



**ИЗБРАННЫЕ ТРУДЫ
ДОКТОРА ХИМИЧЕСКИХ НАУК,
ПРОФЕССОРА О.А. РАЕВСКОГО
В ОБЛАСТИ МОЛЕКУЛЯРНОГО
КОМПЬЮТЕРНОГО ДИЗАЙНА**

Москва
2021

УДК 74; 544
ББК 24.5; 85с
Р16

Избранные труды доктора химических наук, профессора О.А. Раевского в области молекулярного компьютерного дизайна / – М.: РАН, 2021 – 228 с., 35 илл.

СОДЕРЖАНИЕ

О ПРОФЕССОРЕ О.А. РАЕВСКОМ	5
ПРЕДСТАВЛЕНИЕ МОЛЕКУЛЯРНОЙ СТРУКТУРЫ В ВИДЕ СПЕКТРА МЕЖАТОМНЫХ РАССТОЯНИЙ ДЛЯ ИЗУЧЕНИЯ СВЯЗИ «СТРУКТУРА – БИОЛОГИЧЕСКАЯ АКТИВНОСТЬ»	7
ВОЗМОЖНОСТИ И ПЕРСПЕКТИВЫ КОНСТРУИРОВАНИЯ БИОЛОГИЧЕСКИ АКТИВНЫХ ВЕЩЕСТВ	18
НОВЫЕ КССА ДЕСКРИПТОРЫ, РАССЧИТЫВАЕМЫЕ ИЗ СПЕКТРОВ МЕЖАТОМНЫХ ВЗАИМОДЕЙСТВИЙ	49
КЛАССИФИКАЦИОННЫЕ И КОЛИЧЕСТВЕННЫЕ МОДЕЛИ ВЗАИМОСВЯЗИ СТРУКТУРЫ ХИМИЧЕСКИХ СОЕДИНЕНИЙ И ИХ ТОКСИЧНОСТИ ПО ОТНОШЕНИЮ К <i>DAPHNIA MAGNA</i>	56
РАЗВИТИЕ МОДЕЛЕЙ ВЗАИМОСВЯЗИ СТРУКТУРЫ И ТОКСИЧЕСКОГО ДЕЙСТВИЯ ХИМИЧЕСКИХ СОЕДИНЕНИЙ ПО ОТНОШЕНИЮ К <i>GUPPY</i>	71
ФРАКТАЛЬНАЯ РАЗМЕРНОСТЬ ГИСТОГРАММ МЕЖАТОМНЫХ РАССТОЯНИЙ – НОВЫЙ 3D-ДЕСКРИПТОР МОЛЕКУЛЯРНОЙ СТРУКТУРЫ	82
КОМПЬЮТЕРНЫЕ МОДЕЛИ ВЗАИМОСВЯЗИ СТРУКТУРЫ ОРГАНИЧЕСКИХ СОЕДИНЕНИЙ И ИХ ИНГАЛЯЦИОННОЙ ТОКСИЧНОСТИ	94
РЕКУРРЕНТНАЯ МОДЕЛЬ ОСТРОЙ ТОКСИЧНОСТИ В ГОМОЛОГИЧЕСКИХ РЯДАХ ОРГАНИЧЕСКИХ СОЕДИНЕНИЙ ...	103
КОМПЬЮТЕРНЫЕ КЛАССИФИКАЦИОННЫЕ МОДЕЛИ ВЗАИМОСВЯЗИ СТРУКТУРЫ ХИМИЧЕСКИХ СОЕДИНЕНИЙ И ЛЕКАРСТВ С ИХ СПОСОБНОСТЬЮ ПРОНИКАТЬ ЧЕРЕЗ ГЕМАТОЭНЦЕФАЛИЧЕСКИЙ БАРЬЕР	113
ИССЛЕДОВАНИЕ КОЛИЧЕСТВЕННОЙ СВЯЗИ МЕЖДУ ФУНКЦИЕЙ РАДИАЛЬНОГО РАСПРЕДЕЛЕНИЯ И КРИВОЙ ДАВЛЕНИЯ НАСЫЩЕННОГО ПАРА В РЯДУ <i>n</i>-АЛКАНОВ НА ОСНОВЕ МОДИФИЦИРОВАННОЙ ЛИНЕЙНОЙ ДИНАМИЧЕСКОЙ МОДЕЛИ	128
ОЦЕНКА ОСТРОЙ ТОКСИЧНОСТИ ОРГАНИЧЕСКИХ СОЕДИНЕНИЙ ПРИ ВНУТРИВЕННОМ ВВЕДЕНИИ ПО ОТНОШЕНИЮ К МЫШАМ НА ОСНОВЕ МЕЖВИДОВЫХ КОРРЕЛЯЦИЙ, ПАРАМЕТРОВ ЛИПОФИЛЬНОСТИ И ФИЗИКО-ХИМИЧЕСКИХ ДЕСКРИПТОРОВ	138

КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ ПРОНИЦАЕМОСТИ ФИЗИОЛОГИЧЕСКИ АКТИВНЫХ ВЕЩЕСТВ ЧЕРЕЗ ГЕМАТОЭНЦЕФАЛИЧЕСКИЙ БАРЬЕР	149
КЛАССИФИКАЦИОННЫЕ QSAR-МОДЕЛИ ОСТРОЙ ТОКСИЧНОСТИ ОРГАНИЧЕСКИХ СОЕДИНЕНИЙ ПО ОТНОШЕНИЮ К <i>DAPHNIA MAGNA</i>	172
КОМПЬЮТЕРНЫЙ РАСЧЕТ ПРОНИЦАЕМОСТИ ФАРМАКОЛОГИЧЕСКИХ ВЕЩЕСТВ ЧЕРЕЗ ГЕМАТОЭНЦЕФАЛИЧЕСКИЙ БАРЬЕР	179
КЛАССИФИКАЦИОННЫЕ МОДЕЛИ ВЗАИМОСВЯЗИ СТРУКТУРЫ ЛЕКАРСТВЕННЫХ СОЕДИНЕНИЙ И ИХ Р-ГЛИКОПРОТЕИНОВОЙ АКТИВНОСТИ	185
QSAR-МОДЕЛИРОВАНИЕ ОСТРОЙ НЕЙРОТОКСИЧНОСТИ РЯДА ОРГАНИЧЕСКИХ РАСТВОРИТЕЛЕЙ ПО ОТНОШЕНИЮ К ГРЫЗУНАМ	195
QSAR-МОДЕЛИРОВАНИЕ БЛОКАДЫ НМДА-РЕЦЕПТОРА ПОЛИФАРМАКОФОРНЫМИ СОЕДИНЕНИЯМИ НА ОСНОВЕ ПРОИЗВОДНЫХ КАРБАЗОЛА И 1-АМИНОАДАМАНТАНА	203
КОМПЬЮТЕРНЫЕ РЕГРЕССИОННЫЕ МОДЕЛИ Р-ГЛИКОПРОТЕИНОВОГО ТРАНСПОРТА ЛЕКАРСТВЕННЫХ СОЕДИНЕНИЙ	208
ИССЛЕДОВАНИЕ КАНАЛЬНЫХ БЛОКАТОРОВ NMDA-РЕЦЕПТОРА В РЯДУ КОНЬЮГАТОВ МЕТИЛЕНОВОГО СИНЕГО С ИСПОЛЬЗОВАНИЕМ QSAR И МОЛЕКУЛЯРНОГО МОДЕЛИРОВАНИЯ	217



РАЕВСКИЙ ОЛЕГ АЛЕКСЕЕВИЧ
10.05.1940 – 05.03.2019

Олег Алексеевич Раевский начал работать в Институте физиологически активных веществ РАН с июня 1979 года и стал организатором и научным руководителем лаборатории физико-химических исследований, которая в 1989 году была преобразована в отдел компьютерного молекулярного дизайна. Одновременно с 1989 по март 2016 года он был заместителем директора по научной работе.

О.А. Раевский родился 10 мая 1940 года в городе Ростов-на-Дону, где окончил в 1957 году среднюю школу и в 1962 году Химический факультет Ростовского Государственного Университета по специальности «физическая химия» и получил квалификацию химика-исследователя. В период с 1962 по 1979 год работал в Институте органической и физической химии им. А.Е. Арбузова Казанского филиала Академии Наук СССР, занимая последовательно должности лаборанта, аспиранта, младшего научного сотрудника, старшего научного сотрудника.

За указанный период О.А. Раевский защитил диссертации на соискание ученой степени кандидата химических наук (1966) и доктора химических наук (1978). Стал одним из первых в стране председателем Совета молодых ученых.

Область научных интересов Олега Алексеевича была очень разнообразной: физическая химия, спектроскопия, молекулярное моделирование, исследование связи «структура – свойство», конструирование и компьютерный поиск новых физиологически активных соединений с заданными свойствами (лекарств), компьютерный молекулярный дизайн, разработка компьютерных программ и создание баз данных. Им предложены оригинальные расчетные методы, включая оценку количественного вклада водородной связи в свойства и биологическую активность химических соединений, степень сходства химических структур, дискретно-регрессионные модели для установления взаимосвязей «структура – активность».

Олег Алексеевич является автором более 400 научных публикаций, включая два учебных пособия для студентов химических вузов и две монографии: «Свойства химических соединений и лекарств как функции их структуры» (2013) и «Моделирование соотношений “структура – свойство”» (2015). В 1983–1986 годах Олег Алексеевич читал факультативный курс «Конструирование биологически активных соединений» в МХТИ им. Д.И. Менделеева. В 1988 году О.А. Раевскому присвоено ученое звание профессора по специальности физическая химия.

Совместно с сотрудниками О.А. Раевский создал серию компьютерных программ, зарегистрированных в Роспатенте и успешно используемых в исследованиях по компьютерному молекулярному дизайну в ряде отечественных и зарубежных университетов и фирм.

О.А. Раевский являлся одним из организаторов Российской Секции Международного научного общества QSAR and Molecular Modeling Society, председателем которой избирался дважды на пятилетний срок в период с 1996 по 2006 год. С 1994 года являлся членом Правления указанного международного общества. С 1993 по 2005 год О.А. Раевский был членом редколлегии международных научных журналов QSAR & Combinatorial Science и SAR and QSAR in Environmental Research. Он был включен в 5000 персоналий мира в 5-м издании Американского биографического института. Олег Алексеевич был активным участником симпозиумов «Биоинформатика и компьютерное конструирование лекарств», которые проводились в рамках Российских национальных конгрессов «Человек и лекарство». 24 апреля 2013 года О.А. Раевскому было присвоено почетное звание «Основатель научной школы «Компьютерный молекулярный дизайн физиологически активных веществ». Под его руководством было защищено 15 кандидатских и докторских диссертаций.

В.П. Новиков, О.А. Раевский

ПРЕДСТАВЛЕНИЕ МОЛЕКУЛЯРНОЙ СТРУКТУРЫ В ВИДЕ СПЕКТРА МЕЖАТОМНЫХ РАССТОЯНИЙ ДЛЯ ИЗУЧЕНИЯ СВЯЗИ «СТРУКТУРА – БИОЛОГИЧЕСКАЯ АКТИВНОСТЬ»

*Институт физиологически активных веществ РАН, Черноголовка,
Московская область*

Установление и выяснение природы связи «молекулярная структура вещества – его свойство» является стратегическим направлением химической науки. Для фармакологов и химиков-синтетиков особый интерес представляет такое свойство, как биологическая активность, поскольку анализ связи «структура – активность» (ССА) может существенно помочь направленному поиску веществ с заданным биологическим действием, а также установлению механизма этого действия. Прямое моделирование биологического ответа на введенный в систему реагент наталкивается на непреодолимые трудности из-за необходимости учета большого числа многостадийных химических реакций, протекающих в различных средах, а также учета разнообразных физико-химических и биологических процессов (растворимость в разных средах, прохождение через клеточные мембраны, транспорт и т. д.), многие из которых до конца не изучены. Это обстоятельство заставляет исследователей обращаться к статистическим методам установления ССА. Такие методы, первоначально основанные на регрессионном анализе [1], а затем – на различных модификациях теории распознавания образцов [2, 3], получили широкое распространение. Необходимо отметить, что в последнее время применение этих методов и разработка новых в основном направлены не на подтверждение существования ССА, а на выявление молекулярных характеристик, ответственных за проявление той или иной активности.

В основе использования различных алгоритмов теории распознавания образов лежит предположение, что определенный вид активности обусловлен некоторыми общими для данного класса веществ характеристиками молекул. Поэтому успех распознавания практически полностью определяется выбором базиса пространства дескрипторов (молекулярных характеристик), в котором представляется каждый образ (молекулярная структура). В качестве дескрипторов выбирают фрагменты молекулярной структуры [4, 5], которые дополняются дескрипторами, учитывающими взаимное расположение этих фрагментов [6] и физико-химические параметры [7]: молеку-

лярную массу, молекулярную связуемость, молекулярный объем, главные моменты инерции и т. д.

Несмотря на известные успехи этого подхода в анализе ССА, использование фрагментарных дескрипторов сопровождается принципиальными трудностями, которые, на наш взгляд, не позволят на этом пути достичь существенного прогресса в более тонком и глубоком понимании закономерностей ССА. Основные трудности заключаются в следующем.

Во-первых, хорошо известно, что одинаковые структурные фрагменты и группы в различных соединениях могут обладать различной реакционной способностью, которая определяется распределением электронной плотности в молекуле и довольно сложным образом зависит от окружения данного фрагмента, эффектов сопряжения и различных индукционных эффектов. Учет таких различий ведет к неоправданно большому увеличению числа дескрипторов и сильно усложняет алгоритмизацию описания молекулярной структуры. Особую сложность представляет случай, когда разные фрагменты обладают близкими химическими свойствами.

Во-вторых, большую сложность представляет описание пространственной структуры молекулы, которая играет важную роль в процессах транспорта и определения сродства реагента к рецептору.

В-третьих, дополнительную сложность представляет необходимость учета динамики внутримолекулярных движений (колебаний большой амплитуды и конформационных равновесий), в результате которых эффективная структура молекулы может существенно отличаться от равновесной структуры, соответствующей минимуму потенциальной энергии, что вызывает изменение реакционной способности. С необходимостью учета этих факторов при изучении механизма химических реакций исследователи сталкиваются повседневно. Ясно, что это является необходимым и при изучении ССА.

Таким образом, дальнейший прогресс в углублении представлений о природе ССА, по-видимому, должен быть связан с разработкой такого описания молекулярной структуры, которое в идеальном случае учитывало бы пространственное положение и распределение электронной плотности, а также динамику их изменения при внутримолекулярных колебаниях и межмолекулярных взаимодействиях.

В качестве простого описания молекулярной структуры, довольно полно удовлетворяющего требованиям «идеального», мы предлагаем использовать представление пространственного строения молекулы в виде кривой радиального распределения, которая легко может быть вычислена для молекулы любой сложности, если известно ее пространственное строение (декартовы координаты). Аналогичный аппарат, несколько усложненный из-за специфики задачи, применяется при изучении строения молекул методом газовой электронографии [8]. Рассмотрим преобразование молекулярной структуры в функцию радиального распределения, физический смысл такого преобразования и возможные алгоритмы для анализа ССА.

Молекулярное рассеяние и функция радиального распределения

Задача о рассеянии электронов на произвольно ориентированных молекулах решается точно методами квантовой теории рассеяния. Интенсивность рассеянных электронов в зависимости от угла рассеяния может быть представлена в виде двух слагаемых, первое из которых соответствует рассеянию на атомах, составляющих молекулу, а второе определяется взаимным расположением атомов в пространстве, т. е. структурой молекулы, и называется молекулярным рассеянием. В приближении сферической симметрии атомов и отсутствии внутримолекулярных колебаний выражение для молекулярного рассеяния N -атомной молекулы с точностью до постоянного коэффициента имеет вид [8]:

$$M(s) = \sum_{i=1}^N \sum_{j>i}^N C_{ij} \frac{\sin sr_{ij}}{sr_{ij}},$$

где s —угловой параметр: $s = \frac{4\pi}{\lambda} \sin \frac{\theta}{2}$; λ — длина волны электронов; θ — угол рассеяния; r_{ij} – межатомное расстояние в молекуле; C_{ij} – коэффициент, характеризующий рассеивающую способность данной пары атомов и определяющийся их видом. Важным является тот факт, что C_{ij} имеет практическое постоянное значение по всей шкале s и с достаточной степенью точности может быть представлен в виде:

$$C_{ij} = \frac{Z_i Z_j}{\sum_{k=1}^N (Z_k^2 + Z_k)},$$

где Z_i – порядковый номер i -го атома.

$M(s)$ представляет собой затухающую осциллирующую функцию, которая, как правило, полностью затухает при $s > 40\text{--}45 \text{ \AA}^{-1}$. Таким образом, имея весь набор межатомных расстояний, число которых в общем случае $N(N - 1)/2$, по уравнению (1)–(2) можно легко вычислить молекулярное рассеяние.

Следует отметить интересную попытку использовать преобразование молекулярной структуры по уравнениям, аналогичным (1)–(2), для разделения транквилизаторов и седативных средств [9]. Для кодирования структуры использовали точки s , в которых $M(s) = 0$. Для этого интервал $s = 1 - 31 \text{ \AA}^{-1}$, в котором рассчитывались $M(s)$, разбивали на 100 дескрипторов, каждый из которых принимал значение либо 0 [отсутствие пересечения $M(s)$ с осью s], либо 1 (есть пересечение). При использовании бинарного классификатора разделение оказалось намного лучше, чем использование на том же исходном материале фрагментарных дескрипторов [5]. Этот успех, по-видимому, обусловлен лучшим учетом пространственного строения молекул. Тем не менее этот подход не может быть рекомендован в качестве общепотребительного.

тельного для анализа ССА, так как при этом учитывается лишь небольшая часть структурной информации, заключенной в уравнении (1). Опыт структурного анализа электроннограмм показывает, что, как правило, форма и амплитуда $M(s)$ содержат существенно больше информации о структуре, чем точки «зануления». Кроме того, при этом подходе затруднен анализ полученных результатов разделения с точки зрения выделения структурных фрагментов или критических межатомных расстояний, ответственных за данный вид активности.

Учесть всю информацию о структуре и получить физически наглядную интерпретацию преобразования молекулярной структуры позволяет следующий подход. Перепишем уравнение (1) с учетом того, что каждое межатомное расстояние изменяется в процессе внутримолекулярных колебаний и его необходимо усреднить с помощью функции плотности вероятности $P_{ij}(r)$:

$$M(s) = \sum_{i=1}^N \sum_{j>i}^N C_{ij} \int_0^{\infty} P_{ij}(r) \frac{\sin sr}{sr} dr.$$

Используя свойство синус-преобразования Фурье, получаем:

$$\sum_{i=1}^N \sum_{j>i}^N C_{ij} P_{ij}(r)/r = \frac{2}{\pi} \int_0^{\infty} s M(s) \sin sr ds.$$

Левая часть уравнения (4) есть функция радиального распределения $f(r)$, которая по существу представляет собой спектр межатомных расстояний в молекуле. Она может быть вычислена, если известна функция $M(s)$. Поскольку верхний предел интегрирования в уравнении (4) ограничен каким-то предельным значением s_{\max} , для уменьшения волны обрыва обычно используют затухающую функцию $\exp(-bs^2)$, параметр в которой оценивают из соотношения

$$\exp(-bs_{\max}^2) = 0,1.$$

Выражение для кривой радиального распределения приобретает вид:

$$f(r) = \frac{2}{\pi} \int_0^{s_{\max}} s M(s) \exp(-bs^2) \sin sr ds.$$

Введение затухающей функции несколько уширяет пики на кривой $f(r)$, но практически не смещает их положения.

