (активность) и молекулярное моделирование взаимодействия лиганд-мишень для структурной генерации соединений с заданными свойствами» (№

госрегистрации 0120.0 504023) при финансовой поддержке Российского фонда фундаментальных исследований: проект № 97-03-33068-а «Молекулярное моделирование связи структура-активность с учетом трехмерности структур и их способности к образованию водородных связей» и Международного научно-технического центра: проект № 888 «Компьютерное предсказание химической экотоксичности» и проект № 3777 «Развитие и применение КССА моделей токсичности химических соединений по отношению к грызунам».

Цель работы. Разработка количественных моделей «структура – свойство» и «структура – активность» с использованием дескрипторов водородной связи и других молекулярных дескрипторов для компьютерного дизайна органических соединений, включая прогнозирование физико-химических свойств и биологической активности неисследованных соединений и предсказание новых структур с заданными свойствами.

Научная новизна. На основе литературных и собственных экспериментальных данных разработана физико-химическая аддитивно-мультипликативная модель энтальпии и свободной энергии Н-комплексов, сконструирована единая энергетическая шкала водородной связи и создана компьютерная программа для молекулярного дизайна органических соединений с заданной величиной Н-донорной и Н-акцепторной активности.

Создана модель липофильности органических соединений, основанная на линейной комбинации двух физико-химических параметров: молекулярной поляризуемости и Н-акцепторного дескриптора водородной связи.

С использованием молекулярных дескрипторов, в том числе дескрипторов Н-связи, сконструированы оригинальные кластерные, классификационные и регрессионные модели «структура – активность» для компьютерного дизайна органических соединений, включая оценку острой токсичности, ретардантной активности и способности проникновения через гематоэнцефалический барьер неисследованных новых соединений и предсказание структур с заданным уровнем биологической активности.

Впервые разработана рекуррентная модель для описания связи «структура – активность» в гомологических рядах органических соединений, отличающаяся от известных моделей отсутствием дескрипторов.

Впервые предложен, разработан и применен новый 3D дескриптор для оценки фрактальных свойств молекулярных гистограмм межатомных расстояний.

Впервые предложена, разработана и применена линейная динамическая модель для компьютерного молекулярного дизайна 3D структур с заданными свойствами.

Практическая значимость работы. Значительная часть диссертационной работы была выполнена в рамках участия в разработке в ИФАВ РАН двух баз данных по H-комплексобразованию органических соединений. При этом база данных по термодинамическим характеристикам H-комплексов, насчитывающая свыше 13000 записей, нашла использование в качестве источника информации при проведении теоретических и практических исследований в области органической и физической химии. База данных по протонодонорным и протоноакцепторным дескрипторам H-связи, содержащая свыше 50000 записей, а также компьютерная программа HYBOT, созданная на ее основе, служат для количественной оценки способности органических соединений к образованию водородной связи и нашли применение при проведении QSPR/QSAR исследований. Разработанные в диссертации модели могут использоваться как для количественной оценки физико-химических свойств и биологической активности новых неизученных органических соединений, так и для предсказания структур с заданными свойствами.

Личный вклад автора. Все результаты диссертации получены лично автором или в соавторстве при его непосредственном участии. В работах, выполненных в соавторстве, вклад автора состоял в непосредственном участии во всех этапах работы, начиная от постановки задачи и завершая обсуждением результатов. Синтез, очистку и идентификацию новых физиологически активных соединений выполнили сотрудники ИФАВ РАН: Иванов А.Н.,

Соколов В.Б., Чепакова Л.А., Брель В.К., Додонов М.В., Прошин А.Н., Чистяков В.Г. Биологические испытания на ретардантную активность провел Кокурин А.В. (ВНИИСБ РАСХН). Вклад автора в программный комплекс НУВОТ, разработанный совместно с Раевским О.А., Трепалиным С.В. и Раевской О.Е. (ИФАВ РАН), заключался в создании термодинамической базы данных и базы дескрипторов Н-связи.

Автор выражает глубокую признательность доктору химических наук, профессору Раевскому О.А. за ценные научные консультации во время выполнения работы.

Апробация работы. Результаты исследования и основные положения работы были представлены на следующих конференциях, симпозиумах и конгрессах: IX Всесоюзная конференция «Химическая информатика» (Черноголовка, 1992); 12 Международный симпозиум по медицинской химии (Базель, Швейцария, 1992); 9 Европейский симпозиум по количественным соотношениям структура-активность (Страсбург, Франция, 1992); IV Российский конгресс «Человек и лекарство» (Москва, 1997); VI Российский конгресс «Человек и лекарство» (Москва, 1999); 4 Международный симпозиум по компьютерным методам в токсикологии и фармакологии (Москва, 2007); XVII Российский конгресс «Человек и лекарство» (Москва, 2010); 18 Европейский симпозиум по количественным соотношениям структура-активность (Родес, Греция, 2010); XVIII Российский конгресс «Человек и лекарство» (Москва, 2011), XIX Российский конгресс «Человек и лекарство» (Москва, 2012).

Публикации. По теме диссертации опубликовано 40 печатных работ, включая 20 статей в рецензируемых российских и международных журналах, получено 3 свидетельства об официальной регистрации программ для ЭВМ.

Объем и структура диссертации. Работа изложена на 324 страницах машинописного текста, содержит 37 таблиц и 33 рисунка. Диссертация состоит из введения, шести глав, заключения, списка литературы из 427 наименований.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность работы, сформулированы цель и задачи исследования, показана научная новизна и практическая значимость полученных результатов.

Глава 1. Современное состояние в области создания количественных моделей «структура – свойство» и «структура – активность» (на примере острой токсичности органических соединений). Из анализа литературы следует, что во многих публикациях отмечается важная роль водородной связи при формировании биологической активности. Однако при этом способность соединений к образованию Н-связи в разработанных моделях выражается на полуколичественном уровне в виде числа доноров и акцепторов, без учета молекулярного окружения. Очевидно, что качество моделей «структура – активность» может быть улучшено при применении адекватных дескрипторов водородной связи.

Часто используемым дескриптором для конструирования моделей «структура – активность» органических соединений является липофильность в виде коэффициента распределения вещества в системе н-октанол-вода (K_{ow}). Однако физико-химическая природа этого дескриптора выявлена не в полной мере. Кроме того, он является экспериментально определяемой величиной, что создает трудности при моделировании свойств больших рядов соединений. Одним из путей для решения этой проблемы может служить замена K_{ow} другими дескрипторами.

Число глобальных моделей «структура – активность», играющих важную роль при быстрой предварительной оценке биологической активности большого числа разнообразных органических соединений, мало по сравнению с большим числом локальных моделей. Необходимо качественное улучшение и количественное увеличение числа глобальных моделей.

Также можно сделать вывод о недостаточном числе исследований в области развития методологии QSPR/QSAR. В первую очередь, это относится к

разработке новых дескрипторов и статистических моделей «структура – свойство».

Глава 2. Экспериментальная часть. Описаны физико-химические и биологические данные; расчет, преобразование и отбор дескрипторов; меры структурного сходства; статистические методы исследования и валидация моделей; методика проведения спектральных и калориметрических измерений.

Глава 3. Количественные модели «структура – физико-химические свойства» органических соединений

3.1. Аддитивно-мультипликативная модель энтальпии и свободной энергии водородной связи. Монофункциональные соединения

Из литературы были отобраны данные, содержащие сведения о термодинамических характеристиках Н-комплексов и удовлетворяющие следующим условиям: 1) состав 1:1, с одним водородным мостиком; 2) растворитель CCl_4 ; 3) для каждого комплекса имеются значения ΔH (кДж/моль) и ΔG (кДж/моль); 4) ΔH и ΔG рассчитаны на основе экспериментальных данных и соответствуют температуре 298 К. Общее число Н-комплексов, образованных с участием 163 Н-доноров (NH-, OH-, SH-, CH-доноры) и 195 Н-акцепторов (N-, O-, S-, π -акцепторы), составило 936. Для установления соотношения между термодинамическими характеристиками и дескрипторами Н-связи была использована аддитивно-мультипликативная модель:

$$\Delta H = k_1 E_a E_d + k_0 \quad (1)$$

$$\Delta G = k_1^* C_a C_d + k_0^* \quad (2)$$

где k_0 , k_1 , k_0^* , k_1^* – параметры; E_a – энтальпийный протоноакцепторный дескриптор; E_d – энтальпийный протонодонорный дескриптор; C_a – свободноэнергетический протоноакцепторный дескриптор; C_d – свободноэнергетический протонодонорный дескриптор. Для расчета параметров уравнений (1) и (2) необходима фиксация E и C дескрипторов для одного (стандартного) Н-донора и одного (стандартного) Н-акцептора. Эти соединения должны удовлетворять двум требованиям: 1) для них должно быть известно достаточно большое количество Н-комплексов и 2) они должны быть

удобными при экспериментальном применении. В качестве стандартного Н-донора был выбран фенол ($E_d=-2.50$; $C_d=-2.50$), в качестве стандартного Н-акцептора – гексаметилфосфортриамид (ГМФТА) ($E_a=2.50$; $C_a=4.00$). Величины дескрипторов для стандартных соединений были выбраны таким образом, чтобы сделать шкалу с интервалом изменения дескрипторов от 0.00 до ($\cong -5.00$) для Н-доноров и от 0.00 до ($\cong 5.00$) для Н-акцепторов.

Ввиду отсутствия в отобранном массиве данных Н-комплексов с положительной величиной ΔH , коэффициент k_0 в уравнении (1) был положен равным 0. Учитывая, что нижняя граница надежного определения констант Н-комплексов (1:1) соответствует величине приблизительно равной 0.1 л/моль, коэффициент k_0^* (уравнение (2)) был зафиксирован на уровне 5.70 кДж/моль.

При известных величинах ΔH и ΔG для комплексов исследуемого соединения и стандарта с одним и тем же партнером из уравнений (1) и (2) можно рассчитать Н-донорные или Н-акцепторные дескрипторы соединения. Путем итеративного повторения этой процедуры удалось оценить величины дескрипторов для 347 соединений. При этом величины E_d менялись в пределах от -0.34 (C_6HCl_5) до -3.93 (CCl_3COOH), C_d от -0.00 (C_6HCl_5) до -4.78 (CCl_3COOH), E_a от 0.50 (C_6H_6) до 5.21 ($(C_4H_9)_2NH$), C_a от 0.43 (CCl_3CN) до 4.99 (1,4-дiazобикакло[2,2,2]октан). Путем варьирования и минимизации были определены коэффициенты k_1 и k_1^* моделей (1) и (2):

$$\Delta H_{\text{расч}}(\text{кДж/моль}) = 4.96 E_a E_d \quad (3)$$

$$\Delta G_{\text{расч}}(\text{кДж/моль}) = 2.43 C_a C_d + 5.70 \quad (4)$$

Адекватность полученных моделей была подтверждена путем сопоставления экспериментальных и рассчитанных величин ΔH и ΔG :

$$\Delta H_{\text{расч}} = -0.27(\pm 0.45) + 1.00(\pm 0.02) \Delta H_{\text{эксп}} \quad (5)$$

$$n=936; r=0.954; s=2.70; F=9553$$

$$\Delta G_{\text{расч}} = -0.07(\pm 0.12) + 1.00(\pm 0.01) \Delta G_{\text{эксп}} \quad (6)$$

$$n=936; r=0.984; s=1.11; F=28556$$

Статистические характеристики уравнений (5) и (6) свидетельствуют о том, что аддитивно-мультипликативная модель дает возможность реальной

оценки ΔH и ΔG и позволяет надежно рассчитать дескрипторы водородной связи. Величины дескрипторов Н-связи органических соединений являются базой для создания единой энергетической шкалы водородной связи (рис.1). Полная количественная оценка способности органического соединения к образованию Н-связи может быть выполнена на основе четырех дескрипторов: E_a , E_d , C_a , C_d . Из представленных данных ясно, что молекулярное окружение оказывает существенное влияние на величины дескрипторов активных центров. Также можно констатировать, что разные по природе химические соединения могут иметь одинаковую способность к образованию водородной связи.