Вид функций $M(s)$ и $f(r)$

Вид функции $M(s)$, а значит и $f(r)$, определяется способом учета внутримолекулярных колебаний, т. е. видом функции $P_{ij}(r)$. Общая теория изложе-

на в работе [8]. Для использования $f(r)$ в анализе ССА практический интерес представляет рассмотрение трех случаев.

1. Жесткая модель молекулы. В отсутствие колебаний функция $P_{ij}(r)$ представляет собой дельта-функцию:

$$P_{ij}(r) = \delta(r - r_{ij}).$$

Уравнение (3) вырождается в уравнение (1), а пики на кривой $f(r)$, согласно уравнению (4), будут просто линиями при $r = r_{ij}$, высота которых пропорциональна $nZ_i Z_j / r_{ij}$, где n – кратность расстояния r_{ij} в молекуле. При использовании уравнения (6) $f(r)$ будет суммой гауссовых пиков, максимумы которых лежат при $r = r_{ij}$, а полуширина равна $\sqrt{2b}$.

2. Приближение малых гармонических колебаний. Квантово-механическое рассмотрение задачи о малых колебаниях атомов в молекуле приводит к выражению $P_{ij}(r)$ в виде гауссовой функции [8], подстановка которой в уравнение (3) с последующим интегрированием дает в хорошем приближении

$$M(s) = \sum_{i=1}^N \sum_{j>i}^N C_{ij} \exp(-l_{ij}^2 s^2 / 2) \frac{\sin sr_{ij}}{sr_{ij}},$$

где l_{ij} – среднеквадратичная амплитуда колебаний данной пары атомов. Амплитуды колебаний можно вычислить по силовому полю молекулы, используя известные методики [10]. Однако достаточно хорошая оценка для величин l_{ij} может быть сделана по эмпирическим уравнениям $l_{ij} = F(r_{ij})$, которые получены статистической обработкой большого экспериментального материала и представляют собой практически линейную зависимость для различных пар атомов [11, 12].

Кривая $f(r)$ будет суммой гауссовых пиков с максимумами при $r = r_{ij}$ и полушириной $\sqrt{(2b+l_{ij}^2)}$. Последнее обстоятельство имеет важное значение для моделирования гармонических колебаний в молекуле при использовании в расчетах $M(s)$ жесткой модели. Действительно, используя в расчетах $M(s)$ и $f(r)$ уравнений (1), (6) и обрывая $M(s)$ при $s = 40 \text{ \AA}^{-1}$, получаем из уравнения (5) $b = 0,0015 \text{ \AA}^2$. Тогда полуширина пика на кривой $f(r)$ будет порядка $\sqrt{2b} + 0,055 \text{ \AA}$, что соответствует реальным значениям амплитуд колебаний атомов, связанных химической связью [10]. При увеличении межатомных расстояний для $r > 2 - 3 \text{ \AA}$ амплитуды колебаний увеличиваются и могут достигать величины порядка $0,15 \text{ \AA}$ [10]. Таким образом, использование жесткой модели несколько сужает пики на кривой $f(r)$ в области несвязанных расстояний по сравнению с моделью малых гармонических колебаний. При этом положение максимумов не смещается. Поэтому для целей анализа ССА, по-видимому, в достаточно хорошем приближении можно не учитывать малые колебания атомов. Представленные на рисунках 1 и 2 кривые $M(s)$ и $f(r)$ для молекулы бензола иллюстрируют это положение.

3. Колебания больших амплитуд. Этот случай существенно усложняет задачу определения вида функции $P_{ij}(r)$ в общем виде [8]. Однако наиболее часто встречающийся на практике случай крутильных колебаний большой амплитуды, которые приводят к свободному или заторможенному внутреннему вращению и образованию нескольких конформеров, может быть учтен достаточно просто. При внутреннем вращении, например, вокруг связи С–С в молекуле несимметричного трихлорэтана, изменяются только те межатомные расстояния, которые зависят от вращения, т. е. расстояния между атомами водорода и хлора. Для каждого такого расстояния вклад в $M(s)$ может быть учтен с помощью классического распределения:

$$m_{ij}(s) = C_{ij} \int_0^{2\pi} \exp\left[-\frac{V(\varphi)}{RT}\right] \exp[-l_{ij}^2(\varphi)s^2/2] \frac{\sin s_{ij}(\varphi)}{sr_{ij}(\varphi)} d\varphi$$

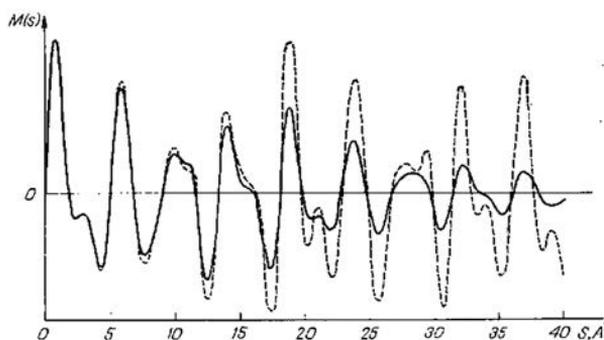


Рис. 1. Молекулярная составляющая для молекулы бензола, рассчитанная для жесткой модели (пунктирная линия) и с учетом внутримолекулярных колебаний (сплошная линия).

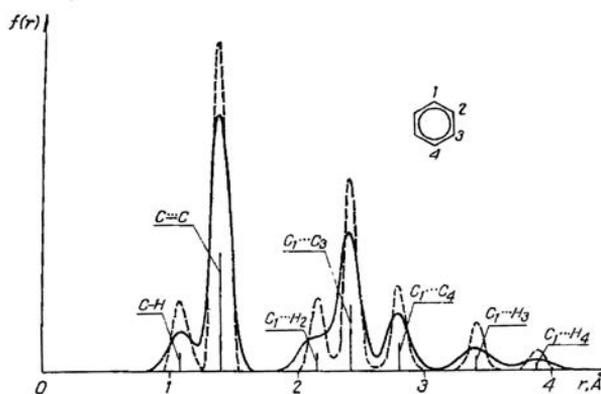


Рис. 2. Кривая радиального распределения для молекулы бензола и отнесение пиков. Пунктирная линия – жесткая модель; сплошная линия – модель с учетом внутримолекулярных колебаний

где φ – угол вращения; $l_{ij}(\varphi)$ – амплитуда колебания остова (этот член может быть опущен, если малые гармонические колебания остова не учитываются); R – универсальная газовая постоянная; T – абсолютная температура; $V(\varphi)$ – потенциал вращения, который в простейших случаях может быть задан в виде:

$${}^2V(\varphi) = V_0 (1 - \cos n\varphi), \quad (10)$$

где V_0 – барьер вращения; n – кратность потенциала. В случае свободного вращения ($V_0 = 0$) и жесткой модели остова ($l_{ij} = 0$) выражение (9) упрощается:

$$m_{ij}(s) = C_{ij} \int_0^{2\pi} \frac{\sin sr_{ij}(\varphi)}{sr_{ij}(\varphi)} d\varphi. \quad (11)$$

Выражения (9) и (11) можно интегрировать численно, рассчитав значения $r(\varphi)$ с определенным шагом по интервалу изменения угла φ . При наличии конформеров молекулярное рассеяние может быть вычислено следующим образом:

$$M(s) = \sum_{k=1}^n w_k M_k(s) \quad (12)$$

где n число конформеров; w_k – масса каждого конформера; $M_k(s)$ – его молекулярное рассеяние. В этом случае возможно расщепление пиков на кривой $f(r)$.

Внутреннее вращение сильно уширяет пики на кривой $f(r)$ для межатомных расстояний, изменяющих свою величину при вращении. Например, в молекуле $\text{Sn}(\text{C}\equiv\text{C}-\text{CF}_3)_4$ существует практически свободное вращение CF_3 -групп [13], что приводит к уширению пиков на кривой $f(r)$ в области $r > 6 \text{ \AA}$ (рисунок 3) соответствующей расстояниям $\text{F} \dots \text{F}$, $\text{C} \dots \text{F}$ между различными трифторпропильными группами.

Возможности использования функции $f(r)$ в анализе ССА

Функция радиального распределения представляет собой спектр межатомных расстояний, интенсивность которого определяется видом атомов, находящихся на данном расстоянии друг от друга, и динамикой изменения этого расстояния в процессе внутримолекулярных колебаний. Можно сказать, что функция $f(r)$ – это то, что «видит» электрон, который взаимодействует с данной молекулой в акте рассеяния. Эта аналогия может быть продолжена на случай взаимодействия рецептора с молекулами введенного в организм вещества, вызывающего определенную биологическую реакцию. Согласно современным представлениям, взаимодействие рецептора происходит не со всей молекулой реагента, а с частями молекулы, комплементарными активным центрам рецептора. При этом определяющими факторами взаимодействия являются расстояния между активными центрами и вид

атомов в молекуле, которые находятся на этих расстояниях. Имеет место своеобразный процесс рассеяния молекул реагента на активных центрах рецептора, т. е. рецептор «видит» молекулу в образе функции $f(r)$. Поэтому представление молекулярной структуры в виде $f(r)$ при анализе ССА методами теории распознавания образов выглядит естественно.

Следует отметить, что в принципе существует возможность модифицировать функцию $f(r)$ с тем, чтобы учесть распределение электронной плотности на каждом из атомов в молекуле. Этот шаг позволит описывать не только вид атомов, взаимодействующих с активными центрами рецептора, но и способность атомов к таким взаимодействиям, которые могут носить характер ион-ионного, диполь-дипольного, ион-дипольного и других видов взаимодействия. Для этого в выражении (2) для коэффициента C_{ij} , определяющего интенсивность пиков на кривой $f(r)$, в качестве Z можно использовать параметр, характеризующий изменение электронной плотности на атоме в данной молекуле по сравнению с изолированным атомом или атомом, находящимся в стандартном состоянии. В простейшем случае этим параметром может быть эффективный заряд на атоме, рассчитанный с использованием различия в электроотрицательностях соседних атомов или методами квантовой химии. По-видимому, можно использовать также другие расчетные или экспериментальные физико-химические параметры, связанные с распределением электронной плотности. Использование таких модифицированных функций радиального распределения требует специального обсуждения. В данном случае важно отметить возможность их построения и использования в анализе ССА.

Представление структуры молекулы в виде функции $f(r)$ позволяет в полной мере использовать методики теории распознавания образов, разработанные для анализа ИК-, ЯМР- и масс-спектров химических соединений [3]. Используя этот аппарат, каждое соединение можно представить в виде вектора X , m компонентов которого есть значения $f(r)$, вычисленные с определенным шагом Δr во всем интервале изменения межатомных расстояний для данной выборки. Задача разделения классов соединений, проявляющих различную активность, сводится к построению гиперплоскости, которая разделяет образы, координаты которых описаны векторами X в m -мерном пространстве. Отбор признаков, наиболее важных для классификации, связан с уменьшением размерности пространства образов. Основным достоинством использования функции $f(r)$ является содержательность анализа отобранных признаков, так как эти признаки соответствуют определенным частям спектра межатомных расстояний. Имея весь набор межатомных расстояний для каждой молекулы, легко установить, какие структурные фрагменты являются определяющими для данного разделения. Это, безусловно, важно для интерпретации закономерностей ССА.

Отбор признаков можно осуществить различными способами. Например, при использовании обучения с учителем и бинарного классификато-

ра разделение на два класса определяется знаком скалярного произведения векторов:

$$s_i = W * X_i, \quad (13)$$

где i – номер образа; W – нормаль к искомой гиперплоскости, называемая весовым вектором. Для неверно классифицированных образов новое значение весового вектора находят итерационным способом:

$$W' = W + cX_i \quad (14)$$

где коэффициент c определяется из уравнения

$$c = -2s_i / (X_i * X_i) = -2(W * X_i) / (X_i * X_i). \quad (15)$$

Итерации повторяют до тех пор, пока не будет правильно классифицировано максимальное число образов. В этом случае отбор признаков может быть сделан путем сравнения компонент векторов, полученных из различных начальных приближений. Компонента признается хорошо коррелирующей и сохраняется, если ее знак остается неизменным. В противном случае она исключается [3].

При использовании классификации по методу наименьших квадратов весовой вектор определяют минимизацией функционала:

$$Q = \sum_{i=1}^n (s_i^* - W * X_i)^2 \quad (16)$$

где s_i^* – истинное значение категории классификации; n – число образов в обучающей выборке. Здесь отбор признаков можно сделать с помощью отделения линейно зависимых уравнений в решаемой системе нормальных уравнений обычными приемами линейной алгебры [14]: разложением по сингулярным числам Якобиана:

$$J_{ij} = \delta(W * X_i) / \delta W_j \quad (17)$$

или анализом собственных значений матрицы коэффициентов нормальных уравнений ($J^T J$).

Существуют и другие способы отбора признаков [3, 15]. Выбор оптимального вычислительного алгоритма, по-видимому, можно будет сделать с накоплением опыта использования $f(r)$ при изучении ССА.

Некоторые замечания о вычислении функции $f(r)$

Вычисление функции $f(r)$ для молекулы любого строения является несложной процедурой, широко используемой в рамках более общего комплекса электронографических программ, при условии, что известны декартовы координаты атомов или набор независимых геометрических параметров, описывающих структуру молекулы (длины связей, валентные и диэдриче-

ские углы) [16]. Источником этих данных служат экспериментальные исследования строения молекул методами рентгеноструктурного анализа, газовой электронографии, микроволновой спектроскопии, нейтронографии. Если такие данные отсутствуют, можно взять геометрические параметры родственных соединений. При этом, как правило, возникают трудности с оценкой значений углов вращения и возникающих конформационных равновесий. С аналогичной проблемой приходится сталкиваться при использовании структурных данных, полученных в разных фазах (например, использование рентгеноструктурных данных для кристаллической фазы при моделировании строения молекул в газе или растворе). В тех случаях, когда установленных экспериментальных закономерностей внутреннего вращения недостаточно для моделирования конформационных особенностей строения молекулы в данной фазе, может возникнуть необходимость в целенаправленных экспериментальных исследованиях или конформационных расчетах.

Таким образом, для вычисления функции радиального распределения необходимо существенно больше знаний о строении молекулы, чем использование фрагментарных дескрипторов. Это может потребовать постановки дополнительных экспериментов и (или) расчетов, необходимых для пространственного описания строения молекулы и динамики ее изменения при колебаниях. Однако эти трудности не являются принципиальными. Вообще говоря, необходимость их преодоления диктуется не способом описания молекулярной структуры, а необходимостью использования более детальной информации о структуре молекулы для дальнейшего развития методов установления и анализа ССА.

ЛИТЕРАТУРА

1. Нижний С.В., Эпштейн Н.А. – Успехи химии, 1978, т. 47, с. 739–742.
2. Голендер В.Е., Розенблит А.Б. Вычислительные методы конструирования лекарств. Рига, 1978.
3. Джурс П., Айзенауэр Т. Распознавание образов в химии. М., 1977.
4. Авидон В.В. – Хим.-фарм. ж., 1974, № 8, с. 22–25.
5. Stuper A., Jurs P.C. – J. Am. Chem. Soc., 1975, v. 97, p. 182–187.
6. Гитлина Л.С., Голендер В.Е., Розенблит А.Б. – Хим.-фарм. ж., 1980, № 7, с. 48–52.
7. Jurs P.C. – В кн.: Международная конференция по применению ЭВМ в химии. Доклады. Новосибирск, 1979, ч. 2, с. 374–401.
8. Вилков Л.В., Анашкин М.Г., Засорин Е.З. и др. Теоретические основы газовой электронографии. М., 1974.
9. Soltzberg L., Wilkins C.L. – J. Am. Chem. Soc., 1977, v. 99, pp. 439–443.
10. Сивин С. Колебания молекул и среднеквадратичные амплитуды. М., 1971.

11. *Mastryukov V.S., Cyvin S.J.* – J. Molec. Struct., 1975, v. 25, pp. 15–25.
12. *Мастрюков В.С., Осина Е.Л., Вилков Л.В.* и др. – Ж. структурн. химии, 1976, т. 17, с. 80–85.
13. *Novikov V.P., Khaikin L.S., Vilkov L.V.* – J. Molec. Struct., 1977, v. 42, pp. 139–146.
14. *Форсайт Дж., Малькольм М., Моулер К.* Машинные методы математических вычислений. М., 1980.
15. *Фу К.* Последовательные методы в распознавании образов и обучении машин. М., 1971.
16. *Andersen B., Seip H.M., Strand T.G.* et al. Acta Chem. Scand., 1969, v. 23, pp. 3224–3234.

О.А. Раевский, А.М. Сапегин

ВОЗМОЖНОСТИ И ПЕРСПЕКТИВЫ КОНСТРУИРОВАНИЯ БИОЛОГИЧЕСКИ АКТИВНЫХ ВЕЩЕСТВ

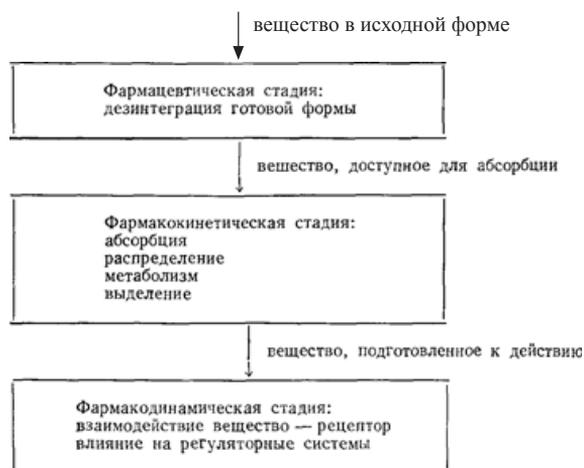
*Институт физиологически активных веществ РАН, Черноголовка,
Московская область*

I. Введение

Научно-технический прогресс сопровождается экспоненциальным ростом числа химических соединений, применяемых в промышленности, сельском хозяйстве, медицине, быту. Химия, благодаря практически беспредельным возможностям синтеза новых соединений, обеспечивает постоянное улучшение полезных свойств используемых веществ. В настоящее время число известных химических соединений оценивается примерно в десять миллионов. Однако все они составляют только ничтожно малую часть бесконечно большого числа возможных соединений. Об этом, в частности, свидетельствуют и данные, приведенные в [1]: если использовать хотя бы пятьдесят хорошо охарактеризованных в органической химии заместителей для подстановки в различном сочетании в семь свободных положений только одной нафтойной кислоты, можно получить $7,8 \cdot 10^{12}$ химических соединений. Отсюда явствует высокая вероятность синтеза соединений с гораздо лучшими, по сравнению с существующими, свойствами. В связи с этим встает очень важная проблема, как подойти к решению задачи получения высокоэффективных препаратов в кратчайший срок и с минимальными затратами. Очевидно, что использование в качестве основы для поиска новых препаратов всесторонних испытаний свойств, т. е. тотального скрининга, приведет к немислимым затратам (при реализации такого подхода для внедрения одного вещества необходим синтез и всесторонние испытания 12–15 тыс. химических соединений). Все изложенное выше ставит химическую науку перед необходимостью выработки четких принципов конструирования веществ с заданными свойствами.

Согласно [2, 3], формирование представлений о структуре вещества было тесно связано с выявлением взаимосвязи между его строением и проявляемыми свойствами: вначале – на качественном уровне [3, 4], а с открытием периодического закона элементов происходит постепенный *переход* к установлению количественных соотношений «структура – свойство» [5].

Особый интерес и значимость проблема установления связи «структура – свойство» приобретает по отношению к поиску биологически активных веществ (БАВ). Традиционные подходы, такие как выделение БАВ из природных источников и синтез их аналогов, изучение средств народной медицины, массовые биологические испытания и т. д., не в состоянии обеспечить быстро растущие потребности в новых препаратах. Поэтому с середины XIX века начинает развиваться новое направление в поиске БАВ – выявление связи между структурой вещества и проявляемой им биологической активностью (ССА). Истоки этого направления связаны с работами Ричардсона, которому впервые удалось на качественном уровне связать структуру исследуемых органических соединений с проявляемой ими физиологической активностью [6], а также, Крум-Брауна и Фрезера, предпринявших первую попытку создания математической модели, описывающей связь «структура – активность» [7]. К настоящему времени (историю развития ССА см. [3]) сформулированы основные принципы установления количественных соотношений «структура – активность» (КССА). Схематически современные представления о механизме действия БАВ, положенные в основу КССА, можно изобразить в следующем виде [8]:



Только на фармакодинамической стадии взаимосвязь структуры вещества и биологического эффекта, вызываемого его введением в организм, проявляется в полной мере и является специфичной. Что же касается других процессов, связанных с трансформацией веществ, ингибированием ферментов, диссипацией веществ на различных органах и т. д., то сведения о них (хотя и постоянно расширяются [9]) носят подчас весьма ограниченный характер, затрудняя тем самым выявление КССА. Таким образом, проблема КССА является частным случаем проблемы связи «структура – свойство». Однако сложности приложения ССС к биологическим объектам заставляют выходить за традиционные рамки проблемы «структура – свойство» и

уделять внимание не только методам выявления связи, совершенствованию экспериментальных исследований структуры и форм ее описания, но и вопросам количественного определения величины биологической активности, детализации описания самих биологических процессов.

В настоящее время число публикаций, посвященных анализу КССА, исчисляется уже тысячами. В специальном журнале QSAR (Quantitative Structure-Activity Relationships), издаваемом с 1981 года, наряду с оригинальными статьями, публикуются многочисленные рефераты работ по КССА, опубликованных в других изданиях, в том числе монографий и обзоров [10–20]. Однако в подавляющем большинстве такие работы носят либо специализированный характер и посвящены какому-либо одному подходу или вопросу – регрессионному анализу [18], использованию ЭВМ [19], приложению КССА к анализу пестицидной активности [20], либо основное внимание уделяется описанию математических методов анализа КССА [16, 17].

В настоящем обзоре акцент сделан на основополагающие принципы конструирования веществ с заданными свойствами [21, 22], составляющими физико-химический фундамент проблемы установления количественных соотношений «структура – активность». При этом предполагается, что исследователь располагает надежной количественной информацией о величине биоактивности соединений [18]. В этом случае задача установления КССА сводится к выбору модели, описывающей взаимодействие вещества с рецептором, и метода выявления связи между корректным (в рамках выбранного метода) представлением структуры соединения и проявляемой им активностью.

II. Молекулярное распознавание – платформа связи «структура – активность»

В настоящее время стало очевидным, что при анализе самых разнообразных свойств химических соединений необходимо рассматривать их межмолекулярные взаимодействия с возможными партнерами. Многочисленные факты, свидетельствующие о том, что определенная молекула A_0 (распознаватель) реагирует или образует комплекс в растворе только с определенным молекулярным эффектором B_0 из смеси иногда очень большого числа возможных эффекторов B_i , легли в основу так называемой «проблемы распознавания в химии» [23–25].

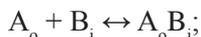
В хронологическом порядке с качественной стороны проблемы распознавания первыми столкнулись химики-аналитики при поиске веществ, эффективно связывающих определенные ионы в многокомпонентных растворах. С развитием химической науки проблема распознавания распространяется на органическую и физическую химию. Здесь уместно упомянуть о фундаментальных работах Ван-Гоффа по созданию основ теории растворов [26], что можно рассматривать как переход к изучению количе-

ственной стороны проблемы распознавания. Как отмечено в [25], при количественном описании процессов точность распознавания химических соединений увеличивается на порядок в последовательности от избирательного донорно-акцепторного взаимодействия или Н-связывания молекул, содержащих по одному электронодонорному и акцепторному центру, и, далее, к избирательному комплексообразованию ионов металлов с полидентатными лигандами, избирательным каталитическим реакциям, вплоть до ферментативных реакций. В этой же последовательности увеличиваются размер и сложность структуры молекул, участвующих в процессе распознавания.

Методологические принципы проблемы распознавания в химии преимущественно изучаются в настоящее время на примерах избирательного комплексообразования макроциклических полиэфиров и криптандов (см., например, [27, 28]). Однако даже для этих не очень сложных соединений стабильность их комплексов зависит от многих факторов, что затрудняет понимание принципов молекулярного распознавания. Поэтому необходимо четко сформулировать основные положения процесса распознавания.

Согласно [23–25], распознаватель A_0 должен связываться с молекулами определенного типа B_0 , т. е. правильным эффектором из раствора, содержащего N типов молекул: $B_0, B_1, B_2, \dots, B_{N-1}$ произвольной концентрации. Обозначим через P_f количество возможных фальшивых распознаваний. Тогда степень распознавания может быть охарактеризована числом правильных распознаваний, зависящим от общего числа возможных эффектов N и доли фальшивого распознавания P_f .

В простейшем случае образования комплекса A_0B_i



с константой реакции

$$\frac{[A_0B_i]}{[A_0][B_i]} = K_{oi} = \exp\left(-\frac{\Delta G}{RT}\right); i = 0, 1, \dots, N - 1$$

число возможных фальшивых распознаваний P_f , будет равно:

$$P_f = \frac{\sum_{i=1}^{N-1} [A_0 B_i]}{[A_0 B_0] + \sum_{i=1}^{N-1} [A_0 B_i]} = \frac{\sum_{i=1}^{N-1} [B_i] K_{oi}}{[B_0] K_{00} + \sum_{i=1}^{N-1} [B_i] K_{oi}},$$

из чего следует, что распознавание будет тем точнее, чем больше величина отношения констант связывания правильного распознавания к фальшивому или, иначе, чем больше разность свободных энергий ($\Delta\Delta G$) образования комплексов $A_0B_i; (\Delta G_{oi})$ и $A_0B_0 (\Delta G_{00})$ соответственно:

$$\Delta\Delta G_i = \Delta G_{oi} - \Delta G_{00}.$$

Таким образом, высокоспецифичное распознавание требует довольно большого различия $\Delta\Delta G$ между способностью распознавателя A_0 взаимо-

действовать с правильным эффектором V_0 и с фальшивыми. При достаточно высоких значениях $\Delta\Delta G$ распознаватель A_0 способен различать один эффектор V_0 из N других эффекторов V_i . Тем самым проблема целенаправленного поиска веществ с заданными свойствами может быть сведена к конструированию молекул, строение которых обеспечивает максимальное связывание с распознавателем, ответственным за рассматриваемое свойство.

Эти положения проблемы распознавания по отношению к биологическим объектам получили дальнейшее развитие в работах Фишера, который еще в 1894 году установил, что на характер взаимодействия субстратов с ферментом существенное влияние оказывает их пространственная совместимость [29]. Это легло в основу так называемого принципа «ключ – замок» или принципа комплементарности. Указанный принцип в настоящее время трансформировался в общепринятое положение, что образование фермент-субстратных комплексов, играющих важную роль в ферментативном катализе, определяется тонкой структурой активного и эффекторного центров и уникальной структурой всей молекулы фермента, обеспечивающими высокую каталитическую активность и специфичность действия биокатализаторов [30].

Другой важный вклад в развитие направления «структура – активность» связан с именем Хэнча, который в шестидесятых годах нашего столетия предложил эмпирическую модель биологической активности, основанную на линейной зависимости свободной энергии рассматриваемого процесса от физико-химических параметров соединений. Эта модель является наиболее распространенным инструментом в работе по оптимизации биологического действия химических соединений [1, 31].

До последнего времени развитие идей Фишера и Хэнча сформировало в задаче «структура – активность» два независимых подхода, связанных с отбором общих элементов структуры в различных соединениях с одним видом активности и с оптимизацией величины биоактивности в гомологических рядах химических соединений. В дальнейшем будет показано, что переход на качественно новый уровень понимания проблемы КССА требует объединения этих подходов в рамках единого системного исследования. Такая постановка задачи подразумевает достаточно полное и корректное описание структуры соединений, включая особенности электронно-топографического и конформационного строения молекул.

III. Корректное описание структуры в рамках использования ЭВМ для установления связи «структура – активность»

Прежде чем приступить к исследованию связи между структурой и активностью, необходимо выбрать способ представления исходных данных [14]. Было затрачено много усилий на подбор доступной экспериментально

или путем расчетов системы параметров, которая достаточно полно описывала бы структуру. В настоящем обзоре рассматриваются четыре основных подхода к описанию структуры соединений.

1. Фрагментарный код суперпозиции подструктур

Фрагментарный код суперпозиции подструктур (ФКСП) был предложен в начале семидесятых годов [32]. При анализе химической структуры на основании этого подхода выделяются потенциально активные центры, названные дескрипторными центрами (ДЦ), к которым отнесены атомы, содержащие подвижные p - и π -электроны или целый электростатический заряд. Это все гетероатомы: азот, кислород, сера, фосфор, галоиды, металлы, а также циклы и пары атомов углеродов, связанные кратными связями. Подробный список ДЦ приведен в [33]. Первоначально для каждого соединения на основании его структурной формулы составлялся набор линейных дескрипторов, каждый из которых представлял собой семизначное число (по две цифры отводились на код дескрипторного центра и длину цепи; признак сопряжения выражался одной цифрой, равной единице при наличии сопряжения, и нулю – в противном случае). Учет специфики циклических фрагментов при кодировании соединений для анализа их сходства проведен в [34]. Свое применение ФКСП нашел в статистических исследованиях информационного массива биологически активных соединений. В дальнейшем был предложен способ кодирования структуры химического соединения, использующий понятие ДЦ, но сохраняющий топологию молекулярной структуры как единого целого в форме матрицы связности дескрипторных центров (МСДЦ) [33]. В этой матрице диагональными элементами являются коды ДЦ, а недиагональными элементами могут быть значения расстояний между соответствующими ДЦ (в случае матрицы геометрии ДЦ) или число связей между ними. Этот подход использован в ряде работ [16, 17] для выявления общих элементов структуры соединений, обладающих конкретным видом биологической активности.

2. Системы топологических и геометрических дескрипторов

Естественным развитием языка описания химических структур является появление систем топологических и геометрических дескрипторов. Среди линейных способов кодировки структур наибольшее распространение получила система Висвессера [35]. Однако в силу ограничений, присущих линейным системам кодирования, о главном из которых уже упоминалось, в последнее время предпочтение отдается кодированию структур с помощью матриц связности с широким набором молекулярных дескрипторов. Это, прежде всего, топологические дескрипторы, включающие в себя: дескрипторы фрагментов, подструктурные дескрипторы, дескрипторы связ-

ности, окружения и т. д. Все они предназначены для более точного описания окружения активных центров молекул [14]. Недостаток этого способа кодировки, так же как и кода ФКСП, – в том, что кодируется не структура химических соединений, а их структурная формула. Некоторого улучшения этого способа описания структуры можно добиться, если наряду с топологическими дескрипторами использовать и геометрические. Последние, описывая общую форму и размер молекул, создают представление о трехмерной структурной диаграмме соединения. Более полному описанию структуры соединений способствуют конформационные расчеты равновесных расстояний между активными центрами методами молекулярной механики [36]. Однако и в этом случае остаются невыясненными электронные свойства активных центров, определяющие их способность к межмолекулярному взаимодействию. Это в существенной мере сказывается на выявлении истинных взаимосвязей между структурой химических соединений и проявляемой ими активностью.

3. Квантовохимические методы, используемые для описания структуры

Квантовохимические расчеты структуры молекул используются в проблеме «структура – свойство» с момента их возникновения и в настоящее время, благодаря прогрессу в области вычислительной техники, получают все большее распространение в исследованиях по установлению связи между структурными параметрами молекул и их свойствами, особенно биологическими [19, 21, 37–39].

Существующие к настоящему моменту программы для ЭВМ охватывают практически весь спектр квантовохимических расчетов, начиная от простых схем полного пренебрежения дифференциальным перекрытием (CNDO) до строгих непараметрических подходов: CNDO/2, CNDO/BV, CNDO/S, INDO, INDO(S), MINDO/2, M1NDO/3, MNDO, VRDDO, *ab initio* STO-39. Однако привлечение квантовохимических расчетов к проблеме ССА сводится в основном к определению наиболее устойчивых конформаций молекул, в которых активные центры доступны для взаимодействия с рецептором, или к оценке молекулярных физико-химических характеристик, связанных с транспортом вещества до биомишени. Перенесение центра «тяжести» квантовохимических расчетов на электронные свойства молекул позволило бы ставить задачу выявления (конструирования) веществ с заданными свойствами среди принципиально новых классов соединений. Такими возможностями не обладают подходы, описанные выше.

Перспективность использования квантовохимических расчетов для проблемы ССА очевидна. И дальнейшее совершенствование вычислительной техники, переход на неэмпирические расчеты [40], безусловно, приведут к возрастанию роли квантовой химии в установлении взаимосвязей «струк-

тура – свойство». В настоящее же время расчет электронного строения больших рядов химических соединений даже при использовании полуэмпирических методов трудно реализуем из-за огромных затрат машинного времени. В связи с этим возникает задача создания такой расчетной схемы, которая была бы достаточно проста и вместе с тем пригодна для корректного описания электронной структуры молекулы в рамках проблемы ССА. В [41] предпринята попытка заполнить матрицу связности дескрипторных центров квантовохимическим содержанием, учитывающим электронное строение молекул.

Следует, однако, отметить, что квантовая химия – далеко не единственный источник информации об электронно-топографическом строении молекул. Более того, в силу перечисленных трудностей, получаемая в результате квантовохимических расчетов информация не всегда достаточно надежна и точна. Поэтому при дальнейшем развитии работ по выявлению взаимосвязей «структура – свойство» наряду с этими расчетами будут совершенствоваться физико-химические подходы к проблеме ССА, прямым образом связанные с экспериментальными исследованиями физико-химических свойств молекул или их фрагментов.

4. Физико-химический подход к проблеме связи «структура – активность»

Начало широкого применения физико-химических характеристик в проблеме ССА связано с работами Хэнча с соавт. [1, 5, 31]. Описан целый набор разнообразных параметров, используемых при расчетах по уравнению Хэнча [18, 42]. Все эти параметры могут быть разделены на три группы: в первую можно включить параметры, связанные преимущественно с гидрофобными свойствами, – логарифм коэффициента распределения ($\lg P$), параметр липофильности (π), параметр тонкослойной хроматографии ($\lg K'$), парахор (Pr), мольный объем (V_m), молекулярный вес (M), логарифм распределения в воде ($\lg C_{aq}$), парциальное давление (p), параметр растворимости Гильдебранда (σ_{π}^{*}), мольная аттракционная постоянная (F_a), энтропийная константа ($\sigma_{\pi}^{*\bullet}$) и др. Во вторую группу входят различные индукционные и резонансные константы (σ^* , σ° , σ^I , σ^R и т. д.), радикальные константы заместителей (E_R), параметры Свайна и Лаптопа (F , R), поляризуемость (α), молекулярная рефракция (MR), дипольный момент (μ), потенциал ионизации (IP), частоты в спектрах (ν), константы ионизации (pK_a , pK_b) и другие характеристики, связанные с электронными свойствами. Наконец, третью группу составляют параметры, характеризующие стерические эффекты внутримолекулярных взаимодействий: стерические параметры Тафта (E_s и E_s°), ван-дер-ваальсовы радиусы (r_v), молекулярные связуемости (χ) и др. Хотя многие из указанных параметров не отражают в чистом виде только гидрофобные, только электронные или только стерические свойства, приве-

денное разделение является весьма полезным, например, в связи с проблемой мультиколлинеарности при выборе параметров модели согласно схеме действия биологически активных веществ.

Несмотря на многочисленные факты успешного применения уравнения Хэнча, следует, однако, подчеркнуть, что этот подход и различные его модификации «работают» только в узких рамках рядов родственных соединений. Замена хотя бы одного активного центра в молекуле приводит к изменению характера внутримолекулярных взаимодействий и, как следствие, к нарушению ранее установленных корреляций между физико-химическими параметрами, описывающими соединение, и проявляемым им свойством. В этом отношении описанные ранее подходы с возможностью отбора общих элементов структуры выглядят предпочтительнее, поскольку допускают изучение широкого спектра соединений. В связи с этим возникает вопрос: нельзя ли подход Хэнча распространить на анализ соединений из разных рядов и классов путем его объединения с методами выделения общих структурных фрагментов? Впервые эта идея в рамках проблемы ССА была реализована в [43], где принимались во внимание следующие обстоятельства. В традиционном дескрипторном анализе сходства структур используется набор ДЦ с их классификацией по типу атомов и характеру связей. Под влиянием внутримолекулярного окружения границы, разделяющие электронные свойства ДЦ, могут стать диффузными, и это повлечет за собой осложнение в разделении активностей. Отсюда становится очевидным, что в основе классификации ДЦ должен лежать не тип атома, а его способность к межмолекулярному взаимодействию.