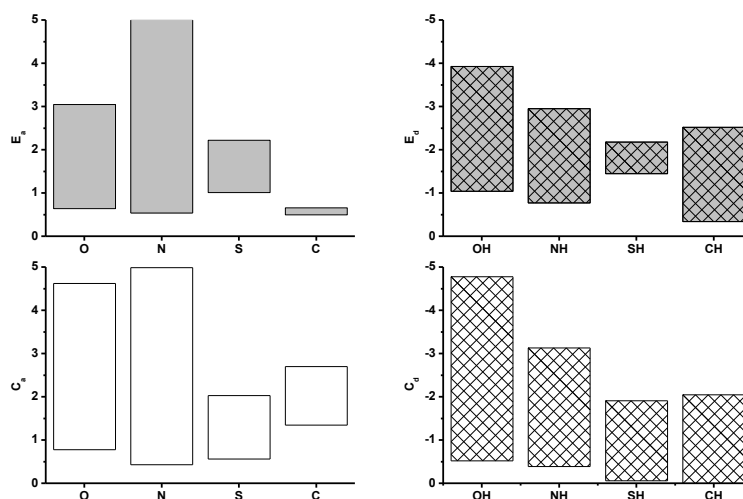


Рис. 1. Единая энергетическая шкала водородной связи

В литературе, кроме данных по Н-комплексам в CCl_4 , приводятся величины ΔH и ΔG в других неполярных растворителях. С этими данными была проведена работа аналогичная рассмотренной выше. Из анализа полученных результатов можно сделать вывод о том, что дескрипторы Н-связи сохраняют приблизительно одинаковое значение в разных растворителях, при этом происходит только изменение коэффициентов k_1 и k_1^* в уравнениях (1) и (2).

3.2. Аддитивно-мультипликативная модель энтальпии и свободной энергии водородной связи. Полифункциональные соединения

Для проверки работоспособности аддитивно-мультипликативной модели Н-связи в случае сложных органических соединений были проведены

экспериментальные исследования с использованием ИК-спектрофотометрии, калориметрии и математического моделирования. Они включали в себя установление стехиометрии, определение взаимодействующих атомов и групп, расчет величин ΔH , ΔG для H-комплексов и оценку величины дескрипторов H-связи. В исследовании принимали участие 22 фосфорилированных и тиофосфорилированных оксима общей формулы: $(R_1)(R_2)P(X)ONC(R_3)(R_4)$ (где $R_1=Me, MeO, EtO, PrO, BuO$; $R_2=MeO, EtO, PrO, i-PrO, BuO, AmO$; $R_3=Me, Et, Pr, CH_2Cl$; $R_4=Me, Et, Pr, Cl$; $X=O, S$), которые можно рассматривать в качестве аналогов инсектицида «валексон» $(EtO)_2P(S)ONC(C_6H_5)CN$ и 10 дифторнитроацетанилидов общей формулы: $(4-R_5)(3-R_6)C_6H_3NHC(O)CF_2NO_2$ (где $R_5=H, CH_3, OCH_3, F, Cl, NO_2$; $R_6=H, CH_3, CF_3, F, NO_2$), которые можно рассматривать в качестве аналогов гербицида «пропанид» $3,4-Cl_2-C_6H_3NHC(O)C_2H_5$.

Вначале были установлены величины ΔH и ΔG при образовании H-комплексов между фосфорилированными оксимами и фенолом, а также между дифторнитроацетанилидами и ГМФТА. Это позволило рассчитать дескрипторы H-связи соответствующих соединений.

Следующим шагом стало изучение комплексов с участием этих соединений. При этом фосфорилированные оксимы выступали в качестве H-акцепторов, а дифторнитроацетанилиды – в качестве H-доноров (табл. 1). При сравнении экспериментальных и рассчитанных на основе аддитивно-мультипликативной модели термодинамических величин можно установить, что в случае ΔH минимальное отклонение (по абсолютной величине) составляет 0.2 кДж/моль, максимальное 1.3 кДж/моль и среднее 0.6 кДж/моль, а для ΔG 0.1 кДж/моль, 1.1 кДж/моль и 0.8 кДж/моль соответственно. Таким образом, можно констатировать вполне удовлетворительное совпадение экспериментальных и рассчитанных термодинамических характеристик H-комплексов.

Таблица 1. Экспериментальные и рассчитанные термодинамические характеристики Н-комплексов фосфорилированных оксимов с дифторнитроацетанилидами (бензол; 298 К; 1:1) и дескрипторы Н-связи (E_a , C_a , E_d , C_d)

№	Н-акцептор		№	Н-донор		-ΔН, кДж/моль		-ΔG, кДж/моль	
	E_a	C_a		$-E_d$	$-C_d$	Эксп.	Расч.*	Эксп.	Расч.*
II	1.71±0.09	3.08±0.11	V	2.71±0.06	2.41±0.07	18.9±1.0	20.2±1.5	10.5±0.6	10.4±1.0
II	1.71±0.09	3.08±0.11	V	2.71±0.06	2.41±0.07	18.9±1.0	20.2±1.5	10.5±0.6	10.4±1.0
II	1.71±0.09	3.08±0.11	VI	2.48±0.02	3.01±0.06	18.9±1.0	18.5±1.1	13.8±0.6	14.4±1.1
III	1.73±0.09	2.84±0.11	V	2.71±0.06	2.41±0.07	19.4±1.0	20.4±1.5	9.98±0.6	9.24±1.0
III	1.73±0.09	2.84±0.11	VI	2.48±0.02	3.01±0.06	18.9±1.0	18.7±1.1	13.5±0.6	12.4±1.1
IV	1.75±0.09	3.04±0.11	V	2.71±0.06	2.41±0.07	19.8±1.0	20.6±1.5	11.0±0.6	10.3±1.0
IV	1.75±0.09	3.04±0.11	VI	2.48±0.02	3.01±0.06	19.6±1.0	18.9±1.1	15.3±0.6	14.2±1.1

Примечание: * ΔН = 4.35 E_a E_d ; ΔG = 5.70 + 2.17 C_a C_d ; (Me)(BuO)P(O)ONC(Me)(Cl) (II); (Me)(BuO)P(O)ONC(Et)(Cl) (III); (Me)(BuO)P(O)ONC(Pr)(Cl) (IV); 4-CH₃-C₆H₄NHC(O)CF₂NO₂ (V); 4-NO₂-C₆H₄NHC(O)CF₂NO₂ (VI).

3.3. Модель липофильности органических соединений

Для количественного описания липофильности было отобрано 71 простое органическое соединение, содержащее по одной Н-акцепторной группе. В состав этих групп входили атомы O, N, S, F, Cl, Br, C. В качестве независимых переменных были использованы молекулярная поляризуемость (α) и свободноэнергетические Н-акцепторные дескрипторы (C_a). Интервал изменения этих величин составлял 3.25÷21.44 Å³ и 0.20÷3.12 соответственно. Величины logK_{ow} менялись в пределах от -0.97 до 4.75. В результате была получена следующая модель:

$$\log K_{ow} = 0.266(\pm 0.006) \alpha - 1.00(\pm 0.05) C_a \quad (7)$$

$$n=71; r^2=0.982; s=0.18; F=3829; q^2=0.981$$

Статистические критерии уравнения (7) свидетельствуют о возможности количественной оценки значений logK_{ow} на основе рассчитанных величин α и C_a . Внешняя валидация модели была проведена путем расчета аналогичной модели с использованием другого ряда из 98 соединений, содержащих Н-акцепторные атомы O, N, Cl, Br, I, C. Интервал изменения величин составлял -1.01÷5.15, 2.25÷28.77 Å³, 0.00÷2.96 для logK_{ow}, α , C_a соответственно. Параметры полученной модели (8) оказались близки к параметрам модели (7):

$$\log K_{ow} = 0.276(\pm 0.008) \alpha - 1.01(\pm 0.07) C_a \quad (8)$$

$$n=98; r^2=0.961; s=0.29; F=1191; q^2=0.959$$

Преобразование уравнения (7) к следующему виду:

$$C_a^* = 0.266 \alpha - \log K_{ow} \quad (9)$$

дает возможность рассчитать дескрипторы C_a^* соединений, с использованием величин α и $\log K_{ow}$.

Имеется много опубликованных данных по экспериментальным величинам $\log K_{ow}$ как простых, так и сложных органических соединений. Это дает возможность на основе уравнения (9) оценить Н-акцепторную способность тех атомов и функциональных групп, которые не могли быть определены в рамках шкалы, основанной на термодинамических параметрах (табл. 2).

Таблица 2. Свободноэнергетические Н-акцепторные дескрипторы, полученные на основе липофильности (C_a^*), свободной энергии (C_a) и их разность (ΔC_a) для органических соединений с несколькими гетероатомами

№	Соединение	C_a^*	C_a	ΔC_a
1	м-Нитробензойная кислота	0.62	0.47	0.15
2	Бензойная кислота	1.03	0.76	0.27
3	Этилтрихлорацетат	1.21	1.11	0.10
4	Уксусная кислота	1.47	1.10	0.37
5	Метилацетат	1.60	1.40	0.20
6	Бутиролактон	2.71	2.05	0.66
7	N,N-Диметилацетамид	3.02	2.76	0.26
8	Мочевина	3.06	2.50	0.56
9	N-Метилацетамид	3.15	2.70	0.45
10	2-Пиперидон	3.34	2.95	0.39

Представленные данные свидетельствуют о том, что в случае сложных органических соединений величина C_a^* превышает значение дескриптора для наиболее сильного Н-акцептора (C_a) и, вероятно, соответствует сумме свободноэнергетических Н-акцепторных дескрипторов в молекуле, в то время как их разность (ΔC_a) характеризует потенциальную возможность к образованию Н-связи слабых Н-акцепторных центров в молекуле. Это открывает широкие возможности для существенного увеличения числа соединений с количественно охарактеризованной Н-акцепторной способностью.

3.4. Компьютерный программный комплекс НУВОТ

Комплекс НУВОТ предназначен для расчета дескрипторов Н-связи и молекулярного дизайна новых органических соединений с заданной величиной Н-донорной или Н-акцепторной активности. Он состоит из трех главных компонентов. Во-первых, это база данных (БД) по термодинамике водородной связи, которая содержит экспериментальные величины ΔH и ΔG Н-комплексов в различных средах. В настоящее время эта БД насчитывает 13688 записей (т.е. сведений об одной паре реакций Н-донор/Н-акцептор).

Другой составной частью НУВОТ является БД дескрипторов, содержащая Н-донорные и Н-акцепторные дескрипторы органических соединений, рассчитанных в рамках общей шкалы водородной связи. Эта БД содержит 4 библиотеки дескрипторов: E_d (1025 дескрипторов), E_a (2060 дескрипторов), C_d (2835 дескрипторов), C_a (46372 дескриптора).

Третьей составной частью НУВОТ является программа, которая рассчитывает дескрипторы органических соединений на основе их 2D структуры с использованием баз данных, содержащих термодинамические характеристики Н-комплексов и дескрипторы Н-связи.

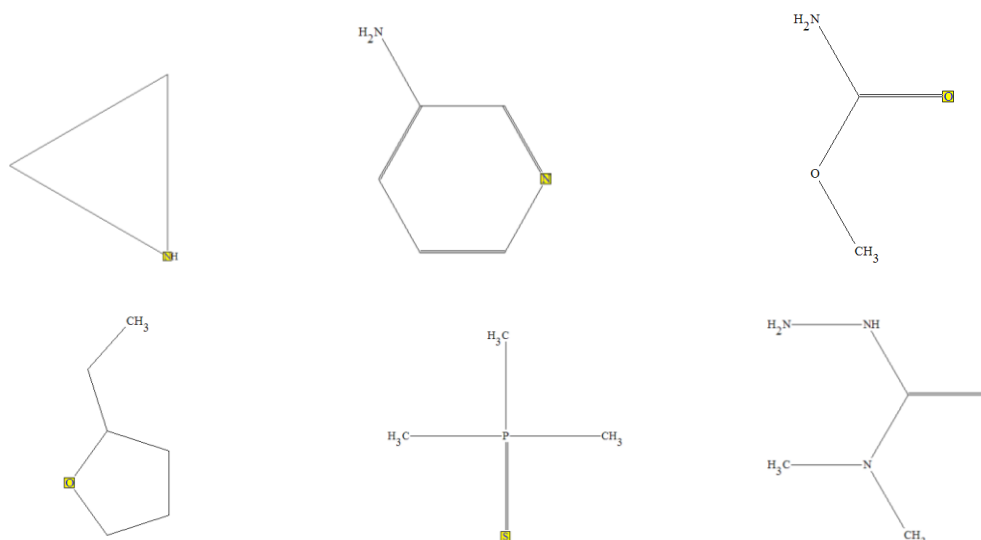
Известно, что водородная связь играет важную роль для проявления многих видов физико-химических свойств и биологической активности, поэтому ее учет при проектировании соединений для целевого органического синтеза является необходимой и важной задачей. Дизайн новых органических соединений с использованием НУВОТ может быть проведен как путем решения прямой задачи (расчет дескрипторов Н-связи для известных структур), так и путем решения обратной задачи (генерация структур/подструктур с известной величиной Н-дескриптора). Пример расчета молекулы диклофенака в рамках прямой задачи приведен на рис. 2. Из представленных данных, в частности, следует, что максимальные величины дескрипторов C_a , E_a , C_d , E_d составляют 1.50, 2.00, -2.50, -2.90 соответственно. При этом произведена полная идентификация Н-активных центров (атомов) в молекуле. Так выделенный атом кислорода является Н-акцепторным центром и характеризуется величинами $C_a=1.42$ и $E_a=1.50$.