Количественной мерой указанного взаимодействия может служить энтальпия комплексообразования (ΔH), пропорциональная свободной энергии (ΔG) взаимодействия молекул, содержащих по одному электроно-донорному или акцепторному центру [25]. К настоящему времени накоплен обширный экспериментальный материал значений ΔH , который позволяет предсказывать термодинамические свойства всего многообразия комбинаций пар донор – акцептор, что дает возможность предложить новую классификацию ДЦ. Для этого необходимо иметь либо экспериментальные данные по ΔH для всех возможных донорных молекул с одним и тем же акцептором, либо найти систему эмпирических параметров, позволяющих оценивать ΔH взаимодействия любой пары донор – акцептор. У первого варианта (см., например, донорные числа Гутмана (44)) есть существенные ограничения: недостаточная растворимость некоторых комплексов или возможность протекания необратимой реакции. Гораздо перспективнее второй вариант, для реализации которого широко используются два уравнения, предложенные в [45, 46]. В последней модификации формулы [45] энтальпия комплексообразования выражается через сумму трех попарных произведений шести типов параметров. Гораздо проще формула, которая в [43] взята за основу классификации донорно-акцепторных взаимодействий.

Согласно [46], энтальпия комплексообразования пропорциональна произведению двух безразмерных величин: фактора i -кислоты (акцептора, P_i) и j -основания (донора, E_j)

$$\Delta H_{ij} = \Delta H_{11} P_i E_j, \quad (1)$$

где H_{11} – коэффициент, определяемый по какому-либо стандартному комплексу. В развитие этого подхода в [43] предложена единая шкала донорных и акцепторных факторов, которая строится на основе следующей модификации уравнения (1):

$$\Delta H_{ij} = | \Delta H_{11} | E_i E_j,$$

где $E_i = -P_i$.

В последующих работах [21, 22, 47] этим подходом были охвачены не только нейтральные, но и заряженные доноры и акцепторы. В [48] на основе этой единой шкалы предложена классификация ДЦ для анализа ССА. Список значений E_i для 70 широко используемых акцепторов и значений E_j для 230 разнообразных электронодонорных соединений представлен в [22]. Даже этот относительно небольшой массив данных дает возможность рассчитать значения энтальпий комплексообразования более чем для 16 тыс. комплексов.

Проведенный анализ больших массивов экспериментальных данных по комплексообразованию позволил установить взаимосвязь между значениями $E_i(E_j)$ и константами заместителей, характеризующих индукционные влияния и эффект сопряжения [49, 50]. Такой подход потребовал расчета значений этих констант для сложных функциональных групп. В связи с этим в [52, 53] выведены эмпирические уравнения, описывающие передачу индукционного эффекта через атомы углерода, кислорода и серы. Заключительным этапом внедрения донорно-акцепторных факторов в практику конструирования биологически активных веществ следует считать создание автоматизированной системы оценки значений $E_i(E_j)$. Недавно авторами этого обзора предложен и реализован на мини-ЭВМ NORD-10 алгоритм расчета $E_i'(E_j)$ с автоматическим определением констант Тафта сложных функциональных групп по введенной в ЭВМ структурной формуле и описанным типам атомов. Тем самым создано математическое обеспечение для реализации физико-химического подхода к описанию и классификации соединений по электронным свойствам дескрипторных центров, используемых для отбора элементов структуры, ответственных за конкретный вид биологической активности. Этот подход отличается ясностью и содержательностью легко интерпретируемых принципов классификации, достаточной для широкого использования простотой и уже

используется, по крайней мере, на качественном уровне, при конструировании веществ с заданными свойствами [54].

Наряду с корректной оценкой электронных свойств ДЦ в проблеме ССА не менее важным является точное описание пространственной структуры молекул с учетом динамики внутримолекулярных движений: таутомерных и конформационных переходов, колебаний большой и малой амплитуд. И здесь можно сразу отметить, что включение в описанную ранее матрицу связности ДЦ в качестве недиагональных элементов значений числа связей между соответствующими ДЦ явно не соответствует современным представлениям о пространственной структуре молекул. Включение в матрицу связности равновесных расстояний [17, 33], использование в качестве дескрипторов моментов инерции и молекулярных объемов [21], конечно же, улучшают ситуацию, но не решают проблемы корректного описания пространственной структуры молекул. В качестве простого способа описания молекулярной структуры, довольно полно удовлетворяющего требованиям «идеального», в [55] предложено использовать представление пространственного строения молекулы в виде кривой радиального распределения, как это принято в газовой электронографии. Можно отметить интересную попытку использования аналогичного преобразования молекулярной структуры для разделения транквилизаторов и седативных средств [56], что оказалось намного эффективнее, чем использование на этом же материале фрагментарных дескрипторов [57].

Описанные способы представления структуры молекул в достаточной степени отражают строение соединений в рамках использования ЭВМ для установления связи «структура – активность»¹. Причем выбор формы описания структуры молекул тесно связан с конкретной задачей исследования (классификация соединений по типу проявляемой биоактивности, информационный поиск родственных соединений, выделение общих элементов структуры, количественный прогноз о величине биоактивности и т. д.), а также с используемым для этой цели методом. С этой точки зрения, очень важным является вопрос об оптимальном соответствии формы описания структуры и выбранного метода обработки структурной информации.

IV. Методы выявления связи «структура – активность»

Существующие многочисленные методы выявления связи «структура – активность» по типу решаемых задач можно разделить на две большие группы [21]: методы, направленные на оптимизацию свойств в ряду родственных соединений, и методы, ориентированные на поиск общих структурных фрагментов в молекулах соединений, проявляющих определенный вид биоактивности.

¹ Raevsky O.A. // QSAR in Drug Des. Toxicology / Eds Hadzi D. et al. Amsterdam: Elsevier, 1987. P. 31.

1. Оптимизация свойств в ряду родственных соединений

В основе этой группы методов лежат различные многопараметрические регрессионные модели, связывающие структурные параметры соединений с их биоактивностью [58–62]. В таблице 1 представлены основные подходы к обработке структурной информации, использующиеся в ССА для оптимизации свойств в ряду родственных соединений.

Успешное применение описанной в предыдущей главе модели Хэнча способствовало появлению ее многочисленных обобщений и модификаций, представляющих своеобразный «скрининг» разнообразных наборов физико-химических и структурных параметров и типов регрессионных моделей. Уже к семидесятым годам текущего столетия было выведено более 2 тысяч регрессионных зависимостей для более чем 20 тысяч химических соединений [63], связывающих гидрофобные, электронные, стерические и другие свойства заместителей с проявляемой веществом биологической активностью. Некоторые из этих соотношений были использованы как для прогнозирования эффективности [64–66], так и для выявления механизма действия биологически активных соединений [18, 67]. Однако немало среди них, по выражению Хэнча [68], и «удивительных уродцев». К ним относятся уравнения, нереальные по физическому смыслу или неэффективные для прогноза величины биоактивности, но обладающие высокими значениями коэффициентов корреляции.

Табл. 1. Основные методы оптимизации свойств в ряду родственных соединений

Вид анализа	Принцип обработки информации	Область применения	Ссылки
Многомерный статистический	Расчет статистических характеристик распределений случайных величин – средних, дисперсий и т. д.	Исключение ошибочных данных и анализ «выбросов»	[72–76]
Многомерное шкалирование	Сокращение размерности многомерного пространства с целью исключения малоинформативных данных и переменных	Отбор соединений с наибольшей информативностью	[73–75]
Корреляционный	Выявление связей между переменными на основе корреляционной матрицы	Проверка гипотезы о независимости параметров модели	[78, 1191]
Факторный	Сжатие информации путем введения новых переменных, являющихся линейными комбинациями исходных	Выявление «структуры» данных – сгруппированности объектов и параметров модели	[79–82]
Регрессионный и дисперсионный	Построение функциональной зависимости между независимыми переменными и откликом (зависимой переменной)	Выявление количественной связи между параметрами структуры и проявляемой биоактивностью соединений	[31,61]–[85–87]

Существование таких уравнений показывает, что высокое значение коэффициента корреляции R , квадрат которого численно равен доле дисперсии функции биоотклика ($\sigma_y^2 - \sigma^2$), «объясненной» регрессионной зависимостью

$$R^2 = (\sigma_y^2 - \sigma^2) / \sigma_y^2,$$

где σ_y – дисперсия функции биоотклика, σ – остаточная дисперсия, является необходимым, но не достаточным условием хорошего качества регрессионной модели. Это объясняется тем, что для регрессионного анализа существует ряд ограничений, нарушение которых приводит к «бедам» регрессионных коэффициентов [69], характеризующих вклад каждого из параметров структуры соединения в величину биоактивности.

Так, в работе [70] обращается внимание на использование без должного обоснования предположения модели Хэнча о нормальном законе распределения для переменных, входящих в эту модель. Анализ 14 возможных законов распределения для переменных в модели Хэнча показал, что независимо от вида характеристики гидрофобности биологически активного соединения наилучшей моделью, описывающей транспорт вещества до биомишени, является показательно-степенной закон распределения величины параметра гидрофобности с явно выраженной асимметричностью. Структура этого распределения свидетельствует о том, что вероятность достижения субстратом места действия возрастает пропорционально его гидрофобности согласно степенному закону, а убывает – по экспоненциальному закону. Наличие асимметрии указывает, что случайные события в данной системе не всегда являются взаимно-независимыми. Наличие корреляций между «независимыми» параметрами является одной из наиболее распространенных причин плохой прогностической способности выведенных соотношений. Появление неэффективных для прогноза КССА моделей, помимо особенностей «техники» регрессионного анализа, связано и с рядом принципиальных трудностей: а) в установлении однотипности механизма взаимодействия исследуемого набора соединений с одними и теми же рецепторами [1]; б) в выборе регрессионной модели, отвечающей механизму действия вещества; в) в выборе переменных, адекватно описывающих анализируемое свойство; г) в использовании данных по биоактивности, полученных разными исследователями (проблема точности и воспроизводимости биологического эксперимента); д) в выявлении зависимости результирующего биологического действия вещества, влияющего на многие физиологические процессы, происходящие в организме.

Перечисленные трудности, а также то обстоятельство, что в настоящее время не существует альтернативы регрессионному анализу как средству получения количественного прогноза величины биоактивности, приводит к мысли о построении регрессионных моделей после предварительного качественного анализа имеющихся данных. В этом случае последовательность анализа КССА могла бы выглядеть следующим образом. На первом этапе предварительного или, как его еще называют

«разведочного»², анализа устанавливается общая структура данных. Для этой цели могут быть использованы как методы наглядного представления многомерного пространства данных, так и автоматические методы классификации и статистической обработки информации. На этом этапе в рамках проблемы «структура – активность» можно решить две задачи: выделить классы родственных соединений и осуществить переход к набору существенно меньшего числа информативных параметров, описывающих структуру соединений. На следующем этапе может быть проверена гипотеза о независимости параметров модели с помощью методов корреляционного анализа [78]. При необходимости, с использованием методов факторного анализа, в регрессионную модель могут быть введены факторы, представляющие линейную комбинацию исходных параметров модели [79–82]. При формировании окончательной регрессионной модели в [18] рекомендуют придерживаться следующих правил: 1) рассматривать на этапе предварительного анализа как можно больше независимых параметров в качестве переменных моделей; 2) выбирать из различных комбинаций переменных те, которые содержат параметры, относящиеся к разным типам взаимодействия (например, $\lg P$ и σ [83]); 3) отбрасывать переменные с низким уровнем значимости регрессионных коэффициентов; 4) при близких значениях коэффициентов корреляции выбирать модель с наименьшим стандартным отклонением; 5) для исключения случайных корреляций включать, по крайней мере, 5–6 результатов биологических испытаний на каждый испытываемый в модели параметр [84]; 6) отдавать предпочтение, при прочих равных условиях, более простым моделям; 7) с целью сокращения и упрощения расчетов, а также во избежание принципиальных ошибок учитывать реальность существования модели с точки зрения рассматриваемого явления.

К этому следует добавить и учет возможностей современного регрессионного анализа по оценке качества полученных регрессионных зависимостей [85–87]. Корректное построение регрессионной зависимости позволяет добиться высокой прогностической способности модели, описывающей действие биологически активного вещества, и тем самым решить задачу оптимизации свойств в ряду родственных соединений. Однако остается открытым вопрос о прогнозе биоактивности соединений с различающейся структурой. Решение этой проблемы связано с поиском общих элементов структуры.

2. Поиск общих элементов структуры биологически активных соединений

В основе поиска общих элементов структуры соединений лежит предположение, что биологическая активность веществ обусловлена наличием в

² Velleman P.F., Hoaglin D.C. Applications, Basics and Computing of Exploratory Data Analysis. Boston: Duxbury Press, 1981. 354 p.

структуре определенных молекулярных подструктур – фармакофоров. Типичным, например, для этого подхода является использование методов распознавания образов (МРО) [87–89] для разделения соединений на активные и неактивные по набору структурных параметров.

Табл. 2. Основные методы выявления общих элементов структуры и классификации соединений

Метод анализа	Принцип обработки информации	Область применения	Ссылки
Дискриминантный	Построение разделяющих гиперповерхностей по обучающей выборке, состоящей из объектов двух классов	Отбор информативных параметров модели описания структуры и классификация соединений на активные и неактивные	[14, 77, 97–101]
Линейная обучающаяся машина	Непараметрический дискриминантный анализ, основанный на исправлении ошибочной классификации через обратную связь в итерационном процессе	То же	[14, 88]
Логико-структурный	Метод дискриминантного анализа, основанный на использовании законов индуктивной логики		[17]
Кластерный	Классификация объектов по признакам сходства или различия с помощью оптимизации целевой функции	Классификация соединений по видам активности	[102–107]
Моделирование пространственной структуры и свойств молекул	Расчет и построение объемных изображений молекул методами квантовой химии и молекулярной механики с помощью молекулярной графики	Выявление стерической специфичности взаимодействия субстрат – рецептор	[110–116, 119]

В этом плане модель Фри – Вильсона [90], основанная на аддитивном вкладе подструктур заместителей в значение величины биоактивности, занимает промежуточное положение между количественными регрессионными моделями и качественными моделями методов распознавания образов. Эта простая модель позволяет существенно сузить область поиска эффективных БАВ. Так, для испытания соединений, где заместители могут находиться в четырех положениях и при изменении их числа в пределах от

двух до пяти, применение метода Фри – Вильсона сокращает число необходимых биоиспытаний со 120 до 11 [18]. Однако из-за использования информации на качественном уровне наличия или отсутствия определенных типов фармакофоров в структурной формуле соединения количественный прогноз величины биоактивности часто оказывается ненадежным. Поэтому предпринимались попытки улучшения этой модели, увязывая ее с механизмом действия веществ [91–94] или вводя в рассмотрение взаимное влияние заместителей [95]. Однако в последнем случае число переменных в модели резко возрастает, и модель становится труднореализуемой.

Для объектов, описываемых большим числом параметров, для поиска общих элементов структуры более эффективными оказываются методы распознавания образов, основанные на автоматической классификации объектов [96] по сходству или различию. Среди них наиболее широко в литературе по проблеме «структура – активность» представлены методы, относящиеся к трем видам анализа (таблица 2).

Дискриминантный анализ [97–101] основан на разделении соединений на активные и неактивные путем построения разделяющей гиперплоскости или другой поверхности в пространстве параметров соединений, описывающих их структуру. В рамках этого анализа проводится отбор наиболее информативных, с точки зрения разделения соединений, параметров, строится разделяющая поверхность и проверяется эффективность реализованной дискриминантной модели на обучающей и экзаменационной выборках соединений. Примером успешного применения дискриминантного анализа могут служить работы [99–101]. •

Логико-структурный анализ [17] использует алгоритмы методов дискриминантного анализа с представлением информации в виде, естественном для исследователя. Многие процедуры обработки информации ориентированы на диалоговый режим работы ЭВМ.

Кластерный анализ [102] состоит в группировке объектов как сгущений точек в многомерном пространстве, описывающих эти объекты, параметров. Этот вид анализа тесно связан с методами многомерного шкалирования (отображением точек многомерного пространства на поверхность меньшей размерности) и численной таксономии (разбиение объектов на заданные классы). Отличительной чертой методов кластерного анализа является отсутствие пустых кластеров. Методы кластерного анализа в проблеме «структура – активность» могут использоваться как самостоятельно [104–106], так и в сочетании с другими методами [77, 107, 108].

Во всех рассмотренных методах автоматической классификации топологические или топографические особенности строения молекул отображались в виде набора индексированных параметров – геометрических дескрипторов, являющихся координатами соединений, как точек в многомерном пространстве параметров. С развитием компьютерной графики в последние годы появилась возможность наглядного трехмерного изображения самих

структур молекул на экране графического дисплея ЭВМ [108–110]. Это положило начало еще одному направлению в исследованиях по ССА – молекулярному моделированию пространственной структуры соединений.

Наиболее плодотворным оказалось использование методов моделирования трехмерной структуры молекул для изучения пространственной специфичности взаимодействия субстрат – рецептор по принципу «ключ – замок» [110], особенно при представлении электронной «формы» молекулы как поверхности вокруг ядерного скелета конформации молекулы [27], полученной с помощью квантовохимического расчета [111] или метода молекулярной механики [36]. Этот подход позволяет обнаружить сходство в электронно-топографических свойствах молекул, не очевидное при других методах анализа. Количественная сторона учета стерической специфичности взаимодействия «субстрат – рецептор» разрабатывалась в регрессионном анализе [112–115], однако в последнее время наметилась тенденция к совместному использованию методов молекулярного моделирования и регрессионного анализа [115, 116], которая в перспективе должна привести к системному подходу к проблеме связи «структура – активность».

3. Системный подход к анализу связи «структура – активность»

Идея системного подхода к анализу ССА состоит в создании единой модели поиска биологически активных соединений, охватывающей как качественную и количественную, так и априорную (экспертную) информацию об изучаемом классе соединений в рамках экспертной системы, включающей соответствующий набор методов выявления ССА. Целесообразность этого продиктована как необходимостью учета всей имеющейся у исследователя информации, так и выбором оптимального метода для ее обработки. Действительно, если, например, сформулировать проблему связывания субстратом рецептора как задачу нахождения функции плотности распределения вероятности комплексообразования по набору параметров, характеризующих структуру субстрата, то станет очевидной связь этой проблемы с задачей восстановления зависимостей в условиях обратной некорректной задачи [87]. Из теории решения подобных задач [117] можно заключить, что качество полученного результата в значительной степени зависит от полноты учета специфики конкретной задачи, выбора соответствующего метода ее решения. Поскольку на разных уровнях и стадиях анализа ССА исследователь располагает информацией, отличающейся как по количеству и качеству, так и по форме ее представления и информативности, становится понятной необходимость использования не какого-то одного метода выявления ССА, а всего комплекса этих методов, причем в такой их последовательности, которая в конечном счете привела бы к количественному прогнозу величины биологической активности неизученного соединения по параметрам, характеризующим его структуру.

Первым шагом в направлении объединения качественных и количественных методов анализа ССА в рамках единой модели стала разработка физико-химического подхода, подробно описанного в главе III. Возможность системного анализа связи «структура – активность» в этом подходе связана с унификацией описания электронно-топографических свойств молекул на базе единой донорно-акцепторной модели взаимодействия «субстрат – «рецептор», где параметры модели могут выступать как индикаторные переменные на этапе автоматической классификации соединений и как количественные переменные при восстановлении регрессионных зависимостей «структура – активность». Это позволяет ставить вопрос об оптимизации свойств в ряду соединений с отличающимися структурными формулами, в том числе принадлежащих разным классам, но имеющих близкие донорно-акцепторные и топографические свойства дескрипторных центров.