Alpha	29.52	maxEa	2.00	Sum(Ea)	9.85	Sum(Cd)/A	-0.127
max(O-)	-0.28	maxCa	1.50	Sum(Ca)	3.45	Sum(Cad)/A	0.244
max(O+)	0.25	maxEd	-2.90	Sum(Co)	-3.75		
Sum(O-)	-0.91	maxCd	-2.50	Sum(Cod)	7.20		
Sum(O+)	0.91	maxEa*maxEd	-5.79	Sum(Ea)/A	0.174		
Sum(O0)	1.83	maxCa*maxCd	-3.75	Sum(Ed)/A	-0.160		
Sum(O-)/A	-0.031	Sum(Ea)	5.13	Sum(Ead)/A	0.334		
Sum(O+)/A	0.031	Sum(Ed)	-4.72	Sum(Co)/A	0.117		

Polarizability	Charge	E	C
1.09	-0.09	2.00	0.88
1.35	0.07	0.08	0.08
1.35	-0.06	0.10	0.08
1.35	-0.08	0.06	0.08
1.35	-0.02	0.08	0.08
1.35	-0.03	0.06	0.08
1.35	-0.02	0.08	0.08
1.35	0.01	0.08	0.04
1.35	-0.02	0.06	0.04
1.35	-0.02	0.06	0.04
1.35	-0.04	0.06	0.04
1.35	-0.04	0.06	0.04
1.35	-0.01	0.06	0.04
2.32	-0.00	0.35	0.10
2.32	-0.00	0.35	0.10
1.06	-0.00		
1.35	0.25		
0.27	-0.28	0.14	0.15
0.39	-1.03	-1.25	-1.25
0.39	0.09	-1.03	-1.25
0.39	0.04		
0.39	0.04		
0.39	0.04		
0.39	0.04		
0.39	0.04		
0.39	0.04		
0.39	0.04		
0.39	0.05		
0.39	0.11	-2.90	-2.50

Рис. 2. N-дескрипторы для молекулы диклофенака

В рамках решения обратной задачи рассмотрим случай, когда для органического синтеза необходимо иметь набор подструктур с заданной величиной N-акцепторной активности $C_a=2.00$. С помощью НУВОТ было найдено 375 соответствующих структур. Некоторые из них приведены ниже.



Из представленных данных очевидно, что одинаковую величину N-акцепторной активности могут иметь различные по природе атомы, с различным типом гибридизации и молекулярным окружением, входящие в состав ациклических, циклических и ароматических структур.

Глава 4. Количественные локальные модели «структура – биологическая активность» органических соединений

4.1. Классификационные и регрессионные модели неполярного наркоза

В работе были использованы величины острой токсичности (LC_{50} , ммоль/л, 96 час., лит. данные) 118 органических соединений с типом токсического действия (MOA, mode of action) неполярный наркоз по отношению к рыбам *Poecilia reticulata* (*Pr*), *Pimephales promelas* (*Pp*) и *Oncorhynchus mykiss* (*Om*). Молекулярные дескрипторы соединений рассчитывали с использованием программ HYBOT, DNESTR и SYBYL. Фрагментные дескрипторы рассчитывали с помощью программы MOLDIVS. Классификация выполнена на основе линейного дискриминантного анализа (LDA) с применением ряда из 4-х HYBOT дескрипторов α , $\sum q$, $\sum C_a$, $\sum C_d$ (LDA_4D), ряда из 10-12 главных компонент (LDA_PC), и метода структурного сходства (SS) (аналог метода k-ого ближайшего соседа) с использованием фрагментных дескрипторов и метрики в виде коэффициентов Танимото.

Отдельные результаты классификации представлены в табл. 3. Высокая эффективность дискриминации была достигнута при использовании метода SS: общая степень правильного распознавания составляла 0.932 в обучающей выборке и 0.914 в тестовой выборке.

Таблица 3. Чувствительность (S_n), специфичность (S_p) и точность (Acc) при бинарной классификации химических соединений с MOA неполярный наркоз (класс 1) и другие MOA (класс 2) по отношению к *P. reticulata* (*Pr*) с использованием методов линейного дискриминантного анализа (LDA) и структурного сходства (SS)

№	Объект	Метод	Обучающая выборка				Тестовая выборка			
			n_1/n_2	S_n	S_p	Acc	n_1/n_2	S_n	S_p	Acc
1	<i>Pr</i>	LDA_4D	72/163	0.472	0.859	0.740	18/40	0.333	0.850	0.690
2	<i>Pr</i>	LDA_PC	72/163	0.639	0.834	0.774	18/40	0.556	0.825	0.741
3	<i>Pr</i>	SS_3	72/163	0.986	0.908	0.932	18/40	0.944	0.900	0.914

Примечание: n_1 – число соединений в первом классе, n_2 – число соединений во втором классе.

Учитывая высокую степень межвидовой корреляции значений токсичности для трех рыб ($r^2=0.94\div 0.99$), возможно построение общей регрессионной модели токсичности:

$$\log(1/LC_{50}) = -1.75(\pm 0.05) + 0.89(\pm 0.03) \log K_{ow} \quad (10)$$

$$n=150; r^2=0.936; s=0.36; F=2171$$

$$\log(1/LC_{50}) = -1.96(\pm 0.09) + 0.258(\pm 0.006) \alpha - 0.88(\pm 0.03) \sum C_a \quad (11)$$

$$n=150; r^2=0.943; s=0.34; F=1211$$

Область применимости (ОП) модели (11): химический класс: хлоралканы, хлорбензолы, эфиры, кетоны, спирты; интервал $\log(1/LC_{50})$: $-2.96 \div 3.15$; α : $3.25 \div 23.43 \text{ \AA}^3$; $\sum C_a$: $0.18 \div 5.77$; $\sum C_d$: $-4.16 \div 0.00$.

Главное преимущество модели (11) по сравнению с традиционной моделью (10) состоит в использовании не экспериментальных, а рассчитанных дескрипторов. Кроме того, уравнение (11) дает возможность количественной оценки вкладов с противоположными эффектами: молекулярной поляризуемости, которая увеличивает токсичность, и H-акцепторной способности, которая уменьшает ее.

4.2. Классификационные и регрессионные модели полярного наркоза

Для исследования была использована выборка, содержащая сведения по острой токсичности (LC_{50} , ммоль/л, 96 час., лит. данные) 134 соединений с МОА полярный наркоз по отношению к трем рыбам *Pr*, *Pp* и *Om*.

Использование метода LDA_4D привело к получению удовлетворительной дискриминации между соединениями (табл. 4). Так степень общего правильного распознавания для *Pr* составила 0.855 и 0.860 в обучающей и тестовой выборках соответственно. Однако метод SS обеспечивал получение более хороших результатов: в обучающей выборке $Acc=0.906$, а в тестовой $Acc=0.947$.

Высокий коэффициент корреляции межвидовой токсичности между *Pr* и *Pp* ($r^2=0.91$) свидетельствует о возможности создания общей модели:

$$\log(1/LC_{50}) = -0.78(\pm 0.06) + 0.72(\pm 0.02) \log K_{ow} \quad (12)$$

$$n=169; r^2=0.860; s=0.37; F=1022; q^2=0.857$$

$$\log(1/LC_{50}) = -1.02(\pm 0.14) + 0.192(\pm 0.007) \alpha - 0.68(\pm 0.04) \sum C_a - 0.11(\pm 0.02) \sum C_d \quad (13)$$

$$n=169; r^2=0.886; s=0.34; F=427; q^2=0.879$$

ОП модели (13): химический класс: нитробензолы, амины, анилины, фенолы; интервал $\log(1/LC_{50})$: $-1.54 \div 3.58$; α : $5.80 \div 30.26 \text{ \AA}^3$; $\sum C_a$: $0.48 \div 4.18$; $\sum C_d$: $-6.52 \div 0.00$.

Таблица 4. Чувствительность (Sn), специфичность (Sp) и точность (Acc) при бинарной классификации химических соединений с МОА полярный наркотик (класс 1) и другие МОА (класс 2) по отношению к *P. reticulata* (*Pr*) с использованием методов линейного дискриминантного анализа (LDA) и структурного сходства (SS)

№	Объект	Метод	Обучающая выборка			Тестовая выборка				
			n_1/n_2	Sn	Sp	Acc	n_1/n_2	Sn	Sp	Acc
1	<i>Pr</i>	LDA_4D	96/138	0.792	0.899	0.855	23/34	0.826	0.882	0.860
2	<i>Pr</i>	LDA_PC	96/138	0.833	0.899	0.872	23/34	0.870	0.882	0.877
3	<i>Pr</i>	SS_1	96/138	0.896	0.913	0.906	23/34	0.957	0.941	0.947

Примечание: n_1 – число соединений в первом классе, n_2 – число соединений во втором классе.

Необходимо отметить, что дескриптор $\log K_{ow}$ не дает информации о механизме действия, потому что он является композитным дескриптором, включающим в себя влияние нескольких факторов. А модели, использующие HUBOT дескрипторы для описания неполярного и полярного наркоза, позволяют заглянуть внутрь механизма неспецифической токсичности.

4.3. Линейный дискриминантно-регрессионный подход

Известно, что в основе кластерного анализа лежит группировка молекул с применением математической модели связи между молекулами с использованием формальных алгоритмов. Как показывает наш опыт и данные других исследователей применение кластерного анализа для создания QSPR/QSAR моделей во многих случаях не приводит к получению значимых результатов. В разработанном нами дискриминантно-регрессионном подходе (ЛДРП) для кластеризации молекул используется неформальная модель связи между структурой и свойствами молекул. Целью ЛДРП является выделение молекулярных кластеров, соответствующих заданной регрессионной модели «структура – свойство». В работе был использован ряд из 293 соединений с известными величинами активности (LC_{50} , ммоль/л, 96 час., лит. данные) и 4

типами МОА (неполярный наркоз, полярный наркоз, реактивный, особый) по отношению к рыбе *Pr*. Интервал изменения величин составлял для $\log(1/LC_{50})$: $-2.94 \div 5.74$, $\log K_{ow}$: $-1.87 \div 6.21$, α : $3.25 \div 39.10 \text{ \AA}^3$, $\sum C_a$: $0.00 \div 8.05$. Для реализации ЛДРП была создана компьютерная программа, в которой модель «структура – свойство» представлена в виде $\log(1/LC_{50}) = a + b \cdot \log K_{ow}$, где a , b – параметры. Основной задачей программы была генерация большого числа молекулярных кластеров путем выделения стартового кластера из двух молекул и циклического повторения 2-х основных шагов: построения для кластера регрессионной модели «структура – свойство» и добавления к этому кластеру ближайшей (по нормали) молекулы к линии регрессии.

Для анализа полученных кластеров был разработан ряд критериев, которые определялись для каждого кластера в отдельности. Первый критерий (МСС – корреляционный коэффициент Мэтьюза) характеризует степень обособленности рассматриваемого кластера молекул от всех остальных молекул. В качестве классификатора в настоящей работе использовался LDA. Второй критерий (М) связан с размером определенного паттерна, выявляемого при анализе формы зависимости МСС от размера кластера N . Еще одним критерием служила величина коэффициента корреляции (R), которая характеризовала степень линейности кластеров молекул.