Современная тенденция к интеграции методов и подходов в рамках проблемы ССА нашла свое отражение и в разрабатываемых в настоящее время комплексах программ.

V. Комплексы программ по выявлению связи «структура – активность»

Рассмотренные методы и подходы к анализу ССА для своей реализации требуют создания соответствующих программных комплексов. Существующие комплексы (пакеты) программ различаются как по назначению и степени охвата решаемой проблемы, так и по ориентации на определенный класс ЭВМ. Наиболее удобными являются пакеты программ, организованные по модульному принципу с диалоговым режимом работы. Как правило, такие комплексы ориентированы на мини-ЭВМ и, благодаря легкости включения новых программ в комплекс, обладают несомненным преимуществом перед другими пакетами программ при создании экспертной системы по проблеме ССА на базе лабораторной мини-ЭВМ.

В качестве примера рассмотрим четыре таких комплекса, в которых реализованы описанные ранее подходы к анализу ССА: автоматической классификации на основе методов распознавания образов (ADAPT), логико-структурный (ТОПЛОГ), молекулярного моделирования (MOLY) и физико-химический (РАСТР) подходы.

В пакете программ ADAPT [14] основной обрабатываемой информацией являются структурные формулы соединений. Любая из них может быть введена в ЭВМ путем построения ее изображения на экране графического дисплея с помощью специального набора команд в интерактивном режиме. Введенные структурные формулы хранятся в виде матриц связности на дисковых накопителях. Предусмотрены возможности для добавления, исключения, изменения и вызова хранимых структур соединений. Анализ ССА проводится на основе методов параметрической и непараметрической ста-

тики и распознавания образов: линейной регрессии, дискриминантного анализа, линейной обучающейся машины и классификации по количеству ближайших соседей.

Система ADAPT реализована на базе 16-битовой мини-ЭВМ MODCOMP/25 и вместе с банком данных занимает около 2 Мбайт памяти на магнитных дисках.

Программный комплекс ТОПЛОГ [17] представляет пользователю более развитую систему ввода, редактирования и преобразования информации к виду, естественному для восприятия исследователем. Это нашло свое отражение в файловой системе комплекса, насчитывающего 14 типов файлов различного назначения. В комплексе реализованы процедуры логико-структурного подхода, подробно описанного в [17]. Отличительной особенностью пакета ТОПЛОГ является наличие алгоритма отбора общих структурных фрагментов молекул на основе анализа граф-представлений структурных формул исследуемых соединений. Пакет состоит из 14 программ общим объемом около 8 тыс. фортрановских операторов и используется на мини-ЭВМНР-1000.

В отличие от описанных комплексов, где основным видом информации является структурная формула соединений, в комплексе MOLY [10, 118] для конструирования БАС используется объемное трехмерное изображение молекулы с отображением ее свойств (например, поверхности электростатического потенциала). Это позволяет с помощью методов молекулярного моделирования получать параметры структуры молекулы, отражающие пространственную специфичность взаимодействия субстратов с рецепторами, которые могут быть использованы для построения регрессионных моделей КССА. Комплекс MOLY содержит около 15 тыс. фортрановских операторов и занимает объем памяти, равный 1 Мбайту.

В комплекс программ РАСТР (распознавание активных структур)³ включены процедуры, реализующие физико-химический подход к выявлению общих по электронно-топографическим свойствам фрагментов БАВ в рамках проблемы «структура – активность». Отличительной чертой комплекса является возможность системного подхода к анализу ССА с использованием на разных этапах исследования всего набора рассмотренных ранее методов (таблицы 1 и 2). Это достигается унифицированным описанием электронно-топографических свойств молекул с позиции единой донорно-акцепторной модели взаимодействия «субстрат – рецептор». Так, например, одним из вариантов работы с пакетом РАСТР может быть следующий. С терминала ЭВМ вводятся структурная формула соединения и информация о типе входящих в нее атомов. С помощью программы INPUT (рисунок 1) эти данные преобразуются в матрицу связности и записываются в файл соединений

³ Раевский О.А., Сапегин А.М. // Хим.-фармацевт. журн. 11987. Т. 21. С. 1338; Сапегин А.М., Раздольский А.Н., Чистяков В.В., Раевский О.А. // Там же. С. 1341.

(FC) на магнитном диске. Информация из этого файла используется для расчета матрицы геометрии дескрипторных центров (МГДЦ) – физико-химического «портрета».

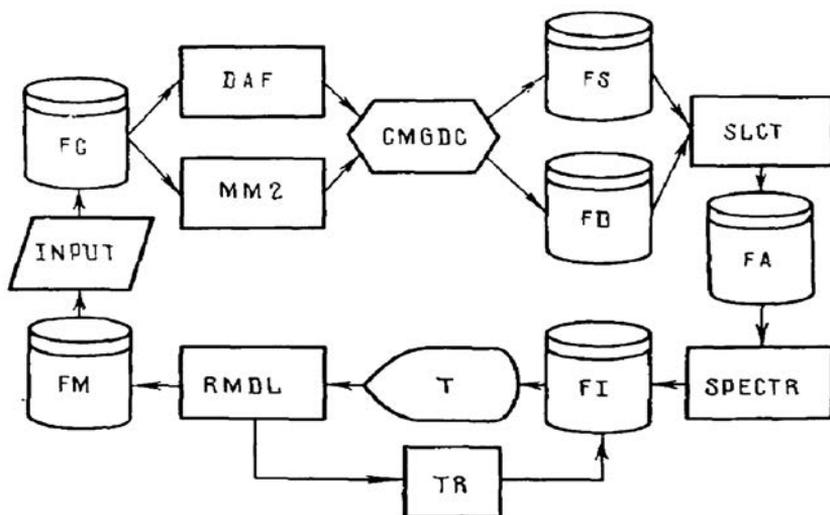


Рис. 1. Блок-схема основного ядра комплекса РАСТР

Программы: INPUT – ввода, DAF – расчета донорно-акцепторных факторов, MM2 – расчета равновесных конформаций молекул методом молекулярной механики, CMGDC – формирования МГДЦ, SLCT – отбора общих структурных фрагментов, SPECTR – построения спектра межатомных расстояний, RMDL – построения регрессионных моделей, TR – преобразования регрессионных моделей фармакофоров в спектр межатомных расстояний; файлы: FC – соединений, FS – структур, FD – структур банка данных, FA – активностей (фармакофоров), FI – изображений (спектров межатомных расстояний), FM – регрессионных моделей молекулы, описывающего донорно-акцепторные и топографические свойства активных центров субстратов. Диагональные элементы МГДЦ кодируются на основе рассчитанных значений донорно-акцепторных факторов (ДАФ) ДЦ в соответствии с унифицированной шкалой донорно-акцепторных взаимодействий. Алгоритм расчета ДАФ ДЦ (программа DAF) построен на анализе дерева граф-представления структурной формулы соединения с использованием банка данных по константам Тафта для различных типов атомов и корреляционных уравнений для расчета констант Тафта сложных заместителей и ДАФ ДЦ [50, 52–53]. Внедиагональные элементы МГДЦ содержат информацию о топографии ДЦ в виде расстояний между соответствующими ДЦ для равновесных конформаций, рассчитанных методом молекулярной механики [36] (программа MM2). Матрица геометрии ДЦ формируется программой CMGDC, которая работает в двух режимах: режиме анализа ССА и режиме записи МГДЦ в банк данных.

С помощью программы SLCT отбираются общие структурные фрагменты для исследуемого (файл FS) и хранимых в банке данных (файл FD) соединений. Процедура поиска общих структурных фрагментов аналогична используемой в комплексе ТОПЛОГ [24]. Отличие состоит лишь в том, что в данном случае под общим структурным фрагментом понимается не строгая тождественность элементов структурной формулы соединений, а их регулируемая исследователем близость по электронно-топографическим свойствам. Выделенные фрагменты являются потенциальными фармакофорами и записываются в файл биологически активных структур (файл FA).

Содержащиеся в файле FA МГДЦ общих структурных фрагментов БАС могут быть представлены (с помощью программы SPECTR) в виде спектра межатомных расстояний, удобного для визуального сравнения структур соединений, или набора структурных параметров для построения регрессионных моделей ССА. Спектры межатомных расстояний хранятся в файле «изображений» (FI).

Таким образом, на описанном выше этапе, решается качественная задача по выявлению потенциальных фармакофоров, т. е. на основе индикаторных параметров донорно-акцепторной модели формируется представление об электронных и пространственных свойствах активных центров субстратов, ответственных за конкретный вид биологической активности.

Следующим важным шагом является установление количественной зависимости между параметрами структуры соединения и величиной биоотклика. В рассматриваемом комплексе РАСТР исходной информацией для построения регрессионных моделей (программа RMDL) является информация из файла FI. Поиск оптимальной регрессионной зависимости осуществляется на основе принципа структурной минимизации среднего эмпирического риска [87]. Оценка прогностической способности модели проводится по экзаменационной выборке данных.

Полученные на этапе регрессионного анализа результаты могут быть повторно использованы в комплексе РАСТР для уточнения модели или построения новой. Эти данные хранятся в файле FM. Благодаря файловой организации передачи информации между программами комплекса он обладает большой гибкостью по отношению к включению стандартных или оригинальных методов обработки информации на любом этапе исследования и может аккумулировать в себе все виды анализа КССА, став основой для построения развитой экспертной системы.

Все программы комплекса написаны на языке ФОРТРАН-77 и используются на мини-ЭВМ NORD-10 (емкость ОЗУ 128 Кбайт).

Рассмотренные комплексы программ показывают, что к настоящему времени программное обеспечение охватывает широкий круг алгоритмов обработки структурной информации, создавая тем самым базу для системного подхода к анализу ССА. Практической реализацией этой идеи могла бы стать разработка проблемно-ориентированной экспертной системы по конструированию веществ с заданными свойствами.

VI. Заключение

Для поиска эффективных биологически активных синтетических соединений на базе основополагающих принципов конструирования веществ с заданными свойствами необходимо решение следующих задач:

- идентификация соединения, взаимодействующего с рецептором;
- установление электронной и пространственной структуры этого соединения;
- корректное описание его структуры в рамках использования ЭВМ;
- получение статистически достоверных данных по биологической активности соединений;
- выбор оптимального метода или группы методов для решения конкретной задачи по выявлению связи «структура – активность».

Некорректный подход к решению любой из указанных задач может мешать установлению «истинной» взаимосвязи «структура – активность» или даже привести к установлению ложных закономерностей. Так, например, отождествление субстрата с введенным в организм веществом (т. е. пренебрежение возможностью химических и биохимических трансформаций вещества в процессе транспорта от места введения до биомишени) подчас затрудняет выявление общих фрагментов структуры среди соединений, проявляющих один и тот же вид биоактивности. Загрубление модели описания электронной и пространственной структуры соединений автоматически скажется на прогностической способности такой модели. Эффективность же последней будет определяться также выбором метода обработки информации, т. е. его соответствия форме представления и уровню информации о структуре и свойствах молекул.

В литературе уже имеются многочисленные примеры правильного предсказания биологической активности соединений (таблица 3) [119]. Однако подавляющее большинство этих работ связано с оптимизацией свойств в ряду родственных соединений. Качественно же новый уровень в проблеме КССА, способный привести к выявлению принципиально новых классов эффективных препаратов, возможен лишь на основе использования методов отбора общих структурных фрагментов среди разнообразных соединений с последующей оптимизацией проявляемых ими полезных свойств.

Это приводит к необходимости развития системного подхода к проблеме ССА, который должен включать в себя, помимо рассмотренных ранее вопросов, обширный раздел химической информатики [120] и искусственного интеллекта [121]. Завершающим этапом системного исследования ССА могло бы стать планирование синтеза и технологических схем получения новых препаратов [122, 123]. Реализация такого подхода, в конечном счете, должна превратить эмпирический путь установления связи «структура – активность» в фундаментальное научное направление, аккумулирующее

в себе достижения химии, физики, биологии и информатики. Тем самым будет создана мощная платформа экономически эффективного конструирования веществ с заданными свойствами для нужд промышленности, сельского хозяйства и медицины.

* * *

За время подготовки обзора к печати появилось значительное число работ, затрагивающих практически все рассмотренные аспекты проблемы целенаправленного поиска и конструирования эффективных БАВ. В первую очередь следует отметить публикацию трудов Шестого европейского симпозиума по КССА [124], а также монографию [125], в которой изложены основы построения систем определения основных видов биоактивности для больших массивов химических соединений.

Особенно интенсивно развивается направление, связанное с трехмерным представлением молекулярной структуры методами молекулярной графики [126–131], основанным на эмпирических [132, 133] и квантовохимических [134, 135] расчетах конформаций молекул или соответствующих рентгеноструктурных данных [136, 137]. Много внимания уделяется совершенствованию моделей описания структуры с помощью наборов дескрипторов [138–149] и наглядному отображению результатов статистического анализа структурной информации [150–153].

К числу принципиально новых тенденций в проблеме «структура – активность», обозначившимся в последнее время, следует отнести углубление физико-химической содержательности моделей КССА [154–157], включение в них, наряду с фармакодинамической, фармакокинетической стадии [158–164] действия БАВ и революционизирующее влияние технологии программирования в экспертных системах [165–171].

Создание экспертной системы КССА могло бы стать основой для формирования своеобразной «программной» теории целенаправленного поиска и конструирования биологически активных веществ.

Табл. 3. Примеры правильных предсказаний связи «структура – активность» [119]

Тип соединения	Вид биоактивности
Бензотиадиазины	Гипотензивное действие
Пропиниламины	Ингибирование моноаминоксидазы
N- (феноксиптил) -циклопропил-амины	То же
β-Карболины	»
Сульфонамиды	Ингибирование карбоангидразы
Карбаноилпиперидины	Ингибирование холинэстеразы

Продолжение таблицы 3

Кетоны	Ингибирование хинотрипсина
Бензилпиридиниевые ионы	Ингибирование комплемента
Стероиды	Связывание с прогестероновым рецептором
Эфиры	Гидролиз папаина
Допаминовые эфиры	Гидролиз ариламидаз
Пептидил-п-нитроанилиды	Субтилизиновый гидролиз
Полициклические ароматические углеводороды	Гидроксילирование оксидазы микросомальной
Изатип-β-изотиосеникарбазоны	Антивирусная канцерогенность
Пиразолы	То же
Производные цитидина	Цитостатическая, иммунодепрессивная
Нитрозоуретаны	Цитостатическая
Митоницины	»
Производные гидантоина	»
Рафинициновые амиды	Антибактериальная
Сульфонамиды	»
Эритромицины	
Хинохалин-1,4-диоксиды	»
Фепантрены	Антимикоплазматическая
Фениламидиноуретаны	»
Нафтохиноны	»
Туберины	Антимикоплазматическая
Метоксихлорины	Инсектицидная
N-гидроксипиридоны	Фунгицидная
Производные пиридина	»
Трифторметансульфонамиды	Гербицидная
Алкиларилкарбонильные соединения	»
Ароматические эфиры	»
Пиперидиноацетанилиды	Фитоактивность
Пиримидоны	Ингибирование реакции Хилла
Уретаны	Гербицидная

Фенилацетамиды	Гипнотическая
Пиридилметаны	Спазмолитическая
Проназины	Нейролептическая
Ароматические и алифатические	Противовоспалительная
Кислоты	
Производные бензотиепина	Нейролептическая
Карбоксилатные эфиры	Противотромбическая
Акрилаты	Токсичность
Медные хелаты	Цитотоксичность
Пираненамины	Иммунодепрессивная
Азапурин-6-оны	»
Фенилоксазолидины	Радиозащитная
Карбоноилэтиламины	Холинолитическая
Триазины	Ингибирование дегидрофолатредуктазы
Бенздиазепины	Антипентилентетразольная активность
Цианоциклогексиламины	Фунгицидная
Пиперидиноацетанилиды	Антисеротининовая
Аналоги соматостатина	Антидиабетическая
Стероиды	Прогестероновая
Азациклоалканы	Альгетическая
Органокремниевые амины	Фунгицидная
Индадионы-1,3	Токсичность
Сульфоны	Антибактериальная
Алкобензилалкиламины	Токсичность
Спиробарбитуровые кислоты	Седативная
Аналоги аденозина	Гипотензивная
Аналоги триметоприма	Антибактериальная

ЛИТЕРАТУРА

1. Хэнч А. // Хим.-фармацевт, журн. 1980. Т. 14. № 10. С. 15.
2. Glasser A.S. // Meth. Find.: Exptl. Clin. Pharmacol. 1984. V. 6. N 10. P. 563.
3. Балобанов А.Е. // Развитие количественных представлений о связи между структурой и биологической активностью химических соединений. М.: Ин-т истории естествознания и техники, 1984. Рук. деп. ВИНТИ № 1639–84.
4. Blake J. // Arch. Gen. Med. III Ser. 1839. V. 6. P. 289.
5. Hansch C. // Drug Metabolism Rev. 1984–1985. V. 15. N 7. P. 1270.
6. Richardson B.W. // Brit. Assoc. Rep. 1864. V. 34. P. 120; 1867. V. 37. P. 47; 1868. V. 38. P. 179.
7. Crum-Brown R., Fraser T. // Trans. Roy. Soc. 1869. V. 25. P. 151; 1869. V. 25. P. 693.
8. Ariens E.J., Simonis A.M. // Top. Curr. Chem. 1974. V. 52. P. 1.
9. Ландау М.А. Молекулярные механизмы действия физиологически активных соединений. М.: Наука, 1981.
10. Computer-Assisted Drug Design / Eds Olson E. S. et al. Washington: Amer. Chem. Soc, 1979.
11. Strategy in Drug Design / Ed. Buisman I. A. K. Amsterdam: Elsevier, 1982.
12. Quantitative Structure-Activity Relationships of Drugs / Ed. Topliss J. G. N. Y.: Acad. Press, 1983.
13. QSAR and Strategies in Design of Bioactive Comopunds / Ed. Seydel J. K., N. Y.: UCH Publ., 1985. 442 p.
14. Стьюпер Э., Брюггер У., Джурс П. // Машинный анализ связи химической структуры биологической активности / Пер. с англ. М.: Мир, 1982. 235 с.
15. Drug Design / Ed. Arines E. J. N. Y.: Acad. Press, 1979. 420 p.
16. Голендер В.Е., Розенблит А.Е. // Вычислительные методы конструирования лекарств. Рига: Зинатне, 1978. 232 с.
17. Розенблит А.Б., Голендер В.Е. // Логико-комбинаторные методы в конструировании лекарств. Рига: Зинатне, 1983. 352 с.
18. Нижний С.В., Эштейн Н.А. // Успехи химии. 1978. Т. 47. № 4. С. 739.
19. Овчинников А.А., Болдырев А.Я. // Успехи химии. 1986. Т. 55. С. 539.
20. Валуева Л.Н., Зацепин В.М., Промоненков В.К. // Применение математических методов для анализа связи молекулярная структура — пестицидная активность. Сер. Химические средства защиты растений. М.: НИИТЭХИМ, 1985. 56 с.
21. Раевский О.А. // Введение в конструирование биологически активных веществ. М.: МХТИ, 1984. 80 с.
22. Раевский О.А., Григорьев В.Ю. // Задачи по курсу «Конструирование биологически активных веществ». М.: МХТИ, 1986. 80 с.
23. Simon Z. // Angew. Chem. Int. Ed. 1974. V. 13. № 11. P. 719.
24. Hayward R.C. // Chem. Soc. Rev. 1983. V. 12. № 3. P. 285.
25. Раевский О.А. // Теорет. эксперим. химия. 1986. № 4. С. 450.
26. Крицман В.А. // Журн. Всесоюз. хим. о-ва им. Д.И. Менделеева. 1975. Т. 20. №6. С. 612.