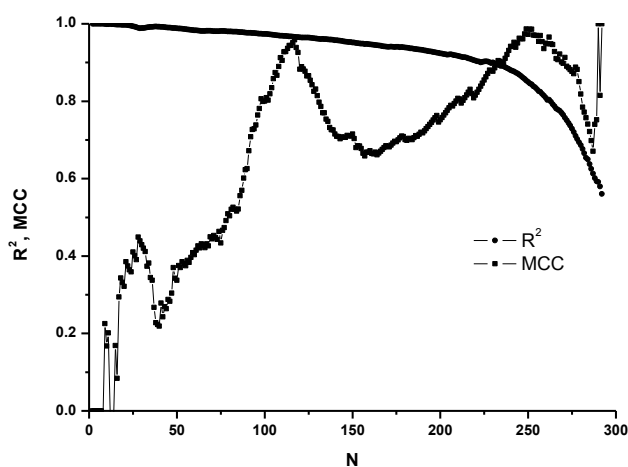


Рис. 3. Зависимость линейного коэффициента корреляции и корреляционного коэффициента Мэтьюза от числа молекул в кластере

На рис. 3 представлена типичная картина, наблюдаемая при использовании ЛДРП. При увеличении числа молекул в кластере R^2 плавно уменьшается от 1.000 (2 молекулы) до 0.560 (292 молекулы). МСС ведет себя более сложным образом. Минимальная величина для стартовой пары молекул, линейно неотделимых от остальных молекул, составляет 0.000, а при размере кластера 292 молекулы МСС=1.000. При этом при увеличении размера кластера наблюдается несколько кластерных пиков, размер которых и положение зависит от стартовой молекулярной пары. Для дальнейшего исследования полученных $293 \cdot 292 \cdot 291 / 2 = 12448398$ молекулярных кластеров был проведен их отбор с использованием следующих граничных значений критериев: $M \geq 25$, $R^2 \geq 0.9$, $MCC \geq 0.9$. В результате отбора было выявлено двенадцать кластеров. Минимальный размер кластера равен 105 молекулам, а максимальный – 123 молекулам. Относительно небольшой диапазон изменения размеров позволил выдвинуть предположение о близости составов полученных групп. Действительно, при анализе содержимого двенадцати кластеров была обнаружена большая группа молекул ($N=80$), которая входит в состав всех кластеров. В результате для выделенной группы из 80-ти молекул было установлено, что 53 молекулы относятся к группе с МОА неполярный наркоз, 19 – к полярному наркозу, 5 молекул являются реактивными и 3 – особыми. Т.е. основную массу молекул составляют молекулы с неспецифическим характером действия (неполярный наркоз + полярный наркоз). Их доля оказывается равной $100 \cdot (53+19)/80 = 90\%$. Учитывая это, мы сконструировали для этого кластера регрессионную модель зависимости токсичности от липофильности:

$$\log(1/LC_{50}) = -1.81(\pm 0.06) + 0.88(\pm 0.02) \log K_{ow} \quad (14)$$

$$n=80; r^2=0.966; s=0.23; F=2199; q^2=0.964$$

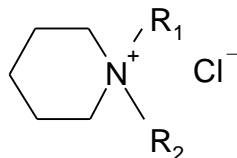
Параметры уравнения (14) очень хорошо соответствуют параметрам уравнения (15), построенного для ряда из 90 молекул с МОА неполярный наркоз:

$$\log(1/LC_{50}) = -1.79(\pm 0.06) + 0.88(\pm 0.02) \log K_{ow} \quad (15)$$

$$n=90; r^2=0.948; s=0.34; F=1596; q^2=0.945$$

4.4. Ретардантная активность солей четвертичного аммония

Для исследования ретардантной активности был использован антигипбереллиновый тест (IC_{50} , моль/л) для ряда из 12 производных холина и триэтилхолина: $(R_1)_3N^+CH_2CH_2-R_2 A^-$ где $R_1=CH_3, C_2H_5$; $R_2=OH, OCH_3, OC_2H_5, OCH_2C_6H_5, F, Cl, Br$; $A=F, Cl, Br$ и 12 хлоридов N,N-диалкилпиперидиниев:



где $R_1=CH_3, C_2H_5, C_3H_7, C_4H_9, CH_2C_6H_5$; $R_2=CH_3, C_4H_9, C_6H_5, CH_2CH_2OCH_3, CH_2CH_2OC_2H_5, CH_2CH_2OCH_2C_6H_5$. В результате QSAR моделирования для ряда диалкилпиперидиниев была получена следующая регрессионная модель:

$$\log(1/IC_{50}) = 14.93(\pm 1.82) + 0.14(\pm 0.04) \alpha - 0.83(\pm 0.19) C_a^{\max} + 1.52(\pm 0.30) E_s \quad (16)$$

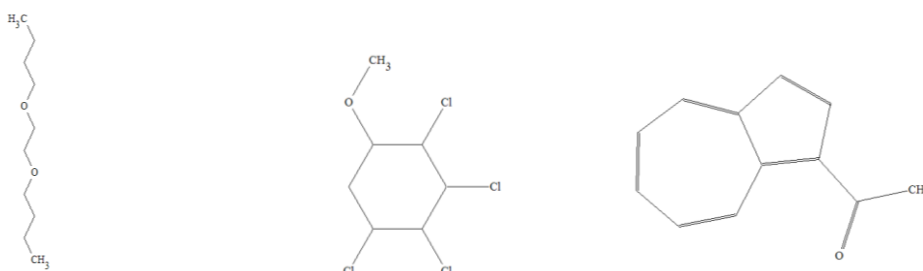
$$n=12; r^2=0.895; s=0.38; q^2=0.788$$

ОП модели (16): химический класс: хлориды N,N-диалкилпиперидиния; интервал $\log(1/IC_{50})$: $2.98 \div 5.95$; α : $14.58 \div 38.21 \text{ \AA}^3$; C_a^{\max} : $0.00 \div 1.49$; $-E_s$: $7.09 \div 10.25$.

Полученные данные свидетельствуют о том, что существенную роль в формировании антигипбереллиновой активности исследованных солей играют поляризуемость, Н-акцепторная способность молекул и стерический параметр. Определяющее влияние стерического фактора связано, по-видимому, с тем, что четвертичный атом азота встроен в пиперидиновое кольцо. Это ограничивает конформационные возможности молекулы при ее взаимодействии с активным центром рецептора и предъявляет повышенные требования к молекулярным свойствам, которые приносят фрагменты R_1 и R_2 .

Представленные в главе 4 локальные модели «структура – активность» могут быть использованы для предсказания биологической активности и для получения информации о структуре соединений с заданной величиной активности. Возьмем, например, модель (11). Оценка величины $\log(1/LC_{50})$ для соединения с известной структурой не представляет сложности: достаточно рассчитать величины α и $\sum C_a$ и подставить их в соответствующее уравнение. Например, для молекулы н-октанола ($\alpha=16.09 \text{ \AA}^3$, $\sum C_a=1.61$) расчетная

величина $\log(1/LC_{50})=0.77$ (эксперимент: 0.87). Решение обратной задачи представляет известные трудности из-за ее некорректности (число неизвестных в модели превышает число уравнений). Один из вариантов ее решения может состоять в следующем. Предположим, нас интересуют структуры с величиной $\log(1/LC_{50})=1.00$ и интервалом изменения α на уровне $20 \div 22 \text{ \AA}^3$. Подставим соответствующие значения в уравнение (11) и определим интервал для $\sum C_a$: $2.50 \div 3.09$. Далее, используя программу НУВОТ (см. раздел 3.4), найдем структуры, соответствующие интервалам. В результате обнаружим 103 структуры, некоторые из которых приведены ниже.



Необходимо отметить, что модель (11) создана с использованием всего 150 соединений, при этом в интервал α $20 \div 22 \text{ \AA}^3$ попадает только восемь из этих соединений. Решение обратной задачи с применением программы НУВОТ позволяет выйти из круга структур обучающей выборки и значительно расширяет возможности для синтеза соединений с заданной величиной активности.

Глава 5. Количественные глобальные модели «структура – биологическая активность» органических соединений

5.1. Модель АЛБТ

Обычно классификационная QSPR/QSAR модель строится для обучающей выборки в целом. В разработанном нами методе АЛБТ для классификации молекул используется большое количество локальных моделей. В случае АЛБТ процесс классификации сводится к следующим основным шагам: 1) задать теоретическую модель связи «структура–активность» (т.е. определить активность, значимые дескрипторы и вид математической связи между ними); 2) сформировать кластеры молекул в пространстве дескрипторов; 3) для каждого кластера

создать модель «структура – активность»; 4) на основе модели провести дискриминацию молекул.

Исследование было проведено на основе ряда из 9618 органических соединений с известной активностью (LD_{50} , ммоль/кг, внутривенно, 24 час., мыши, лит. данные). В качестве модели «структура – активность» использовалось уравнение вида $\log(1/LD_{50}) = a + b \cdot \alpha + c \sum C_a$, где a , b , c – параметры. В качестве сравнительных методов классификации были использованы такие методы, как линейный дискриминантный анализ (LDA), к-ближайший сосед (KNN), структурное сходство (SS), опорные вектора (SVM), нейронная сеть (ANN).

Таблица 5. Чувствительность (S_n), специфичность (S_p) и точность (Acc) при бинарной классификации химических соединений с MOA неполярный наркоз (класс 1) и другие MOA (класс 2) по отношению к мышам с использованием методов линейного дискриминантного анализа (LDA), к-ближайшего соседа (KNN), структурного сходства (SS), опорных векторов (SVM), нейронных сетей (ANN) и анализа локальной базовой токсичности (АЛБТ)

№	Метод	Обучающая выборка				Тестовая выборка			
		n_1/n_2	S_n	S_p	Acc	n_1/n_2	S_n	S_p	Acc
1	LDA	2811/4884	0.291	0.881	0.666	757/1166	0.284	0.890	0.652
2	KNN_7	2811/4884	0.605	0.807	0.733	757/1166	0.594	0.813	0.727
3	SS_5	2811/4884	0.587	0.798	0.721	757/1166	0.568	0.810	0.715
4	SVM	2811/4884	0.855	0.947	0.913	757/1166	0.567	0.852	0.739
5	ANN	2811/4884	0.406	0.853	0.690	757/1166	0.394	0.846	0.668
6	АЛБТ	1959/5715	0.992	1.000	0.998	509/1405	0.650	0.887	0.824

Примечание: n_1 – число соединений в первом классе, n_2 – число соединений во втором классе.

В табл. 5 представлены результаты бинарной классификации 9618 органических соединений с использованием различных методов в пространстве из HYBOT дескрипторов (LDA, KNN, SVM, ANN), фрагментных дескрипторов (SS), α и $\sum C_a$ дескрипторов (АЛБТ). Следует отметить, что среди использованных методов самые худшие результаты показывает метод LDA, а самые лучшие – АЛБТ. Так, в случае LDA, величина Acc, составляет 0.666 (обучающая выборка) и 0.652 (тестовая выборка). При использовании АЛБТ

Асс достигает 0.998 и 0.824 соответственно. Подходы, основанные на анализе ближайшего соседа (KNN_7 и SS_5), показывают промежуточные результаты.

ОП модели АЛБТ: химический класс: различные химические классы; интервал $\log(1/LD_{50})$: $-2.23 \div 6.73$; α : $3.25 \div 132 \text{ \AA}^3$; $\sum C_a$: $0.00 \div 30.60$.

5.2. Модель АМР

При создании глобальных QSPR/QSAR моделей обычной практикой является формирование молекулярных кластеров структурно родственных ближайших соседей на основе различных мер молекулярного сходства, например, коэффициентов Танимото (T_c) с дальнейшим использованием статистических методов для оценки активности соединений. В результате наших исследований был разработан АМР подход, в котором наряду с мерами структурного сходства используются меры физико-химического сходства, основанные на дескрипторах. Алгоритм модели АМР может быть описан в виде четырех шагов. 1. Для каждой анализируемой молекулы определялась величина структурного сходства с остальными молекулами. Величины T_c упорядочивались по убыванию. На основе задаваемого минимального граничного значения T_c формировались первичные кластеры ближайших структурных соседей. 2. Для всех молекул рассчитывался ряд дескрипторов. 3. С использованием дескрипторов и первичного структурного кластера для анализируемой молекулы рассчитывался вторичный кластер. Для этого из молекул первичного кластера выбирались такие пары ближайших соседей, которые имели значение дескриптора больше и меньше, чем у анализируемой молекулы. Размер вторичного кластера регулировался путем задания числа молекулярных пар (NP). 4. Активность анализируемой молекулы рассчитывалась как арифметическое среднее токсичностей молекулярных пар.

Для исследования был использован ряд из 10241 органического соединения с известной активностью (LD_{50} , ммоль/кг, внутривенно, 24 час., мыши, лит. данные). Дескрипторы рассчитывали с использованием программы НУВОТ. Наилучшие результаты были получены на основе дескриптора $\sum C_{ad}/\alpha$.

Путем варьирования значений T_c ($0.00 \div 1.00$) и NP ($1 \div 5$) был проведен расчет величин $\log(1/LD_{50})$. Качество полученных моделей оценивалось путем сравнения экспериментальных и рассчитанных величин токсичности с использованием уравнения:

$$\log(1/LD_{50})_{\text{эксп}} = a_0 + a_1 \log(1/LD_{50})_{\text{расч}} \quad (17)$$

Интервал изменения параметров и статистических характеристик полученных уравнений был следующим: a_0 : $-0.02 \div 0.15$; a_1 : $0.77 \div 1.10$; n : $491 \div 10228$; r^2 : $0.456 \div 0.783$; s : $0.33 \div 0.52$; F : $1567 \div 9672$; q^2 : $0.456 \div 0.773$. При возрастании границы структурного сходства и числа пар, свободный член уравнения (17) приближается к 0, а угловой коэффициент приближается к 1. В то же самое время, коэффициент линейной корреляции возрастает, а стандартное отклонение убывает. Однако, этот процесс сопровождается уменьшением числа соединений, вовлеченных в расчет.

Почти для всех химических соединений изучаемого ряда могут быть рассчитаны величины токсичности на основе модели AMP с использованием $T_c \geq 0.00$ и $NP=2$. Только 14 соединений из 10241 не имеют двух пар соседей с большим и меньшим значением выбранного дескриптора ($\sum C_{ad}/\alpha$). Параметры уравнения (17) в этом случае удовлетворительные и стандартное отклонение сравнимо с ожидаемыми экспериментальными ошибками измерения токсичности. Однако результаты расчета могут быть улучшены на основе каскадного расчета активности, в котором AMP модели появляются последовательно.

На самом высоком уровне А (табл. 6) граница структурного сходства $T_c \geq 0.80$ и число вовлеченных в расчет структурных пар $NP=2$. Этому уровню соответствует 491 соединение (4.79% от общего числа соединений). При этом токсичность этих соединений может быть оценена со стандартным отклонением $s=0.33$. Следующий уровень В включает в себя соединения с $T_c \geq 0.70$ и $NP=2$, исключая соединения уровня А. Для таких 832 соединений (8.12% от общего числа соединений) стандартное отклонение равно 0.39. Подобным образом были получены модели для последующих уровней с исключением соединений предыдущего уровня.

Использование каскадной модели AMP (уровни А-Е) позволило оценить острую токсичность 7085 соединений (около 69% от анализируемого ряда данных) по отношению к мышам со стандартным отклонением, сравнимым с

экспериментальной ошибкой определения токсичности. Для оставшихся 3142 соединений (уровень F) стандартное отклонение равно 0.64, что эквивалентно приблизительно одной трети интервала токсичности для 97% изученных соединений.

Таблица 6. Модель AMP: параметры и статистические характеристики уравнения $\log(1/LD_{50})_{\text{эксп}} = a_0 + a_1 \log(1/LD_{50})_{\text{расч}}$ при различных уровнях структурного сходства (T_c) и числе структурных пар $NP=2$

Уровень	T_c	a_0	a_1	n	r^2	s	F	q^2
A	≥ 0.80	0.03(± 0.02)	1.00(± 0.03)	491	0.762	0.33	1567	0.756
B	≥ 0.70	0.04(± 0.02)	0.96(± 0.02)	832	0.673	0.39	1708	0.671
C	≥ 0.60	0.01(± 0.02)	0.97(± 0.02)	1291	0.673	0.38	2654	0.671
D	≥ 0.50	0.04(± 0.01)	0.94(± 0.02)	2010	0.581	0.44	2785	0.580
E	≥ 0.40	0.10(± 0.01)	0.82(± 0.02)	2461	0.442	0.48	1947	0.440
F	≥ 0.00	0.16(± 0.01)	0.83(± 0.02)	3142	0.337	0.64	1593	0.335

Статистические характеристики обучающего ряда (8193 соединений) и тестового ряда (2048 соединений) близки к характеристикам общей модели. В частности, величины s составили $0.26 \div 0.35$, 0.40, $0.37 \div 0.40$, $0.43 \div 0.48$, $0.46 \div 0.52$, $0.60 \div 0.65$ для уровней A, B, C, D E, F соответственно.

ОП модели AMP: химический класс: различные химические классы; интервал $\log(1/LD_{50})$: $-3.14 \div 4.35$; T_c : $0.00 \div 1.00$; $\sum C_{ad}/\alpha$: $0.00 \div 1.35$.

5.3. Модель ЛРМПК

Модель локальных регрессионных моделей в перекрывающихся кластерах (ЛРМПК) стала результатом дальнейшего развития концепции дискретно-непрерывных QSAR моделей. Для ее тестирования был использован тот же ряд из 10241 соединения, дескрипторы и мера структурного сходства, что и в модели AMP. Алгоритм ЛРМПК может быть представлен следующим образом.

1. Для рассматриваемого (i -го) соединения из выборки, содержащей n соединений, рассчитываются меры структурного сходства (T_c) с остальными $n-1$ соединениями. Полученные величины ранжируются по убыванию. Процедура повторяется для каждого соединения. В результате получается n отдельных рядов соединений.
2. Фиксируется минимальный порог структурного сходства T_c , и в каждом из n рядов отбираются структурно-родственные соединения, имеющие T_c выше фиксированного уровня. Таким

образом, формируются n перекрывающихся кластеров. 3. Для всех анализируемых соединений рассчитываются дескрипторы. 4. Для каждого кластера производится отбор дескрипторов на основе парных коэффициентов линейной корреляции и генерируются отдельные локальные регрессионные QSAR модели с использованием полного перебора всех возможных комбинаций из одного ($M=1$), двух ($M=2$) или трех ($M=3$) дескрипторов. Выбирается лучшее уравнение на основе критерия Фишера (F). В работе использовались линейные регрессионные модели при наличии в кластере ($5M + 1$) и более структурных соседей. Это обеспечивает наличие минимум 5 соединений на один дескриптор в процессе перекрестного контроля с исключением по одному. 5. Полученные QSAR уравнения используются для расчета активности рассматриваемых соединений.

Из данных табл. 7 следует, что при задании в качестве граничного значения $T_c \geq 0.3$ и с учетом необходимого минимального числа структурных соседей, было получено 7759 кластеров, для которых было сконструировано 6425 однопараметровых, 1176 двухпараметровых и 158 трехпараметровых регрессионных уравнений для расчета токсичности выбранных соединений. Для оставшихся 2482 соединений не было обнаружено достаточного числа структурных соседей, необходимого для оценки токсичности с помощью описанного метода. В полученном уравнении свободный член и угловой коэффициент оказались близкими к идеальным (0.01 ± 0.01 и 0.99 ± 0.01 соответственно), а стандартное отклонение (0.51) - на уровне ошибки экспериментального измерения этого свойства для мышей разного возраста, пола с учетом условий их содержания и методики измерений.

Таблица 7. Модель ЛРМПК: параметры, статистические характеристики уравнения $\log(1/LD_{50})_{\text{эксп}} = a_0 + a_1 \log(1/LD_{50})_{\text{расч}}$ при различных уровнях структурного сходства (T_c) и распределение числа моделей при использовании одного ($M=1$), двух ($M=2$) и трех ($M=3$) дескрипторов

№	T_c	a_0	a_1	n	r^2	s	F	q^2	$M=1$	$M=2$	$M=3$
1	≥ 0.3	$0.01(\pm 0.01)$	$0.99(\pm 0.01)$	7759	0.434	0.51	5958	0.434	6425	1176	158
2	≥ 0.4	$0.06(\pm 0.01)$	$0.91(\pm 0.01)$	5011	0.530	0.46	5646	0.529	4074	810	127
3	≥ 0.5	$0.06(\pm 0.01)$	$0.90(\pm 0.01)$	3157	0.621	0.40	5170	0.620	2514	540	103
4	≥ 0.6	$0.09(\pm 0.01)$	$0.87(\pm 0.02)$	1748	0.621	0.40	2862	0.618	1457	260	31
5	≥ 0.7	$0.15(\pm 0.02)$	$0.76(\pm 0.02)$	924	0.618	0.38	1490	0.613	814	99	11
6	≥ 0.8	$0.10(\pm 0.03)$	$0.88(\pm 0.03)$	346	0.659	0.37	666	0.643	272	74	0

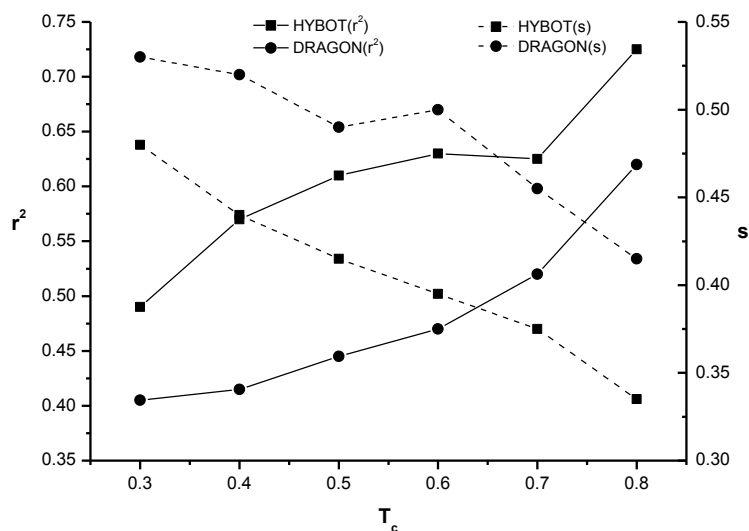


Рис. 4. Зависимость значений коэффициента корреляции (r^2) и стандартного отклонения (s) моделей ЛРМПК от величины порога структурного сходства (T_c) при использовании HYBOT и DRAGON дескрипторов

Для сравнения были проведены расчеты токсичности на основе дескрипторов, полученных с помощью компьютерной программы DRAGON. Очевидно, что порог структурного сходства имеет существенное влияние на точность расчетов (рис. 4). Статистические критерии 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. Также очевидно, что лучшие результаты расчетов получаются при использовании дескрипторов программы HYBOT.

Валидация модели ЛРМПК выполнена с использованием обучающей и тестовой выборок (см. модель AMP). При этом были получены близкие статистические характеристики, что свидетельствует об устойчивости анализируемой модели.

ОП модели ЛРМПК: химический класс: различные химические классы; интервал $\log(1/LD_{50})$: $-3.14 \div 4.35$; T_c : $0.00 \div 1.00$.

5.4. Проницаемость ФАВ через гематоэнцефалический барьер

Важным механизмом, обеспечивающим транспорт ФАВ через ГЭБ, является пассивная диффузия. Для количественного моделирования этого механизма была использована модель AMP (см. раздел 5.2). В качестве меры биологической активности были использованы литературные данные о $\log BB$ (десятичный логарифм отношения концентраций вещества в мозге и крови при пассивной диффузии через ГЭБ) для 298 соединений. Этот ряд разбивался на обучающую (239 соединений) и тестовую (59 соединений) выборки. В качестве внешней тестовой выборки использовались также данные по $\log BB$ для пяти дополнительных соединений, не включенных в общую выборку. Для описания структуры ФАВ использовали дескрипторы, рассчитанные на основе программы NYBOT, а также молекулярные фрагменты и меры структурного сходства, рассчитанные с помощью программы MOLDIVS.

В результате компьютерного моделирования было выделено 3 лучших модели на основе одиночных дескрипторов $\sum C_{ad}$, $\sum C_{ad}/\alpha$, и $\sum E_{ad}$. При использовании обучающей выборки из 298 соединений для вышеперечисленных моделей были получены величины r^2 и s равные 0.594 и 0.41, 0.588 и 0.41, 0.577 и 0.42 соответственно. На финальной стадии работы была получена согласованная модель с использованием в качестве конечного результата среднего арифметического рассчитанных значений из 3-х моделей по отдельным дескрипторам. Применение такого подхода привело к получению модели со следующими статистическими характеристиками: $r^2=0.641$ и $s=0.38$. Последняя цифра близка к значению s равному 0.37, полученному при корреляции экспериментальных значений $\log BB$ одних тех же соединений, взятых из разных источников. Эта величина хорошо согласуется с литературными данными, согласно которым граница варибельности величин $\log BB$, полученных из разных источников, составляет $0.3 \div 0.5$ логарифмических

единиц. Это свидетельствует о том, что полученная финальная модель имеет стандартное отклонение на уровне экспериментальной ошибки $\log BB$.