27. *Lehn J.M.* // Recherche. 1981. N 127. P. 1213.
28. *Kellog R.M.* // Recherche. 1984. N 156. P. 819.
29. *Fischer E.* // Chem. Ber. 1894. B. 27. N 3. S. 2985.
30. *Берцов Т.Т., Коровкин Б.Ф.* // Биологическая химия. М.: Медицина, 1983. С. 137.
31. *Hansch C., Fujita T.* // J. Amer. Chem. Soc. 1964. V. 86. P. 1616.
32. *Авидон В.В., Лексина Л.А.* // НТИ. Сер. 2. 1974. № 3. С. 22.
33. *Гитлина Л.С., Голендер В.Е., Дрбоглав В.В. и др.* // Методы представления и обработки структурной информации для анализа связи структура — активность. Препринт. Рига: Ин-т орг. синтеза АН ЛатвССР, 1981. 74 с.
34. *Авидон В.В., Козлова С.П., Аролович В.С.* // НТИ. Сер. 2. 1974. № 12. С. 21.
35. *Wiswesser W.J.* // A Line-Formula Notation. N. Y.: T. J. Cro. Co, 1954.
36. *Буркет У., Эллинджер Н.* // Молекулярная механика / Пер. с англ. М.: Мир, 1986. 364 с.
37. *Kier L.B., Hall L. H.* // Molecular Connectivity in Chemistry and Design. N. Y.: Acad. Press, 1976. 257 p.
38. *Kaufman J.J.* // Int. Quant. Chem. 1979. V. 26. P. 221.
39. *Лужков В.Б., Богданов Г.Н.* // Успехи химии. 1986. Т. 55. № 1. С. 3.
40. *Rolleit J.S.* // J. Mol. Struct. 1985. V. 130. P. 1.
41. *Димогло А.С.* // Хим.-фармацевт. журн. 1985. № 3. С. 438.
42. *Motos I.* // Lecture Notes in Chemistry. 1980. V. 15. P. 2.
43. *Раевский О.А., Новиков В.Я.* // Хим.-фармацевт. журн. 1982. № 5. С. 583.
44. *Бургер К.* // Сольватация, ионные реакции и комплексообразование в неводных растворах / Пер. с англ. М.: Мир, 1984. 256 с.
45. *Kroeger M.K., Drago R.S.* // J. Amer. Chem. Soc. 1983. V. 103. P. 3250.
46. *Иогансен А.В.* // Теорет. и эксперим. химия. 1971. № 7. С. 302, 312.
47. *Раевский О.А., Григорьев В.Ю., Соловьев В.П.* // Хим.-фармацевт. журн. 1984. № 5. С. 578.
48. *Раевский О.А., Авидон В.В., Новиков В.П.* // Хим.-фармацевт. журн. 1982. № 8. С. 968.
49. *Мартынов И.В., Раевский О.А.* // Вестн. АН СССР. 1983. № 7. С. 99.
50. *Мартынов И.В., Раевский О. А.* // Докл. АН СССР. 1982. Т. 265. № 3. С. 664.
51. *Мартынов И.В., Раевский О.А.* // Журн. Всесоюз. хим. о-ва им. Д. И. Менделеева 1983. Т. 28. № 6. С. 716.
52. *Мартынов И.В., Раевский О. А.* // Докл. АН СССР. 1980. Т. 254. № 6. С. 1404.
53. *Мартынов И.В., Раевский О.А.* // Докл. АН СССР. 1985. Т. 283. № 2. С. 417.
54. *Тюрина Л.А., Семенов В.А., Кадыров Ч.Ш. и др.* // Агрехимия. 1986. № 1. С. 99.
55. *Новиков В.Л., Раевский О.А.* // Хим.-фармацевт. журн. 1982. № 5. С. 574.
56. *Soltzberg L.J., Wilkins C.L.* // J. Amer. Chem. Soc. 1977. V. 99. P. 439.

57. *Stuper A.J., Jurs P.C.* // *J. Amer. Chem. Soc.* 1975. V. 97. P. 182.
58. *Kubinyi H.* // *Arzneim-Forsch.* 1979. V. 29. N 8. P. 1067.
59. *Harisch C., Clayton J.M.* // *J. Pharm. Sci.* 1973. V. 62. P. 1.
60. *Szentpaly L.* // *J. Amer. Chem. Soc.* 1984. V. 106. P. 6021.
61. *Hansch C.* // *Drug Intell. Clin. Pharm.* 1982. V. 16. N 5. P. 391.
62. *Mager P.P.* // *Med. Res. Rev.* 1983. V. 3. N 4. P. 435.
63. *Hansch C.* // *J. Med. Chem.* 1976. V. 19. P. 1.
64. *Dekker W.H., Selling H.A., Overeem J.C.* // *J. Agr. Food Chem.* 1975, V 23. P. 785.
65. *Pursell W.P.* // *Eur. J. Med. Chem.* 1975. V. 10. P. 335.
66. *Goodford P.G., Norrington F. E.* // *Quant. Struct.-Act. Relat./Ed. Tichy M., Budapest: Akad. Kiado, 1979. P. 79.*
67. *Marchevka V.C., Ebner N.A., Sehon K.D., Hanna P.E.* // *J. Med. Chem.* 1985. V. 28. N 1. P. 18.
68. *Montgomery J.A., Mayo J.G., Hansch C.* // *J. Med. Chem.* 1974. V. 17. P. 477.
69. *Мостеллер Ф., Тьюки Дж.* // *Анализ данных и регрессия / Пер. с англ. М.: Финансы и статистика, 1982. Ч. 1, 2.*
70. *Фруммин Г.Т., Мухоморов В. К.* // *Хим.-фармацевт. журн.* 1985. № 3. С. 180.
71. *Barth A., Franke R., Orlick M.* // *Quant. Struct.-Activity Analysis // Eds Franke R. et al. Berlin: Akad.-Verlag, 1978. P. 395.*
72. *Dillon W.P., Goldstein M.* // *Multivariate Analysis. Methods and Applications. N. Y.: Wiley, 1984.*
73. *Dove S., Stretch W.J., Franke R.* // *J. Med. Chem.* 1980. V. 23. P. 1456.
74. *Dove S., Coats E., Scharfenberg P., Franke R.* // *J. Med. Chem.* 1985. V. 28. N 4. P. 447.
75. *Streich W. J., Dove S., Franke R.* // *J. Med. Chem.* 1980. V. 23. P. 1452.
76. *Wootton R.* // *J. Med. Chem.* 1983. V. 26. P. 275.
77. *Wold S., Dunn W.J.* // *J. Chem. Inf. Comput. Sci.* 1983. V. 23. N 1. P. 6.
78. *Жданов Ю.А., Минкин В.И.* // *Корреляционный анализ в органической химии. Ростов-на-Дону: РГУ, 1966.*
79. *Иберла К.* // *Факторный анализ / Пер. с англ. М.: Статистика, 1980.*
80. *Weiner M.L., Weiner P.H.* // *J. Med. Chem.* 1973. V. 16. P. 655.
81. *Franke R.* // *Farmaco.* 1978. V. 34. P. 545.
82. *Franke R., Dove S., Kiihne R.* // *Eur. Med. Chem.* 1979. V. 14. P. 363.
83. *Craig P. N.* // *J. Med. Chem.* 1971. V. 14. P. 680.
84. *Topliss J.G., Costello R.J.* // *J. Med. Chem.* 1972. V. 15. P. 1066.
85. *Себер Дж.* // *Линейный регрессионный анализ / Пер. с англ. М.: Мир, 1980. 456 с.*
86. *Moriguchi I., Komatsu K., Matsushita Y.* // *Quant. Struct.-Act. Relat.* 1984. V. 3. N 3. P. 106.
87. *Алгоритмы и программы восстановления зависимостей / Под ред. Вапника В.Н. М.: Наука, 1984. 816 с.*
88. *Джуре П., Айзенauer Т.* // *Распознавание образов в химии / Пер. с англ. М.: Мир, 1977. 230 с.*

89. Henry D.R., Jurs P.C., Denny W.A. // *J. Med. Chem.* 1982. V. 25. P. 899.
90. Free S. M., Wilson J.W. // *J. Med. Chem.* 1964. V. 7. P. 395.
91. Fujita T., Ban T. // *J. Med. Chem.* 1971. V. 14. P. 148.
92. Cammarata A., Yau S.J. // *J. Med. Chem.* 1970. V. 13. P. 93.
93. Clayton J.M., Pursell W.P. // *J. Med. Chem.* 1969. V. 12. P. 1087.
94. Craig P.N. // *J. Med. Chem.* 1972. V. 15. P. 144.
95. Корецкы J., Воцек K. // *Experientia.* 1967. V. 23. P. 125.
96. Айвазян С.А., Бежаева З.И., Староверов О.В. // Классификация многомерных наблюдений. М.: Статистика, 1974. 238 с.
97. Martin Y.C., Holland J.B., Jarboe C.H., Plotnikoff N. // *J. Med. Chem.* 1974. V. 17. P. 409.
98. Henry D. R., Block J. // *Eur. Med. Chem.* 1980. V. 15. P. 133.
99. Нигматуллин Р.С., Осипов А.Л., Пузаткин А.П., Коптюг В.А. // *Хим.-фармацевт. журн.* 1985. № 2. С. 179.
100. Jurs P.C., Stouch T.R., Czerwinski M., Narvaez J.N. // *J. Chem. Inf. Comput. Sci.* V. 25. P. 296.
101. Шульман В.С. // *Хим.-фармацевт. журн.* 1984. № 9. С. 1087.
102. Дюран Б., Оддел П. Кластерный анализ / Пер. с англ. М.: Статистика, 1977. 128 с.
103. Hansch C., Unger S.H., Forsythe A.B. // *J. Med. Chem.* 1973. V. 16. P. 1217.
104. Adamson G.N., Bawden D. // *Chem. Inf. Comput. Sci.* 1981. V. 21. P. 204.
105. Wold S. // *Pattern Recognition.* 1976. V. 8. P. 127.
106. Willett P. // *Chem. Inf. Comput. Sci.* 1984. V. 24. № 1. P. 29.
107. Randic M. // *Int. Quan. Chem. Quan. Biol.* 1984. N 11. P. 137.
108. *Advances in Drug Research.* V. 14 / Ed. Testa B. N. Y.: Acad. Press, 1985. 339 p.
109. Andy M. // *Comput. Graph. Int. Conf. Proc. L.: Pinner,* 1983. P. 435.
110. *Macromolecular Structure and Specificity: Computer Assisted Modelling and Applications* / Eds Kataraghaven B. et al., N. Y.: Acad. Sci., 1985. 209 p.
111. Hopfinger A.J. // *Pharm. Int.* 1984. V. 5. N 9. P. 224.
112. Chose A.K., Crippen G.M. // *J. Med. Chem.* 1985. V. 28. P. 333.
113. Beddel C.R. // *Chem. Soc. Rev.* 1984. V. 13. N 3. P. 279.
114. Hopfinger A.J. // *J. Med. Chem.* 1983. V. 26. P. 990.
115. Selassie C.D., Fang Z.-X., Li R.-L., Hansch C. // *J. Med. Chem.* 1986. V. 29. P. 621.
116. Кузьмин В.Е., Крутиус С.В. // *Хим.-фармацевт. журн.* 1986. № 7. С. 791.
117. Тихонов А.Н., Гончарский А.В., Степанов В.В., Ягола А.Г. Регуляризирующие алгоритмы и априорная информация. М.: Наука, 1983. 200 с.
118. Dyott T.M., Stuper A.J., Zander G.S. // *J. Chem. Inf. and Comput. Sci.* 1980. V. 20. N 1. P. 28.
119. Franke R. // *Theoretical Drug Design Methods.* Berlin: Akad.-Verlag. 1984. 415 p.
120. Rush J.E. // *J. Chem. Inf. Comput. Sci.* 1985. V. 89. P. 4426.
121. Clementi E.J. // *J. Phys. Chem.* 1985. V. 89. P. 4426.
122. Johnson A.P. // *Chem. in Brit.* 1985. V. 21. N 1. P. 59.
123. Зефирова М.С., Гордеева Е.В. // *Успехи химии.* 1987. Т. 56. С. 1753.

124. QSAR in Drug Design and Toxicology / Eds Hadzi D. et al. Amsterdam: Elsevier, 1987. 375 p.
125. Баренбойм Г.М., Маленков А.Г. Биологически активные вещества. Новые принципы поиска. М.: Наука, 1986. 363 с.
126. Liebman M.N. // J. Mol. Graphics. 1986. V. 4. N 1. P. 61.
127. Ripka W.C. // Nature. 1986. V. 321. P. 93.
128. Rose G.D., Lee R.H. // Biophys. J. 1986. V. 49. N 1. P. 83.
129. White D.N.J., Pearson J.E. // J. Mol. Graphics. 1986. V. 4. N 3. P. 134.
130. Hansch C., Klein T.E. // Acc. Chem. Res. 1986. V. 19. N 12. P. 392.
131. Meyer E.F., Bode W. // QSAR in Drug Des. Toxicol/Eds Hadzi D. et al. Amsterdam: Elsevier, 1987. P. 247.
132. Tomasi J. // Ibid. 1987. P. 269.
133. Labanowski J., Motoc I., Naylor C.B. et al. // QSAR. 1986. V. 5. P. 138.
134. Cummins P.L., Rendell A.P.L., Swanton D.J. et al. // Int. Rev. Phys. Chem. 1986. V. 5. N 2/3. P. 139.
135. Kikuchi Q. // QSAR. 1987. V. 6. P. 179.
136. Fujita T. // QSAR. 1987. V. 6. N 2. P. 54.
137. Hoi W.G.J. // Angew. Chem. Int. Ed. Engl. 1986. V. 25. N 9. P. 767.
138. Kier L.B., Hall L.H. // Molecular Connectivity in Structure-Activity Analysis. N. Y.: Wiley, 1986. 262 p.
139. Kier L.B. // QSAR. 1986. V. 5. N 1. P. 7.
140. Kier L.B. // QSAR. 1987. V. 6. N 1. P. 8.
141. Kier L.B. // QSAR. N 3. P. 117.
142. Verloop A. The STREIMOL Approach to Drug Design. N. Y.: Marcel Dekker, 1987.
143. Kramer C.R. // Z. Phys. Chem. 1986. V. 267. N 2. P. 277.
144. Trinajstić N., Klein D.J., Randić M. // Int. J. Quant. Chem. Quant. Chem. Symp. V. 20. P. 699.
145. Dunn W.J., Wold S. // Trends Anal. Chem. 1986. V. 5. N 3. P. 53.
146. Химические приложения топологии и теории графов / Пер. с англ., под ред. Кинга Р. М.: Мир, 1987. 560 с.
147. Ramaswamy N. // Chem. Inf. Comput. Sci. 1987. V. 27. N 2. P. 82.
148. Ghose A.K., Crippen G.M. // J. Comput. Chem. 1986. V. 7. N 4. P. 565.
149. Young G., Lien E.J., Guo Z. // QSAR. 1986. V. 5. N 1. P. 12.
150. McFarland J.W., Gans D.J. // J. Med. Chem. 1986. V. 29. N 4. P. 505.
151. Derde M.P., Massart D.L. // Anal. Chim. Acta. 1986. V. 191. P. 1.
152. Advances in Multivariate Statistical Analysis. 3 / Ed. Gupta A.K. Dordrecht: Reidel, 1986, 402 p.
153. McFarland J.W., Gans D.J. // QSAR in Drug Des. Toxicol. / Eds Hadzi D. et al. Amsterdam: Elsevier, 1987. P. 25.
154. Lopes de Compadre R.L., Pearlstein R.A. et al. // Ibid. P. 149.
155. Janssen L.H.M. // Ibid. P. 191.
156. Morris J.J., Hughes L.R., Taylor P. // Ibid. P. 204.
157. Goldblum A. // Ibid. P. 207.
158. Bruns A., Wahnsiedler W., Weitzel P. et al. // Ibid. P. 103.
159. Tichy M. // Ibid. P. 109.

160. Grunewald G.L., Sail D.J., Monn J.A. // *Ibid.* P. 118.
161. Peace E.A., Rord M.G., Greenwood R. et al. // *Ibid.* P. 331.
162. *Ley sen J.E., Gommeren W.* // Drug dev. Res. 1986. N 8. P. 119.
163. *Balaz S., Sturdic E., Augustin J.* // Biophys. Chem. 1986. V. 24. P. 135.
164. *Jakes S.E., Willett P.* // *J. Mol. Graphics.* 1986. V. 4. N 1. P. 12.
165. Artificial Intelligence Applications in Chemistry / Eds Pierce T. H. et al. Washington: Amer. Chem. Soc. 1986.
166. *Ayscough P.B., Chinnick S.J., Dybowski P.* // Chem. Ind. 1987. V. 15. P. 515.
167. *Derde M.P., Buydens L., Guns C. et al.* // Anal. Chem. 1987. V. 59. N 14. P. 1868.
168. *Darvas E.* // *QSAR in Drug Des. Toxicol* / Eds Hadzi D. et al. Amsterdam: Elsevier. 1987. P. 37.
169. *Элти Дж., Кумбс М.* // Экспертные системы: концепции и примеры / Пер. с англ. М.: Финансы и статистика, 1987. 191 с.
170. Экспертные системы. Принципы работы и примеры / Пер. с англ., под ред. Форсайта Р.М.: Радио и связь, 1987. 224 с.
171. Построение экспертных систем / Пер. с англ., под ред. Хейеса-Рота Ф. и др. М.: Мир, 1987. 441 с.