На заключительном этапе для тестирования модели АМР были привлечены данные по проницаемости пяти соединений, которые не вошли в обучающую выборку (табл. 8). Это три вещества с пассивным механизмом транспорта (лоперамид, хлорпромазин, клонидин) и два вещества с активным механизмом транспорта (терфенадин, ранитидин). Из представленных данных следует, что не только соединения с пассивным механизмом транспорта, но и соединения, имеющие существенный вклад активного транспорта, хорошо описываются моделью АМР. Это позволяет предположить, что разработанный метод расчета может быть применен к химическим соединениям, проникающим через ГЭБ с помощью различных механизмов.

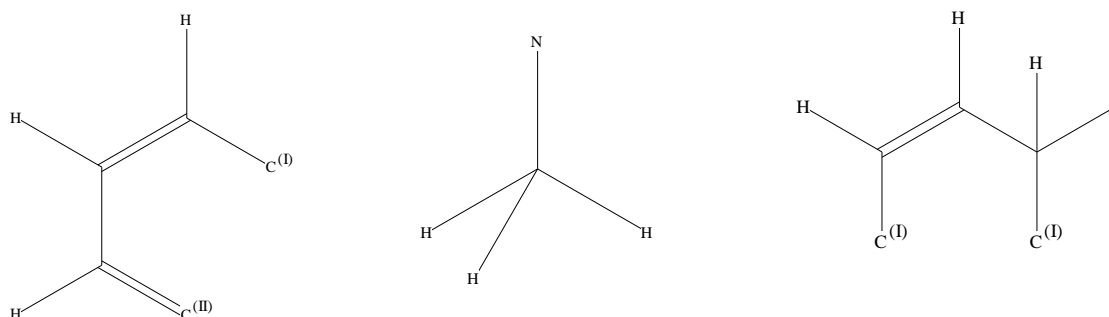
Таблица 8. Экспериментальные ($\log BB_{\text{эксп}}$) и рассчитанные ($\log BB_{\text{расч}}$) величины проницаемости ФАВ через ГЭБ на основе АМР моделей с дескрипторами: $\sum E_{\text{ад}}$ – сумма энтальпийных Н-акцепторных и Н-донорных дескрипторов, $\sum C_{\text{ад}}$ – сумма свободноэнергетических Н-акцепторных и Н-донорных дескрипторов, α – молекулярная поляризуемость (Å^3).

№	ФАВ	$\log BB_{\text{эксп}}$	$\log BB_{\text{расч}}$			
			$\sum E_{\text{ад}}$	$\sum C_{\text{ад}}$	$\sum C_{\text{ад}}/\alpha$	$\sum E_{\text{ад}} + \sum C_{\text{ад}} + \sum C_{\text{ад}}/\alpha$
1	Лоперамид	0.77	0.45	0.45	0.94	0.61
2	Хлорпромазин	0.84	0.70	1.02	1.02	0.91
3	Клонидин	0.19	-0.14	0.25	0.25	0.12
4	Терфенадин	0.64	0.45	0.45	0.70	0.53
5	Ранитидин	-1.23	-1.27	-1.03	-1.03	-1.11

ОП модели АМР: химический класс: различные химические классы, полифункциональные соединения; интервал $\log BB$: $-2.00 \div 1.44$; T_c : $0.00 \div 1.00$; $\sum E_{\text{ад}}$: $0.00 \div 25.80$; $\sum C_{\text{ад}}$: $0.00 \div 28.85$; $\sum C_{\text{ад}}/\alpha$: $0.00 \div 1.13$.

На основе глобальных моделей «структура – активность», представленных в главе 5, может проводиться компьютерный дизайн органических соединений путем решения прямой и обратной задачи. Пример успешного решения прямой задачи на основе модели АМР представлен в табл.

8. Обратная задача (предсказание структур с заданными свойствами) из-за своей некорректности может быть решена путем введения дополнительных условий. Введем, например, следующие ограничения: $\log BB: 1.0 \div 1.5$, $T_c \geq 0.5$; $NP=1$; дескриптор: $\sum C_{ad}$. В результате в обучающей выборке найдем 2 соответствующих соединения, для которых проведем анализ молекулярных фрагментов. В результате был установлен набор из 30 молекулярных фрагментов, 19 из которых являются общими. Некоторые из этих фрагментов изображены ниже:



При изменении ограничений модели АМР набор общих молекулярных фрагментов также меняется.

Глава 6. Перспективные разработки в области создания количественных моделей «структура – свойство» и «структура – активность»

6.1. Рекуррентная модель

Для анализа количественных соотношений «структура – активность» были использованы рекуррентные уравнения вида:

$$T_n = A + BT_{n-1} \quad (18)$$

где T_n – активность текущего n-го члена гомологического ряда, T_{n-1} – активность предшествующего члена гомологического ряда, A и B – параметры. В качестве меры активности использовались литературные данные по LC_{50} и LD_{50} 85 органических соединений из 7 гомологических рядов по отношению к 6 организмам: бактерии *Vibrio fisheri*, инфузории *Tetrahymena pyriformis*, рыбам *Poecilia reticulata* и *Pimephales promelas*, мышам и крысам. Полученные уравнения характеризуются высокими статистическими показателями.

Например, в большинстве случаев, $r^2 > 0.9$. Часть представленных зависимостей имеет предельное значение активности ($B < 1$), другая часть не имеет предела ($B > 1$). При этом коэффициент A варьирует от 0.13 до 0.67.

Полезность математических моделей для описания активности можно связать с качеством экстраполяции данных. С этой целью мы сопоставили три модели: рекуррентную, линейную и параболическую. В двух последних моделях в качестве независимой переменной фигурировало число метиленовых групп в гомологическом ряду. Для сравнения были выбраны данные по активности с учетом количества молекул в ряду и с максимальным представлением всех использованных типов химических соединений и биологических объектов. В результате было отобрано 5 комбинаций гомологический ряд/биообъект. Устойчивость и предсказательная способность моделей была оценена с помощью кросс-валидации и расчетом стандартных отклонений s . При этом было установлено, что рекуррентная модель имела минимальную величину $s = 0.33$.

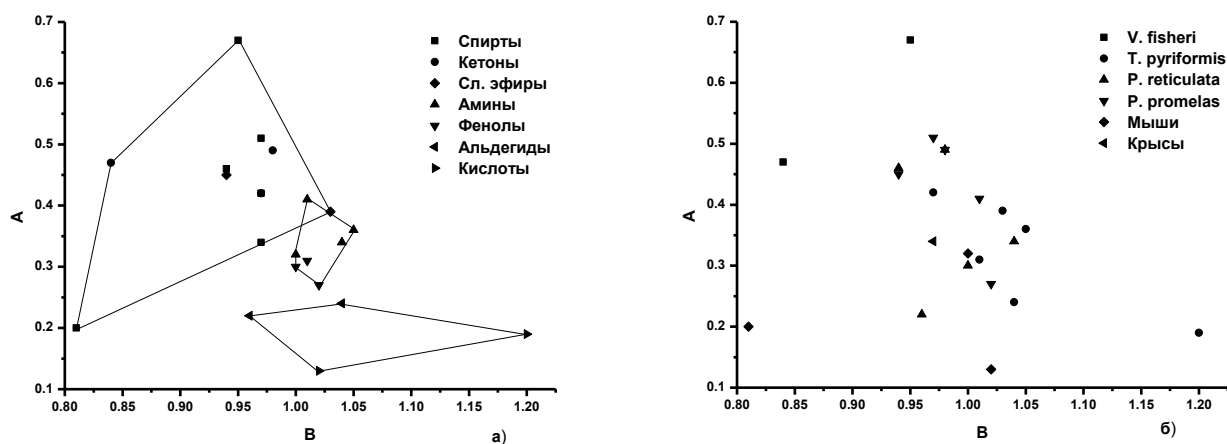


Рис. 5. Параметры рекуррентных уравнений на плоскости А-В с учетом типа гомологического ряда (а) и вида биологического объекта (б)

Рекуррентные модели могут быть выражены в виде точек на плоскости А-В с учетом типа гомологического ряда и вида биологического объекта. На рис. 5а отчетливо видны три кластера. В первый кластер входят спирты, кетоны и

сложные эфиры (МОА: неполярный наркоз). Во второй кластер – амины и фенолы (МОА: полярный наркоз). Третий кластер состоит из альдегидов и кислот (другие МОА). Таким образом, представленный метод графического анализа позволяет различать гомологические ряды между собой, относя их к различным типам действия.

Расстояние между точками (рис. 5б) было использовано для анализа близости биообъектов между собой. Обращает на себя внимание уникальность данных для бактерии *V. fisheri* ввиду отсутствия в ближайших окрестностях аналогичных моделей активности. Следует отметить близость параметров рекуррентных уравнений, описывающих активность химических соединений по отношению к рыбам *P. reticulata* и *P. promelas*. Из представленных данных также следует сходство моделей активности для водных организмов *T. pyriformis* и *P. reticulata*, а также *T. pyriformis* и *P. promelas*.

ОП рекуррентной модели: химический класс: спирты, кетоны, эфиры, амины, фенолы, альдегиды, кислоты; интервал $\log(1/LC_{50})$: $-2.96 \div 3.69$; $\log(1/LD_{50})$: $-2.17 \div 1.09$.

6.2. Фрактальный дескриптор

Для описания геометрической структуры соединений были использованы гистограммы (дискретные функции распределения) межатомных расстояний, которые представлялись в виде плоских диаграмм, состоящих из набора закрашенных прямоугольников (баров). Ось абсцисс разбивалась с определенным шагом (разрешением) ΔR на интервалы, количество которых зависело от значений межатомных расстояний R , меняющихся от минимального значения R_{\min} до максимального значения R_{\max} . На оси ординат откладывалась частота f , т.е. количество межатомных расстояний, попадающих в заданный интервал, и варьирующая в виде целых чисел от 0 до f_{\max} . Эта ось также разбивалась на интервалы с определенным шагом Δf . Полученные гистограммы оцифровывались путем их представления в виде квадратных бинарных матриц, где значение 1 соответствовало закрашенным участкам гистограммы, а 0 – всему остальному.

Фрактальная размерность определялась «клеточным» методом с помощью компьютерной программы, путем покрытия оцифрованного изображения клетками переменной величины и подсчетом числа клеток, которые пересекают закрашенные области гистограмм. Локальная фрактальная размерность рассчитывалась по уравнению:

$$\log N(L) = \log C - D \log L \quad (19)$$

где: $N(L)$ – минимальное число клеток со стороной L , необходимых для покрытия фрактала; D – «клеточная» фрактальная размерность (со знаком -); C – константа. Пример расчета представлен на рис. 6. Подобным образом были проведены расчеты для 73 органических молекул, относящихся к 6 различным классам: алканы, спирты, эфиры, кетоны, амины, ароматические соединения. При этом число атомов N_a меняется от 5 до 50, а максимальная длина межатомных расстояний R_{\max} варьирует в пределах от 1.82 Å до 20.57 Å. Минимальная величина фрактальной размерности D составляет 1.00 (метан), а максимальная – 1.58 (1,2:5,6-дибензантрацен).

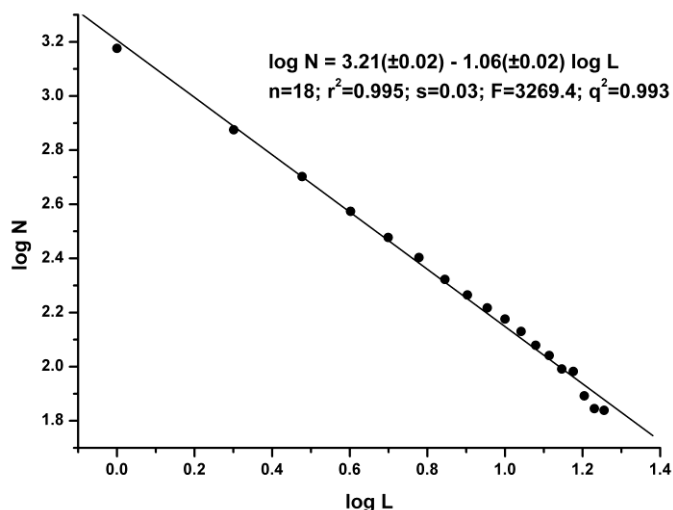


Рис. 6. Зависимость числа клеток (N) от их размера (L) (уравнение (19)) для молекулы метанола

Анализ полученных данных свидетельствует о существовании тенденции между ростом числа атомов в молекуле и фрактальной размерностью. Однако, наблюдающиеся значительные отклонения от тренда, говорят о том, что число

атомов не является единственным фактором. Например, четыре изомерных спирта с общей брутто-формулой $C_4H_{10}O$: 1-бутанол(I), 2-метил-1-пропанол(II), 2-бутанол(III) и 2-метил-2-пропанол(IV) имеют по 15 атомов, но различаются по фрактальной размерности: D при этом меняется от 1.20 до 1.38. Наблюдающееся явление может быть объяснено, если принять во внимание точечные группы симметрии молекул. Действительно, соединения (I) и (IV) относятся к группе C_s и имеют $D=1.25$ и 1.20 соответственно. Напротив, соединения (II) и (III) принадлежат к группе C_1 и их фрактальные размерности равны 1.33 и 1.38 соответственно. Количественной мерой, отражающей количество атомов в молекуле и принадлежность молекул к той или иной точечной группе симметрии, в первом приближении, может являться число баров в гистограммах межатомных расстояний N_b . С учетом этого для соединений с различными точечными группами симметрии были сконструированы следующие модели:

$$C_{2v}: D = 0.97(\pm 0.03) + 0.81(\pm 0.06) \log(N_b/N_a) \quad (20)$$

$$n=14; r^2=0.935; s=0.04; F=172; q^2=0.889$$

$$C_s: D = 0.91(\pm 0.03) + 0.70(\pm 0.06) \log(N_b/N_a) \quad (21)$$

$$n=23; r^2=0.851; s=0.04; F=120; q^2=0.810$$

$$C_1: D = 0.91(\pm 0.06) + 0.63(\pm 0.08) \log(N_b/N_a) \quad (22)$$

$$n=18; r^2=0.785; s=0.03; F=58.5; q^2=0.682$$

Уравнения (20)-(22) свидетельствуют о существовании корреляционной связи между фрактальной размерностью и нормированным числом баров. Эта связь становится понятной, если проанализировать физический смысл величины N_b/N_a . По сути, она представляет собой число баров, которое приходится на один атом, т.е. характеризует плотность заполнения пространства гистограммы. Поэтому фрактальная размерность, являясь количественной мерой степени изрезанности или сложности пространственной структуры молекул (в пределе, крайняя изрезанность – это полностью заполненное пространство), оказывается в статистической связи с величиной N_b/N_a .

Полезность нового предлагаемого дескриптора может быть продемонстрирована на примере установления количественных соотношений

между фрактальной размерностью и плотностью для гомологического ряда алканов (пентан, гексан,...эйкозан, пентакозан, триаконтан), которые при нормальных условиях представляют собой жидкости или твердые тела:

$$d_4^{20} \text{ (г/см}^3\text{)} = -0.21(\pm 0.05) + 0.70(\pm 0.03) D \quad (23)$$

$$n=18; r^2=0.965; s=0.01; F=442; q^2=0.956$$

Полученное регрессионное уравнение свидетельствует о существовании тесной корреляционной связи между соответствующими микро- и макромолекулярными характеристиками. В изученном ряду соединений возрастание сложности молекулярной структуры сопровождается увеличением плотности вещества.

6.3. Линейная динамическая модель

Для описания структуры химических соединений была использована предложенная в литературе функция радиального распределения в приближении жесткой модели молекулы (отсутствие внутримолекулярных колебаний):

$$f(R) = \frac{2}{\pi} \int_0^{s_{\max}} s M(s) \exp(-bs^2) \sin sR ds \quad (24)$$

здесь $f(R)$ - функция радиального распределения, R - межатомное расстояние, $M(s)$ - функция молекулярного рассеяния, s - угловой параметр.

В качестве моделируемого свойства выступала обратная функция давления насыщенного пара от температуры $T=f(P)$. Количественную связь между исследуемыми кривыми устанавливали на основе линейной динамической модели с одним входом, одним выходом и импульсной переходной функцией. Эта модель описана в литературе и предназначена для анализа случайных стационарных процессов. Непосредственно для расчетов использовалось уравнение:

$$Y(f) = H(f) X(f) \quad (25)$$

здесь $X(f)$ и $Y(f)$ – преобразования Фурье для входной и выходной кривой соответственно, а $H(f)$ - частотная характеристика. Частотную характеристику вычисляли путем усреднения по всем кривым на основе соотношения для взаимного спектра входного и выходного процессов:

$$S_{xy}(f) = H(f) S_{xx}(f) \quad (26)$$

здесь $S_{xy}(f)$ – взаимная двусторонняя спектральная плотность, $S_{xx}(f)$ – двусторонняя спектральная плотность входной кривой.

Качество исследуемой модели оценивали с помощью функции когерентности, изменяющейся в пределах от 0 до 1 и вычисляемой по формуле:

$$\gamma_{xy}^2(f) = \frac{|S_{xy}(f)|^2}{S_{xx}(f)S_{yy}(f)} \quad (27)$$

где $S_{yy}(f)$ – двусторонняя спектральная плотность выходной кривой. В дальнейшем для сравнения различных моделей использовали среднее значение γ^2 , полученное путем усреднения функции когерентности по всем частотам. Расчеты выполняли с использованием модифицированной компьютерной программы на основе быстрого преобразования Фурье. В качестве входных и/или выходных процессов выступали $f(R)$ и обратная функция давления насыщенного пара от температуры $T=f(P)$. Оцифровку зависимостей проводили путем дискретизации независимых переменных с использованием 16, 32, 64, 128, 256 точек (уровни дискретизации m) и вычислением соответствующих значений функций.

Учитывая разный физический смысл входных и выходных кривых, для получения статистически значимых корреляционных зависимостей использовали модификацию линейной динамической модели. Для этого на вход системы подавали участки кривых различной длины. При решении прямой задачи (модель «структура–активность») проводили варьирование $f(R)$ в качестве входной кривой путем деления интервала R (0 – 16 Å) на 160 частей и парной комбинации полученных значений R_{\min} и R_{\max} . При этом длина выходной кривой $T=f(P)$ оставалась неизменной и была максимальной ($P_{\min}=10$ мм Hg, $P_{\max}=1500$ мм Hg). Аналогичный подход использовали для входной кривой $T=f(P)$ при решении обратной задачи (модель «активность–структура»). В этом случае интервал P (10 – 1500 мм Hg) делили на 50 частей и комбинировали величины P_{\min} и P_{\max} . Длина кривой $f(R)$ была постоянной ($R_{\min}=0$ Å, $R_{\max}=16$ Å). Полученные модели сортировали по величине

стандартного отклонения и отбирали те из них, которые имели минимальное значение. Для исследования был использован ряд из 10 н-алканов (пропан, бутан, ... додекан).

В табл. 9 приведены характеристики лучших моделей типа «структура – активность», полученных путем перебора 12880 реализаций различной длины на каждом уровне дискретизации. Учитывая, что модель 1 имеет минимальную величину $s=5.0$ и максимальное значение $\gamma^2=0.988$, она была выбрана в качестве финальной модели типа «структура-активность». С использованием этой модели были рассчитаны функции $T=f(P)$ для 10 исследуемых соединений:

$$T_{\text{расч}} = -3.5(\pm 1.7) + 1.01(\pm 0.01) T_{\text{эксп}} \quad (28)$$

$$n=160; r^2=0.997; s=5.0; q^2=0.996$$

Уравнение (28) свидетельствует об адекватности модели 1.

Таблица 9. Порядковый номер модели (№), уровни дискретизации (m), минимальная ($R_{\text{мин}}$) и максимальная ($R_{\text{макс}}$) величина межатомного расстояния, стандартное отклонение (s) и среднее значение функции когерентности (γ^2) для моделей «структура – активность»

№	m	$R_{\text{мин}}, \text{Å}$	$R_{\text{макс}}, \text{Å}$	s	γ^2
1	16	1.1	13.9	5.0	0.988
2	32	0.8	12.7	5.4	0.944
3	64	1.4	15.3	5.2	0.976
4	128	0.0	15.5	6.6	0.956
5	256	0.0	15.6	6.6	0.933

В рамках рассматриваемой динамической модели также может решаться «обратная» задача QSPR/QSAR, т.е. предсказание структур с заданными свойствами. С этой целью была сконструирована специальная модель «активность - структура», где в качестве входа используется зависимость $T=f(P)$ переменной длины, а в качестве выхода – полноразмерная $f(R)$. Учитывая, что точность описания функции радиального распределения по оси ординат (интенсивность) на уровне величины 0.06 представляется приемлемой, а максимальное разрешение по оси абсцисс (межатомное расстояние) достигается при $m=256$ и составляет 0.06 Å , в качестве рабочей модели «активность – структура» была использована модель 10 (табл. 10).

Таблица 10. Порядковый номер модели (№), уровни дискретизации (m), минимальная ($P_{\text{мин}}$) и максимальная ($P_{\text{макс}}$) величина давления насыщенного пара, стандартное отклонение (s) и среднее значение функции когерентности (γ^2) для моделей «активность - структура»

№	m	$P_{\text{мин}}$, ММ Hg	$P_{\text{макс}}$, ММ Hg	s	γ^2
6	16	1470.2	1500.0	0.018	0.992
7	32	10.0	39.8	0.047	0.893
8	64	1470.2	1500.0	0.072	0.919
9	128	1470.2	1500.0	0.059	0.943
10	256	1470.2	1500.0	0.061	0.940

С использованием этой модели было получено следующее уравнение:

$$f(R)_{\text{расч}} = -0.001(\pm 0.001) + 1.00(\pm 0.01) f(R)_{\text{эксп}} \quad (29)$$

$$n=2560; r^2=0.971; s=0.061; q^2=0.966$$

Параметры и коэффициенты регрессионного уравнения (29) свидетельствуют о высоком качестве модели 10. Это демонстрирует и сопоставление рассчитанных и экспериментальных значений функции радиального распределения для молекулы гептана (рис. 7) с хорошим соответствием как по абсциссе, так и по ординате. Таким образом, задача предсказания структур с заданными свойствами может быть решена путем конструирования специальной модели «активность – структура», расчета на ее основе $f(R)$ для молекулы с заданными свойствами и сравнения с $f(R)$ модельного соединения с известной 3D структурой.

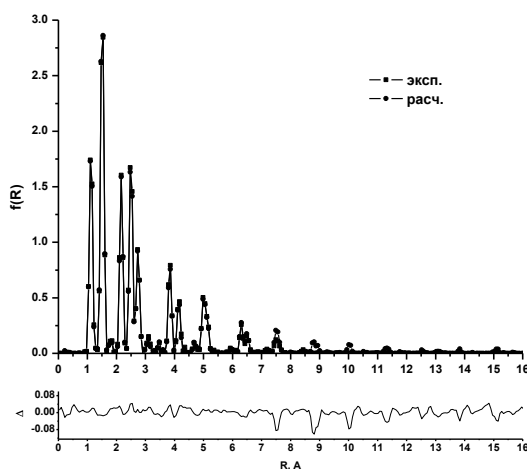


Рис. 7. Экспериментальные и рассчитанные по модели 10 значения функции радиального распределения и их разность (Δ) для молекулы гептана при уровне дискретизации $m=256$

ВЫВОДЫ

1. Разработана аддитивно-мультипликативная модель для описания энтальпии и свободной энергии Н-комплексов. На основе этой модели создана единая энергетическая шкала водородной связи и рассчитаны дескрипторы, количественно характеризующие способность соединений к образованию Н-связи.
2. Разработана количественная физико-химическая модель для описания липофильности органических соединений с использованием поляризуемости и протоноакцепторных дескрипторов Н-связи.
3. На основе стандартных и оригинальных подходов с применением дескрипторов водородной связи сконструированы и протестированы модели «структура–активность» для компьютерного дизайна органических соединений, включая оценку острой токсичности, ретардантной активности и способности проникновения через гематоэнцефалический барьер неисследованных новых соединений и предсказание структур с заданным уровнем активности.
4. Разработана рекуррентная модель активности в гомологических рядах соединений. Выявлено, что с помощью этой модели может быть установлен тип действия ряда и близость биообъектов между собой.
5. Для описания структуры химических соединений создан и применен новый дескриптор: фрактальная размерность. Установлено, что предлагаемый дескриптор зависит от числа атомов и точечной группы симметрии молекул и может быть использован при исследовании связи «структура – свойство (активность)».
6. Разработан новый подход для установления количественной связи между структурой и свойствами (активностью) химических соединений на основе линейной динамической модели, который может быть использован в качестве рабочего инструмента при проведении QSAR/QSPR исследований.
7. На основе теоретических разработок получен ряд практических результатов: создана БД по термодинамическим характеристикам Н-комплексов; сформирована БД дескрипторов Н-связи.