О.А. Раевский, С.В. Трепалин, А.Н. Раздольский

НОВЫЕ КССА ДЕСКРИПТОРЫ, РАССЧИТЫВАЕМЫЕ ИЗ СПЕКТРОВ МЕЖАТОМНЫХ ВЗАИМОДЕЙСТВИЙ

*Институт физиологически активных веществ РАН, Черноголовка,
Московская область*

Введение

Свойства химических соединений, в том числе биологическая активность, определяются характером их взаимодействий с другими соединениями или биологическими объектами [1, 2]. Описание межмолекулярных взаимодействий возможно посредством термодинамических параметров на основе известного соотношения:

$$\Delta G = G_f - G_i \quad (1)$$

где G – свободная энергия, индексы f и i относятся к конечному и исходному состоянию.

Строгое количественное описание межмолекулярных взаимодействий требует знания не только разности свободных энергий (ΔG), но также разности энтальпии (ΔH) и разности энтропии (ΔS), связанных между собой соотношением:

$$\Delta G = \Delta H - T\Delta S, \quad (2)$$

где T – температура реакции. Экспериментальное определение величин ΔG обычно осуществляется посредством измерения констант (K) реакции (комплексообразования), связанных со свободной энергией соотношением:

$$\Delta G = \Delta G^0 - RT \ln K, \quad (3)$$

где ΔG^0 – разность свободной энергии реакции в стандартных условиях.

Следует отметить, что большинство процедур расчета энергий реакций и взаимодействий основывается на предположении об аддитивности вкладов различных типов межмолекулярных взаимодействий [1]:

$$\Delta G = \Delta G_{\text{solvent}} + \Delta G_{\text{conf}} + \Delta G_{\text{int}} + \Delta G_{\text{motion}}, \quad (4)$$

где $\Delta G_{\text{solvent}}$ – вклад растворителя, ΔG_{conf} – изменение свободной энергии за счет конформационных изменений во взаимодействующих молекулах, ΔG_{int} – изменение свободной энергии за счет специфических взаимодей-

ствий, ΔG_{motion} – изменение свободной энергии за счет изменения трансляционной, вращательной и колебательной составляющих.

Специфические взаимодействия рассчитываются на основе предположения об аддитивности вкладов различных типов таких взаимодействий. В настоящее время считается, что основной вклад в специфические взаимодействия вносят стерический (st), электростатический (el) эффекты и водородная связь (hb). В рамках этого предположения свободная энергия специфических взаимодействий может быть представлена следующим образом [3]:

$$\Delta G_{\text{int}} = \Delta G_{\text{st}} + \Delta G_{\text{el}} + \Delta G_{\text{hb}} \quad (5)$$

В настоящее время существует множество исследовательских и коммерческих компьютерных программ, предназначенных для расчетов как энергий специфических взаимодействий, так и дескрипторов межмолекулярных взаимодействий, которые используются в процедурах моделирования количественных связей «структура – активность» (КССА). Здесь, прежде всего, необходимо упомянуть о программах CoMFA [4] и GRID [3]. Обе указанные программы основываются на расчетах энергий взаимодействий какого-либо пробного атома, располагаемого поочередно во всех узлах решетки, в которую помещаются молекулы обучающей выборки.

В качестве альтернативы вышеупомянутым программам ComFa и GRID можно рассматривать подходы, использующие отображение пространственной структуры и способности к межмолекулярным взаимодействиям посредством спектра межатомных расстояний или межатомных взаимодействий. В настоящей работе описан подход, который впервые описывает стерические, электростатические взаимодействия и водородные связи посредством спектров соответствующих взаимодействий.

Описание дескрипторов

При использовании спектров межатомных расстояний молекулярная структура представляется в виде кривой радиального рассеяния, которая легко может быть вычислена для молекулы любой сложности, если известны ее декартовы координаты. Аналогичный аппарат применяется при изучении строения молекул методом газовой электронографии.

Функция радиального распределения, $M(s)$, в приближении сферической симметрии атомов и в отсутствие внутримолекулярных колебаний представляет собой спектр межатомных расстояний, интенсивность которого определяется видом атомов, находящихся на данном расстоянии друг от друга, и динамикой изменения этого расстояния в процессе внутримолекулярных колебаний:

$$M(s) = \sum_{i=1}^N \sum_{j>1}^N C_{ij} \cdot \frac{\sin sr_{ij}}{sr_{ij}} \quad (6)$$

где s – угловой параметр: $s = (4\pi/\lambda)\sin Q/2$, λ – длина волны электронов, Q – угол рассеяния, r_{ij} – межатомное расстояние в молекуле, C_{ij} – коэффициент, характеризующий рассеивающую способность данной пары атомов.

Можно сказать, что функция радиального распределения, получаемая в процессе электронографического эксперимента, – это то, что «видит» электрон, который взаимодействует с молекулой в акте рассеяния.

В публикации [5] эта аналогия была впервые продолжена на случай взаимодействия рецептора с молекулами введенного в организм вещества. Как известно, взаимодействие рецептора происходит не со всей молекулой реагента, а с частями молекулы, комплементарными активным центрам рецептора. При этом определяющими факторами взаимодействия являются расстояния между активными центрами и способность к межмолекулярным взаимодействиям атомов, находящихся на этих расстояниях в молекулах реагента. Имеет место своеобразный процесс рассеяния молекул реагента на активных центрах рецептора, т. е. рецептор «видит» молекулу в образе функции радиального распределения. При таком подходе существует возможность модифицировать функцию радиального распределения таким образом, чтобы количественно учесть способность атомов реагента к различным межмолекулярным взаимодействиям. В публикациях [5, 6] было высказано предложение учитывать вместо типа атомов их донорно-акцепторную способность. Использование функции радиального взаимодействия в качестве дескриптора молекулярной структуры при исследовании взаимосвязей «структура – активность» описано позднее также в публикациях [7–17].

Конкретное воплощение подхода осуществлено нами в специальном программном комплексе. Первоначальная исследовательская версия программы описана в публикациях [18–21]. Подробное описание архитектуры программы и ее использование представлено в сообщениях [22, 23].

При использовании указанной программы рассчитываются спектры:

1. стерических взаимодействий атомов,
2. взаимодействий положительно заряженных атомов между собой,
3. взаимодействий отрицательно заряженных атомов между собой,
4. взаимодействий положительно заряженных атомов с отрицательно заряженными,
5. взаимодействий доноров водородной связи между собой,
6. взаимодействий акцепторов водородной связи между собой,
7. взаимодействие доноров и акцепторов водородной связи.

Реализованные спектры межатомных взаимодействий сохранили лишь формальную аналогию с кривыми радиального распределения. В обоих случаях каждая пара рассматриваемых центров дает линию. Колебания атомов трансформируют линии в полосы. Однако в спектрах межмолекулярных взаимодействий интенсивность полос определяется лишь локальными характеристиками рассматриваемых центров, в то время как в кривых ради-

ального распределения существует нормировка интенсивностей по расстоянию между центрами.

Суперпозиция полос всех возможных пар центров формирует конечный спектр межатомных взаимодействий.

На рисунке 1 представлены спектры взаимодействий акцепторов водородной связи в некоторых макроциклических полиэфирах. В принципе, каждую *точку каждого вида спектра* можно рассматривать в качестве дескриптора межмолекулярных взаимодействий. При разумных значениях интервала расстояний 1,1–20,0 Å и шага в 0,1 Å для каждого вида спектров получается 190 дескрипторов. А семь видов спектров позволяют описать возможные стерические и кулоновские взаимодействия, а также водородную связь посредством 1340 дескрипторов. Примеры использования спектров межатомных взаимодействий для установления взаимосвязей «структура – свойства» описаны нами в [18–21] и обсуждаются в деталях в сообщениях, посвященных описанию программы [22] и ее применению [23].

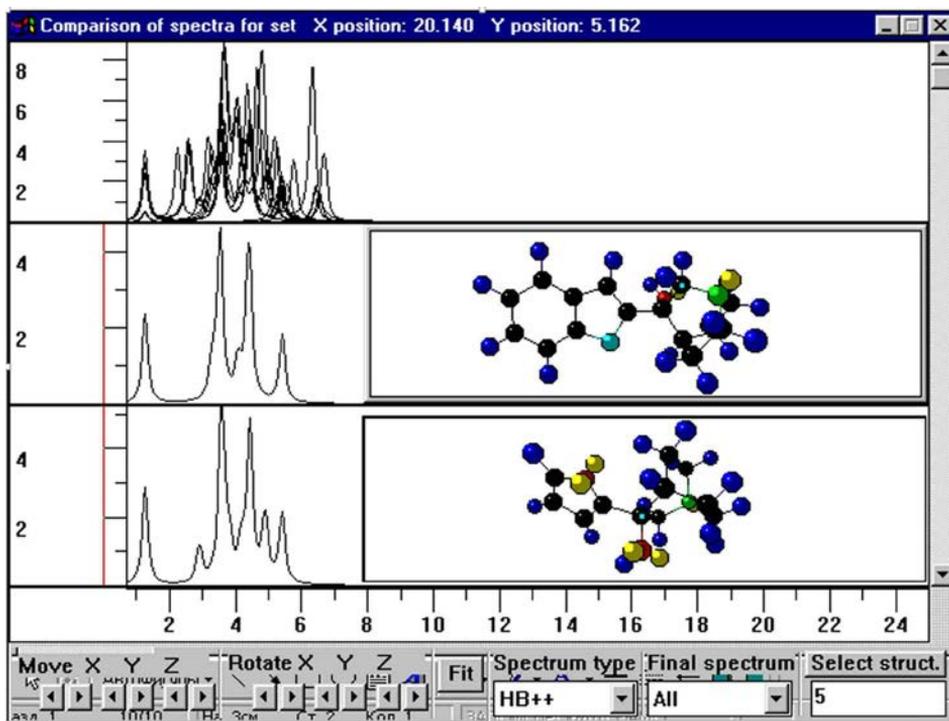


Рис. 1. Пример спектров взаимодействий водородно связанных акцепторных центров

Спектры межатомных взаимодействий исключительно удобны и для оценки сходства молекулярной структуры. В отличие от уже существующих подходов сопоставления трехмерных структур [24–26], спектры межмоле-

кулярных взаимодействий уже при их построении имеют единую систему отсчета и не требуют специальных дополнительных процедур для проведения сопоставления.

Степень перекрытия спектров может быть оценена посредством индекса сходства (SI):

$$SI (\%) = 100 (1 - DIS) \quad (7)$$

$$\text{где } DIS = [\sum(Y1(i) - Y2(i))] / 2 \times [(\sum Y1(i))^2 + (\sum Y2(i))^2] \quad (8)$$

где $Y1(i)$ и $Y2(i)$ – интенсивности спектров первого и второго соединений при i -расстоянии.

Пример наложения спектров межатомных взаимодействий и оцененные значения индексов сходства представлены на рисунке 2.

Предложенные новые КССА дескрипторы, рассчитываемые из спектров межатомных взаимодействий, включают в себя информацию как о пространственной структуре химических соединений, так и о способности их атомов к взаимодействиям. Можно полагать, что эти свойства приведут к широкому их использованию при выявлении реальных взаимосвязей «структура – активность».

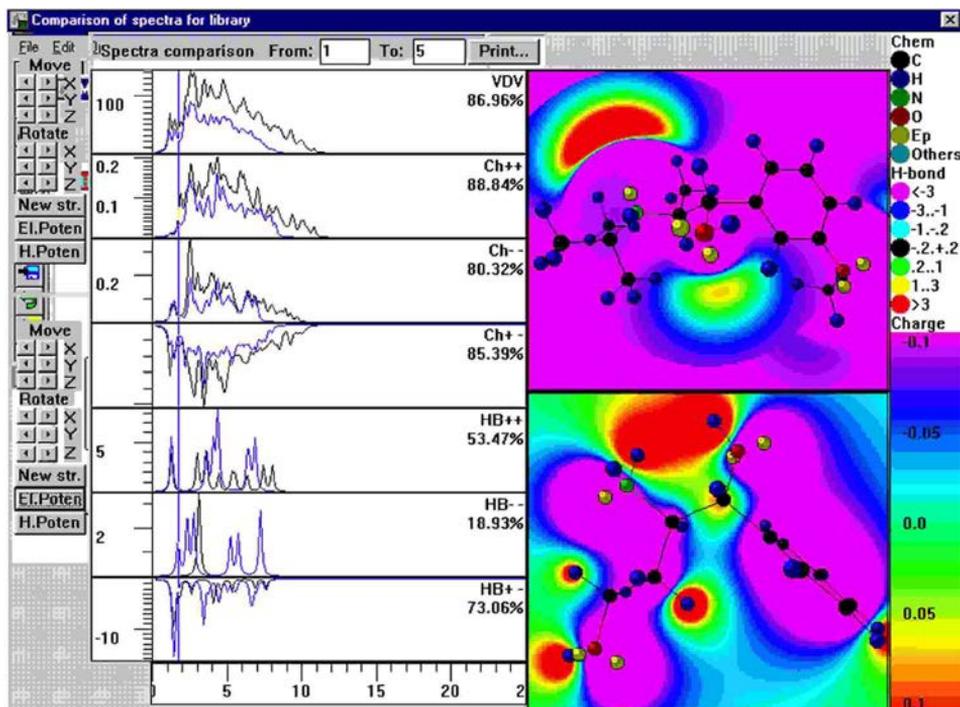


Рис. 2. Пример спектров межмолекулярных взаимодействий и их индексов сходства

Работа была выполнена благодаря поддержке Миннауки (проект N 0201.07.201), Федеральной целевой научно-технической программы «Исследования и разработки по приоритетным направлениям науки и техники гражданского назначения» (подпроект «Компьютерное конструирование лекарственных средств») и РФФИ (проект 97-03-33068а).

ЛИТЕРАТУРА

1. *Ajay, M. Murcko.* J.Med.Chem. 1995. **38**, 4953.
2. *M. Charton.* In Lipophilicity in Drug Action and Toxicology.(eds. V.Pliska, B.Testa, H.Waterbeemd). VCH,Weinheim. 1996. p.387.
3. *Goodford P.J.* J.Med.Chem. 1985. **28**, 849.
4. *Cramer III R.D., Patterson D.E., Bunce J.D.* J.Am.Chem.Soc. 1988. **110**, 5959.
5. *Новиков В.П., Раевский О.А.* Хим.-фарм.ж. 1982. **16**, 574.
6. *Raevsky O.A.* In QSAR in Drug Design and Toxicology.(eds. D. Hadzy, B. Jerman-Blazic).Elsevier, Amsterdam. 1987. p.31.
7. *Csorvassy I., Tozser L.* In QSAR:Rational Approaches to the Design of Bioactive Compounds. (esd. C.Silipo, A.Vittoria).Elsevier, Amsterdam. 1991. p.193.
8. *King J.W., Kassel R., King B.B.* Int. J. Quant. Chem.:Quant. Biol. Simp. 1990. **17**, 27.
9. *King J.W., Kassel R.J.* Int.J. Quant. Chem.: Quant. Biol. Simp. 1991. **18**, 289.
10. *King J.W., Kassel R.J.* Int. J. Quant. Chem.: Quant. Biol.Simp. 1992. **19**, 179.
11. *King J.W.* Int. J. Quant. Chem.: Quant. Biol. Simp. 1993. **20**, 139.
12. *Famini G.R., Kassel R.J., King J.W., Wilson L.Y.* Quant. Struct. — Act. Relat. 1991. **10**, 344.
13. *King J.W.* Int. J. Quant. Chem.: Quant. Biol. Simp. 1994. **21**, 209.
14. *Кучкаев Б.И., Князев Б.А., Курочкин В.К., Петрунин В.А.* Хим.-фарм. журн. 1992. **26**, 60.
15. *Кучкаев Б.И., Князев Б.А., Курочкин В.К., Петрунин В.А.* Хим.-фарм. журн. 1997. **31**, 30.
16. *Schuur J.H., Selzer P., Gasteiger J.* J.Chem. Inf. Comput .Sci. 1996. **36**, 334.
17. *Gasteiger J., Sadowski J., Schuur J., Selser P., Steinhauer L., Steinhauer V.* J. Chem. Inf. Comput. Sci. 1996. **36**, 1030.
18. *Raevsky O., Dolmatova L., Grigor'ev V., Lisyansky I., Bondarev S.* In QSAR and Molecular Modelling: Concepts, Computational Tools and Biological Applications. (eds. F. Sanz, J. Giraldo, F. Manaut). R. Prous, Barcelone, 1995. p.241.

19. *Raevsky O.A.* J. Phys. Org. Chem. 1997. **10**, 405.
20. *Raevsky O.A.* In Computer-Aided Lead Finding and Optimization. Current Tools for Medicinal Chemistry. (eds. H. Waterbeemd, B. Testa, G. Folkers). WILEY-VCH, Weinheim, 1997. p.36.
21. *Лукоянов Н.В., Раевский О.А.*. Вопросы медицинской химии. 1998. **44**, 185.
22. *Трепалин С.В., Раздольский А.Н., Раевский О.А.* Хим.-фарм. ж. 2000. **34**, 23.
23. *Раздольский А.Н., Трепалин С.В., Раевский О.А.* Хим.-фарм. ж. 2000. **34**, 26.
24. *Barakat M.T., Dean P.M., Comput J. Aided Mol.Design.* 1990. **4**, 295.
25. *Blaney F., Finn P., Phippen R., Wyatt M., Mol J. Graph.* 1993. **11**, 98.
26. *Perkins T.D.J., Dean P.M.* In Computer-Aided Lead Finding and Optimization. Current Tools for Medicinal Chemistry. (eds. H. Waterbeemd, B. Testa, G. Folkers). WILEY-VCH, Weinheim, 1997. p. 423.

О.А. Раевский¹, А.Н. Раздольский¹, В.Д. Тонкопий^{1,2}, И.В. Иофина^{1,2},
А.О. Загребин^{1,2}

КЛАССИФИКАЦИОННЫЕ И КОЛИЧЕСТВЕННЫЕ МОДЕЛИ ВЗАИМОСВЯЗИ СТРУКТУРЫ ХИМИЧЕСКИХ СОЕДИНЕНИЙ И ИХ ТОКСИЧНОСТИ ПО ОТНОШЕНИЮ К *DAPHNIA MAGNA*

¹Институт физиологически активных веществ РАН, Черноголовка,
Московская область

²Институт озераведения РАН, Санкт-Петербург

Проблемы воздействия химических соединений на окружающую среду являются очень серьезными в условиях непрерывного повышения объемов производства и использования химических соединений в промышленности, сельском хозяйстве и медицине. К сожалению, даже в случае широко используемых соединений лишь для малой их части имеются данные по токсическому воздействию на живые организмы экосистем. По этой причине создание количественных моделей «структура – активность» (КССА) могут быть полезны для предварительной оценки токсичности как уже используемых, так и новых химических соединений.

Одним из важных элементов водной экологии являются рачки дафниды, вследствие их способности фильтровать и очищать воду. В то же время это один из старейших и широко используемых тест-объектов для оценки чистоты и качества воды из-за способности рачков пропускать через себя огромное по отношению к собственному весу количество воды и высокой чувствительности *Daphnia* к различным ксенобиотикам. Именно по этой причине этот тест-объект используется во многих странах [1].