Основные публикации по теме диссертации**Статьи в отечественных и зарубежных журналах**

1. Раевский О.А., Григорьев В.Ю., Соловьев В.П. Моделирование связи структура – активность. II. Оценка электронодонорных и акцепторных функций активных центров в молекулах физиологически активных веществ // Хим.-фармац. журн. 1989. Т. 23. № 11. С. 1294-1300.
2. Григорьев В.Ю., Соловьев В.П., Додонов М.В., Брель В.К., Мартынов И.В., Раевский О.А. Электронодонорные и акцепторные функции физиологически активных и модельных соединений. XIV. Электроноакцепторная способность группы N-H дифторнитроацетанилидов // Журн. общ. химии. 1989. Т. 59. № 4. С. 846-851.
3. Григорьев В.Ю., Соловьев В.П., Раевский О.А., Иванов А.Н., Соколов В.Б., Мартынов И.В. Электронодонорные и акцепторные функции физиологически активных и модельных соединений. XV. Использование мультипликативного правила факторов для оценки термодинамических характеристик водородной связи дифторнитроацетанилидов с фосфорилированными оксимами // Журн. общ. химии. 1990. Т. 60. № 3. С. 527-530.
4. Гафуров Р.Г., Чистяков В.Г., Кокурин А.В., Григорьев В.Ю., Прошин А.Н., Раевский О.А., Муромцев Г.С., Мартынов И.В. Связь химической структуры и активности регуляторов роста растений. Сообщение 1. Влияние стерического параметра катиона на антигиббереллиновую активность хлоридов диалкилпиперидиниев // Агрехимия. 1991. № 12. С. 81-85.
5. Raevsky O.A., Grigor'ev V.Yu., Kireev D.B., Zefirov N.S. Correlation Analysis and H-bond Ability in Framework of QSAR // J. Chim. Phys. Phys.-Chim. Biol. 1992. Vol. 89. № 7-8. P. 1747-1753.
6. Raevsky O.A., Grigor'ev V.Yu., Kireev D.B., Zefirov N.S. Complete Thermodynamic Description of H-Bonding in the Framework of Multiplicative Approach // Quant. Struct.-Act. Relat. 1992. Vol. 11. № 1. P. 49-63.

7. Раевский О.А., Григорьев В.Ю. Количественное описание липофильности органических соединений на основе поляризуемости и акцепторной способности к образованию водородной связи // Хим.-фармац. журн. 1999. Т. 33. № 5. С. 46-49.
8. Гафуров Р.Г., Григорьев В.Ю., Прошин А.Н., Чистяков В.Г., Мартынов И.В., Зефиоров Н.С. Роль молекулярных параметров солей четвертичного аммония в формировании их антигипбереллиновой (ретардантной) активности // Докл. РАН. 2004. Т. 394. № 5. С. 710-714.
9. Gafurov R.G., Grigor'ev V.Yu., Proshin A.N., Chistyakov V.G., Martynov I.V., Zefirov N.S. The Importance of Molecular Parameters of Quaternary Ammonium Salts in Their Antigibberelin (Retardant) Activity // Russ. J. Bioorg. Chem. 2004. Vol. 30. № 6. P. 592-598.
10. Raevsky O.A., Grigor'ev V.Yu., Raevskaja O.E., Schaper K.-J. Physicochemical Properties/Descriptors Governing the Solubility and Partitioning of Chemicals in Water-Solvent-Gas systems. Part.1. Partitioning between Octanol and Air // SAR QSAR Environ. Res. 2006. Vol. 17. № 3. P. 285-297.
11. Raevsky O.A., Grigor'ev V.Yu., Weber E.E., Dearden J.C. Classification and Quantification of the Toxicity of Chemicals to Guppy, Fathead Minnow and Rainbow Trout. Part 1. Nonpolar Narcosis Mode of Action // QSAR Comb. Sci. 2008. Vol. 27. № 11-12. P. 1274 – 1281.
12. Raevsky O.A., Grigor'ev V.Yu., Dearden J.C., Weber E.E. Classification and Quantification of the Toxicity of Chemicals to Guppy, Fathead Minnow, and Rainbow Trout. Part 2. Polar Narcosis Mode of Action // QSAR Comb. Sci. 2009. Vol. 28. № 2. P. 163-174.
13. Раевский О.А., Григорьев В.Ю., Тихонова О.В. Развитие моделей взаимосвязи структуры и токсического действия химических соединений по отношению к Guppy // Хим.-фармац. журн. 2009. Т. 43. № 3. С. 3-7.
14. Raevsky O.A., Grigor'ev V.Yu., Modina E.A., Worth A.P. Prediction of Acute Toxicity to Mice by the Arithmetic Mean Toxicity (AMT) Modelling Approach // SAR QSAR Environ. Res. 2010. Vol. 21. № 3-4. P. 265-275.

15. Raevsky O.A., Grigor'ev V.Y., Liplavskaya E.A., Worth A.P. Prediction of Acute Rodent Toxicity on the Basis of Chemical Structure and Physicochemical Similarity // Mol. Inf. 2011. Vol. 30. № 2-3. P. 267-275.
16. Григорьев В.Ю., Раевский О.А. Фрактальная размерность гистограмм межатомных расстояний – новый 3D дескриптор молекулярной структуры // Журн. общ. химии. 2011. Т. 81. № 3. С. 353-359.
17. Григорьев В.Ю., Раевский О.А. Рекуррентная модель острой токсичности в гомологических рядах органических соединений // Хим.-фармац. журн. 2011. Т. 45. № 7. С. 28-33.
18. Раевский О.А., Григорьев В.Ю., Липлавская Е.А., Вортс А.П. Расчет острой токсичности химических соединений при их внутривенном введении мышам на основе локальных регрессионных моделей в перекрывающихся кластерах (ЛРМПК) // Биомед. химия. 2012. Т. 58. № 5. С. 489-500.
19. Григорьев В.Ю., Раевский О.А. Исследование количественной связи между функцией радиального распределения и кривой давления насыщенного пара в ряду n-алканов на основе модифицированной линейной динамической модели // Журн. физ. химии. 2012. Т. 86. № 3. С. 417-422.
20. Григорьев В.Ю., Раздольский А.Н., Раевский О.А. Исследование острой токсичности химических соединений по отношению к *Guppy* с использованием линейного дискриминантно - регрессионного подхода // Хим.-фармац. журн. 2012. Т. 46. №8. С. 7-11.

Другие публикации

21. Raevsky O.A., Grigor'ev V.Yu., Solov'ev V.P., Kireev D.B., Sapegin A.M., Zefirov N.S. Drug Design H-bonding Scale // In: QSAR: Rational Approaches to the Design of Bioactive Compounds / Silipo C., A. Vittoria A., Eds. Amsterdam: Elsevier Science Publishers, 1991. P. 135-138.
22. Григорьев В.Ю., Раевская О.Е., Раевский О.А. База данных «Термодинамика водородной связи» // Тезисы IX Всесоюзной конференции «Химическая информатика». Черногловка, МО. 1992. С. 33.

23. Raevsky O.A., Grigor'ev V.Yu., Kireev D.B., Lukoyanov N.V., Glinka R., Brzezinska E. QSAR Models of β - Adrenergic Activity of Some Substituted Phenethylamines // Proceedings of the 12-th International Symposium on Medicinal Chemistry. Basel. 1992. P. P-037A.
24. Raevsky O.A., Grigor'ev V.Yu., Mednikova E. QSAR H-Bonding Descriptors // Proceedings of the 9-th European Symposium on Structure-Activity Relationships: QSAR and Molecular Modelling. Strasbourg. 1992. P. O-22.
25. Raevsky O., Dolmatova L., Grigor'ev V., Lisyansky I., Bondarev S. Molecular Recognition Descriptors in QSAR // In: QSAR and Molecular Modelling: Concepts, Computational Tools and Biological Applications / Sanz F., Giraldo J., Eds. Barselona: Prous Science Publ., 1995. P. 241-245.
26. Раевский О.А., Григорьев В.Ю., Трепалин С.В., Шапер К. Расчет гидрофобности на основе поляризуемости и водородной связи // Тезисы IV Российского конгресса «Человек и лекарство». Москва. 1997. С. 288.
27. Раевский О.А., Григорьев В.Ю., Трепалин С.В., Герасименко В.А., Раздольский А.Н., Шапер К. Оптимизация компьютерного конструирования лекарств на основе физико-химических параметров // Тезисы VI Российского конгресса «Человек и лекарство». Москва. 1999. С. 463.
28. Раевский О.А., Григорьев В.Ю., Трепалин С.В. Свидетельство об официальной регистрации программы для ЭВМ НУВОТ (Hydrogen Bond Thermodynamics) № 990090 от 26 февраля 1999 г., Москва, Федеральная служба по интеллектуальной собственности, патентам и товарным знакам.
29. Григорьев В.Ю. Расчет вклада водородной связи в энтальпию испарения и нормальную точку кипения органических соединений // Рукопись деп. ВИНТИ 04.06.99. № 1820-В99. 22 с.
30. Григорьев В.Ю. Комплексный метод оценки протонодонорной способности фенолов // Рукопись деп. ВИНТИ 19.07.2000. № 2012-В00. 13 с.
31. Раевский О.А., Скворцов В.С., Григорьев В.Ю., Трепалин С.В. Свидетельство об официальной регистрации программы для ЭВМ Программный комплекс НУВОТ (Hydrogen Bond Thermodynamics) in UNIX

№ 2002610496 от 4 апреля 2002 г., Москва, Федеральная служба по интеллектуальной собственности, патентам и товарным знакам.

32. Grigor'ev V.Yu., Weber E.E., Raevsky O.A. Classification SAR Models of Chemicals Toxicity to Guppy, Fathead Minnow and Rainbow Trout on the Basis of Structure Similarity // Fourth International Symposium on Computational Methods in Toxicology and Pharmacology Integrating Internet Resources. Moscow. 2007. P. 99.
33. Tikhonova O.V., Skvortsov V.S., Grigor'ev V.Yu., Raevsky O.A. 3D-QSAR Models of Irreversible and Reversible Inhibitors of Acetylcholinesterase // Fourth International Symposium on Computational Methods in Toxicology and Pharmacology Integrating Internet Resources. Moscow. 2007. P. 162.
34. Raevsky O.A., Grigor'ev V.Yu., Modina E.A., Raevskaya O.E., Liplavsky Ja.V. Classification- and Regression-Based QSAR of Chemical Toxicity on the Basis of Structural and Physicochemical Similarity // Workshop on "Satisfying REACH Requirements in Predictive Toxicology". Rome. 2009. P. 12-13.
35. Григорьев В.Ю., Раевский О.А. Рекуррентная модель острой токсичности гомологического ряда нормальных спиртов // Материалы XVII Российского национального конгресса «Человек и лекарство». Москва. 2010. С. 600.
36. Раевский О.А., Григорьев В.Ю., Модина Е.А. Компьютерные модели взаимосвязи структуры органических соединений и их острой токсичности // Материалы XVII Российского национального конгресса «Человек и лекарство». Москва. 2010. С. 708-709.
37. Raevsky O.A., Grigor'ev V.Yu., Modina E.A., Worth A.P. Classification- and Regression-Based QSAR of Acute Chemical Rodent Toxicity // Proceedings of the 18-th European Symposium on Quantitative Structure-Activity Relationships. Rhodes. 2010. P. 77-78.
38. Григорьев В.Ю., Раевский О.А. Гистограмма межатомных расстояний – новый 3D дескриптор молекулярной структуры // Материалы XVIII Российского национального конгресса «Человек и лекарство». Москва. 2011. С. 552.

39. Липлавская Е.А., Григорьев В.Ю., Раевский О.А. Предсказание острой токсичности органических соединений по отношению к грызунам на основе структурного и физико-химического сходства // Материалы XVIII Российского национального конгресса «Человек и лекарство». Москва. 2011. С. 612.
40. Раевский О.А., Трепалин С.В., Григорьев В.Ю., Солодова С.Л., Ярков А.В., Раевская О.Е. Свидетельство о государственной регистрации программы для ЭВМ Программный комплекс для оценки и предсказания проницаемости физиологически активных веществ через гематоэнцефалический барьер № 2012614339 от 16.05.2012, Москва, Федеральная служба по интеллектуальной собственности.