В настоящей работе проведено моделирование взаимосвязи структуры химических соединений и их токсичности по отношению к *Daphnia magna* на основе собственных экспериментальных данных и современных методов КССА.

Экспериментальная биологическая часть

Определение токсичности (LC50) в опытах на дафниях

Для определения токсичности использовали культуру *Daphnia magna* двух-трехдневного возраста, генетически однородных, что позволяет минимизировать возможность разброса результатов. Необходимое для определения токсичности количество дафний определяли числом изучаемых концентраций вещества. При определении токсичности дафнии помещались в стаканчики с 50 мл раствора различных концентраций исследуемого вещества (по 10 рачков в каждый сосуд). Растворы готовили на воде, используемой для культивирования дафний (водопроводной дехлорированной) при температуре 18 °–19 °С. Растворы веществ использовали в 5–10 возрастающих по логарифмической шкале концентрациях. Каждая серия опытов сопровождается контрольными испытаниями.

В качестве основного показателя токсичности использовалась выживаемость рачков. Токсичность исследуемых соединений определяли по величине LC50 – концентрации соединений, приводящей к гибели 50% гидробионтов при 24-часовой инкубации. Величины LC₅₀ определяли методом пробит-анализа Литчфилда и Уилкоксона в модификации В. Прозоровского [2].

Гибель рачков отмечали при наступлении неподвижности (иммобилизации): дафнии лежат на дне стакана, плавательные движения отсутствуют и не возобновляются при покачивании стакана. При обработке результатов острого эксперимента проводили сравнение выживаемости в опытных и контрольных сосудах.

Хорошо известно, что воздействие химических соединений на *Daphnia magna* может протекать по различным механизмам физиологического действия. В качестве таких механизмов обычно рассматривают общий наркоз, полярный наркоз, ингибирование ацетилхолинэстеразы, а также воздействие на другие ферменты и биологические мишени (специфическое действие). По этой причине в процессе количественной оценки токсичности производилось и отнесение химического соединения к тому или иному классу на основе поведения рачков. Результаты классификации и величины LC₅₀ для всех изученных 119 соединений приведены в таблице 1. Отметим, что из общего числа изученных 119 соединений 24 вещества проявили токсичность по типу общего токсического действия, 18 – полярного наркоза, 37 являются ингибиторами холинэстераз и 40 – специфического действия.

Моделирование КССА

Моделирование проводилось с использованием компьютерных программ, созданных в Отделе компьютерного молекулярного дизайна ИФАВ

РАН. Программа HYBOT (HYdrogen BOnd Thermodynamic) [3,4] рассчитывает 32 физико-химических дескриптора, связанных со стерическими, электростатическими и водородно-связанными взаимодействиями атомов. Программа MOLDIVS (MOLecylar DIVersity and Similarity) [5,6] оценивает сходство структурных формул соединений, выстраивая все молекулы, входящие в обучающую выборку, в порядке уменьшения величины индекса сходства Танимото (Tc):

$$Tc = N(A\&B) / [N(A) + N(B) - N(A\&B)],$$

где: N(A) – число фрагментов в молекуле А,

N(B) – число фрагментов в молекуле В,

N(A&B) – число общих фрагментов в молекулах А и В.

Липофильность изученных соединений оценивалась программой Slipper (Solubility, LIPophilicity & PERmeability) [7].

Результаты и их обсуждение

На первом этапе КССА-исследования изученного ряда соединений целесообразно было провести дискриминацию соединений с общим токсическим действием от остальных веществ. Это связано с тем, что 60% промышленных загрязнителей окружающей среды проявляют себя как вещества общего токсического действия. Таким образом, надежное отнесение соединения к веществам общего токсического действия может помочь избежать дальнейших дорогостоящих затрат на проведение экспериментов по изучению токсичности.

Результаты дискриминации между веществами общего токсического действия и всеми остальными, полученными программой MOLDIVS, приведены в таблице 2 (строки 1–4). Очевидно, что указанная дискриминация может быть четко реализована на основе первого структурно родственного соединения. В этом случае процент правильного распознавания составляет 91,7% для соединений общего токсического действия и 90,5% для соединений с иным типом действия. Использование же других родственных соединений (2–5-й сосед) не улучшает результаты, что свидетельствует о недостаточно большом размере обучающей выборки для использования нескольких родственных структур.

После надежной дискриминации между веществами общего токсического действия и остальными было целесообразно ставить вопрос о количественной модели описания токсичности соединений общего токсического действия. Обычно это делается с использованием параметра липофильности (выраженного коэффициентом распределения веществ в системе октанол/вода [$\log P$]) [8] в рамках регрессионного анализа. Действительно, исполь-

зование этого параметра привело к хорошему количественному описанию токсичности изученных 24 соединений общего токсического действия:

$$\log(1/LC_{50}) = -2,25(0,12) + 0,65(0,07) \log P \exp \quad (1)$$
$$N = 24, R = 0,89, SD = 0,44, F = 87,95,$$

где N – число соединений, R – коэффициент корреляции, SD – среднее квадратичное отклонение, F – критерий Фишера.

Уравнение 1, хотя и имеет хорошие статистические критерии, обладает существенным недостатком, связанным с использованием экспериментально определяемого параметра – липофильности. Это означает, что для расчета токсичности необходимо вещество синтезировать и измерить его коэффициент распределения октанол/вода. Поэтому мы решили заменить экспериментальный параметр $\log P$ на этот же параметр, рассчитываемый программой Slipper.

Полученные результаты отражены уравнением 2:

$$\log(1/LC_{50}) = -2,29(0,13) + 0,64(0,08) \log P_{\text{Slipper}} \quad (2)$$
$$N = 24, R = 0,87, SD = 0,49, F = 67,81.$$

Сравнение уравнений 1 и 2 показывает возможность использования рассчитанных программой Slipper значений $\log P$ для оценки общего токсического эффекта.

В наших предыдущих исследованиях было установлено, что параметр липофильности является композитным дескриптором и включает в себя конкуренцию двух эффектов – стерического и водородной связи [9, 10]. По этой причине мы использовали также молекулярную поляризуемость (Alpha), описывающую стерические взаимодействия, и сумму водородно-связанных акцепторных факторов (SumCa), которая рассчитывается программой HUVOT, для их корреляции с токсичностью этих же 24 соединений. В результате было получено следующее уравнение:

$$\log(1/LC_{50}) = -2,66(0,43) + 0,19(0,03) \text{Alpha} - 0,39(0,16) \text{SumCa} \quad (3)$$
$$N = 24, R = 0,80, SD = 0,61, F = 18,73.$$

Из уравнения 3 следует, что увеличение размера молекул приводит к возрастанию токсического действия и, наоборот, наличие атомов с большой способностью к образованию водородной связи уменьшает токсичность веществ этого типа действия. Таким образом, проведенная дискриминация соединений с общим токсическим действием по отношению к соединениям с иным типом действия и количественная модель, выраженная уравнением 3, основанная на использовании физико-химических рассчитываемых параметре-

тров, позволяет четко идентифицировать соединения общего токсического действия и рассчитывать их величины LC_{50} для рачков *Daphnia magna*.

На рисунке представлена зависимость токсичности от липофильности, рассчитанной программой Slipper, для всех изученных соединений. Очевидно, что для соединений с действием по типу полярного наркоза, ингибирования холинэстераз и специфического типа действия токсичность по отношению к *Daphnia magna* формируется за счет вкладов как механизма общего токсического действия, так и дополнительного приращения, связанного с иным типом действия.

На следующем этапе работы была проведена классификация соединений с механизмом полярного наркоза по сравнению с соединениями, проявляющими иные механизмы действия. Результаты дискриминации изученных нами соединений этого типа действия приведены в таблице 2 (строки 5–8). Данные таблицы 2 подтверждают возможность использования структурного сходства для дискриминации веществ, действующих по типу полярного наркоза. Процент правильного распознавания таких соединений по их первому структурному соседу составил 83,3%, а соединений с иным типом действия правильно идентифицируются с 93,1%.

Поскольку часто дискутируется вопрос о возможности отдельной идентификации веществ, действующих по механизмам общего токсического действия и полярного наркоза, была проведена дискриминация соединений указанных двух механизмов методом структурного сходства. Результаты этой работы представлены в таблице 2 (строки 9–12). Очевидно, что дискриминация веществ, действующих по этим двум механизмам, может успешно проводиться используемым методом структурного сходства (100% правильного распознавания для соединений с общим типом токсического действия и 89,9% – для соединений с типом действия полярного наркоза).

Также была проведена работа по выделению в отдельные группы соединений – веществ, ингибирующих функционирование холинэстераз в организме *Daphnia magna*, и веществ, действующих специфически на другие рецепторы изучаемого биообъекта. Для этого также использовался метод ближайшего структурного сходства. Результаты приведены в таблице 2 (строки 13–16 и 17–20). Очевидно, что используемый КССА-подход достаточно хорошо проводит дискриминацию между соединениями с антихолинэстеразным механизмом (правильное распознавание 81,1%) и соединений с иным механизмом действия (здесь правильное распознавание достигает величины 91,5%). Что же касается дискриминации между соединениями, отнесенными к специфическим по отношению ко всем другим, результаты не столь впечатляющи: хотя 92,4% соединений с иным типом действия здесь правильно классифицируются, но только две третьих (65%) веществ специфического действия распознаются правильно. Здесь следует отметить, что понятие вещества специфического типа действия является в настоящее время недостаточно четко формализованным. К соединениям этого класса

относятся вещества, действующие на различные мишени внутри организма *Daphnia magna*. Можно предположить, что дальнейшее исследование механизмов действия химических соединений на *Daphnia magna* приведут к более четкой классификации типов действия и, следовательно, дадут возможность построения более стабильных, физиологически обоснованных КССА-моделей токсичности.

Таким образом, на основе использования метода структурного сходства удалось построить стабильные классификационные модели для дискриминации соединений с механизмом общего токсического действия по отношению ко всем остальным соединениям; соединений с механизмом полярного наркоза по отношению ко всем остальным; соединений, ингибирующих холинэстеразы, по отношению к соединениям с другими механизмами; а также соединений, проявляющих общее токсическое действие по отношению к веществам, проявляющим механизм полярного наркоза. Совместное использование классификационной и регрессионной (уравнение 3) моделей, основанное на указанных физико-химических дескрипторах, позволяет четко идентифицировать вещества, действующие по общему токсическому механизму, количественно оценивать величины LC_{50} и анализировать вклады в токсичность стерических взаимодействий и водородной связи. Что же касается количественных моделей для токсичности соединений, действующим по иным механизмам, то их построение, очевидно, должно основываться не только на двухмерных структурных и физико-химических параметрах, но и учитывать пространственное расположение взаимодействующих атомов. Исследования в этой области будут являться предметом нашей дальнейшей работы.

Работа выполнена при финансовой поддержке МНТЦ (проект 0888).

Табл. 1. Токсичность ($\log [1/LC_{50}]$), рассчитанная липофильность ($\log P$ Slipper) и тип токсического действия изученных веществ

	Наименование	Брутто-формула	CAS	$\log(1/LC_{50})$	$\log P$ Slipper	Тип действия
1	Этанол	C2 H6 O	64-17-5	-2.90	-0.12	Общий
2	Метанол	C H4 O	67-56-1	-2.84	-0.63	Общий
3	2-Этоксизанол	C4 H10 O2	110-80-5	-1.60	-0.35	Общий
4	2-Метоксизанол	C3 H8 O2	109-86-4	-2.74	-0.80	Общий
5	Этиленгликоль	C2 H6 O2	107-21-1	-2.81	-1.45	Общий
6	Четыреххлористый углерод	C Cl4	56-23-5	-0.64	2.36	Общий
7	1,2-Дихлорэтан	C2 H4 Cl2	1300-21-6	-1.46	1.46	Общий
8	Дихлорметан	C H2 Cl2	75-09-2	-1.69	1.67	Общий
9	n-Бутанол	C4 H10 O	71-36-3	-2.09	1.21	Общий
10	2-Метил-1-пропанол	C4 H10 O	78-83-1	-2.15	0.79	Общий
11	n-Пропанол	C3 H8 O	71-23-8	-2.54	0.50	Общий
12	2-Пропанол	C3 H8 O	67-63-0	-2.72	0.06	Общий
13	Амиловый спирт	C5 H12 O	71-41-0	-1.34	1.73	Общий
14	1-Гексанол	C6 H14 O	111-27-3	-1.01	2.24	Общий
15	n-Гептанол	C7 H16 O	111-70-6	0.25	2.55	Общий
16	Ацетон	C3 H6 O	67-64-1	-2.34	-0.10	Общий
17	Трихлорэтилен	C2 H Cl3	79-01-6	0.13	2.58	Общий
18	Бензол	C6 H6	71-43-2	-0.82	2.13	Общий
19	2-Метил-бензонитрил	C8 H7 N	529-19-1	-0.12	2.12	Общий
20	Бутан-2-он	C4 H8 O	78-93-3	-1.68	0.62	Общий
21	Циклогексанол	C6 H12 O	108-93-0	-1.72	1.16	Общий
22	2-Бутокси-этанол	C6 H14 O2	111-76-2	-1.55	0.67	Общий
23	Дифенилметанон	C13 H10 O	119-61-9	-0.61	3.43	Общий

Продолжение таблицы 1

24	о-Ксилен	C8 H10	95-47-6	-0.70	3.08	Общий
25	Пентахлорофенол	C6 H Cl5 O	87-86-5	3.03	3.43	Полярный
26	Пиридин	C5 H5 N	110-86-1	2.50	0.68	Полярный
27	Анилин	C6 H7 N	62-53-3	2.16	0.99	Полярный
28	3-Метилфенол	C7 H8 O	108-39-4	0.24	2.02	Полярный
29	4-Этилфенол	C8 H10 O	123-07-9	0.14	2.55	Полярный
30	2-Аллилфенол	C9 H10 O	1745-81-9	0.80	2.65	Полярный
31	2,4-Дихлорофенол	C6 H4 Cl2 O	120-83-2	0.91	2.97	Полярный
32	2-Метилфенол	C7 H8 O	95-48-7	0.48	1.88	Полярный
33	Нафтол	C10 H8 O	1321-67-1	1.43	2.67	Полярный
34	Фенол	C6 H6 O	108-95-2	0.65	1.53	Полярный
35	О-Толуламин	C7 H9 N	95-53-4	0.83	1.32	Полярный
36	пара-Толуидин	C7 H9 N	106-49-0	0.35	1.37	Полярный
37	Бензиламин	C7 H9 N	100-46-9	0.93	-1.11	Полярный
38	2,5-Дихлоранилин	C6 H5 Cl2 N	95-82-9	1.05	2.83	Полярный
39	3,4-Дихлоранилин	C6 H5 Cl2 N	95-76-1	0.92	2.72	Полярный
40	4-Этиламин	C8 H11 N	589-16-2	0.98	1.80	Полярный
41	п-Пропиламин	C3 H9 N	107-10-8	0.45	-2.28	Полярный
42	2-Метоксэтиламин	C3 H9 N O	109-85-3	0.55	-2.83	Полярный
43	Ареколин	C8 H13 N O2	63-75-2	0.87	0.94	АнтиХЭ
44	4-азабицикло[2,2,2]окт-2-ил ацетат	C9 H15 N O2	827-61-2	0.78	-1.17	АнтиХЭ
45	этил 4-нитрофенилэтилфосфонат	C10 H14 N O5 P	546-71-4	4.99	1.64	АнтиХЭ
46	О-Ацетилхолин	C7 H16 N O2	51-84-3	0.30	2.68	АнтиХЭ
47	Галантамин	C17 H21 N O3	357-70-0	1.74	0.04	АнтиХЭ

**Избранные труды доктора химических наук, профессора О.А. Раевского
в области молекулярного компьютерного дизайна**

Продолжение таблицы 1

	Наименование	Брутто-формула	CAS	log(1/LC ₅₀)	logP Slipper	Тип действия
48	Диметоат	C5 H12 N O3 P S2	60-51-5	3.96	1.87	АнтиХЭ
49	Диазинон	C12 H21 N2 O3 P S	333-41-5	3.92	4.19	АнтиХЭ
50	Дихлофос	C4 H7 Cl2 O4 P	62-73-7	5.66	1.47	АнтиХЭ
51	Дизопропилфлюорофосфат	C6 H14 F O3 P	55-91-4	3.77	0.03	АнтиХЭ
52	Малаксон	C10 H19 O7 P S	1634-78-2	4.05	-0.11	АнтиХЭ
53	Карбофуран	C12 H15 N O3	1563-66-2	3.09	2.40	АнтиХЭ
54	Никотин	C10 H14 N2	54-11-5	2.11	0.04	АнтиХЭ
55	Оксотреморин	C12 H18 N2 O	70-22-4	1.53	-0.03	АнтиХЭ
56	Паратион	C10 H14 N O5 P S	56-38-2	4.91	3.57	АнтиХЭ
57	Диэтил-р-нитрофенилфосфат	C10 H14 N O6 P	311-45-5	4.61	2.04	АнтиХЭ
58	Спартейн	C15 H26 N2	90-39-1	2.25	-0.67	АнтиХЭ
59	Пропоксур	C11 H15 N O3	114-26-1	2.91	1.74	АнтиХЭ
60	Пилокарпин	C11 H16 N2 O2	92-13-7	0.77	1.75	АнтиХЭ
61	Трихлорофон	C4 H8 Cl3 O4 P	52-68-6	4.26	0.61	АнтиХЭ
62	Тетраэтилдифосфат	C8 H20 O7 P2	107-49-3	5.51	-0.54	АнтиХЭ
63	Такрин	C13 H14 N2	321-64-2	2.81	0.90	АнтиХЭ
64	Фенитропион	C9 H12 N O5 P S	122-14-5	3.98	3.31	АнтиХЭ
65	Физостигмин	C15 H21 N3 O2	57-47-6	3.06	1.66	АнтиХЭ
66	Фосдрин	C7 H13 O6 P	298-01-1	4.73	-0.45	АнтиХЭ
67	Хлорпирифос	C9 H11 Cl3 N O3 P S	2921-88-2	4.29	5.05	АнтиХЭ
68	Зарин	C4 H10 F O2 P	107-44-8	4.60	-0.70	АнтиХЭ
69	Зоман	C7 H16 F O2 P	96-64-0	5.18	0.86	АнтиХЭ
70	S-2(Дизопропиламино)этил О-этил метилфосфонгоат	C11 H26 N O2 P S	50782-69-9	4.35	0.05	АнтиХЭ

Продолжение таблицы 1

71	N-(диметокси)фосфорилокси) этаннимидонилхлорид	C4 H9 Cl N O4 P		3.22	0.18	АнтиХЭ
72	N-(диэтоксифосфорилокси) этаннимидонилхлорид	C6 H13 Cl N O4 P		2.51	0.71	АнтиХЭ
73	N-(дипропилокси)фосфорилокси) этаннимидонилхлорид	C8 H17 Cl N O4 P		2.87	1.42	АнтиХЭ
74	N-(диизопропилокси)фосфорилокси) этаннимидонилхлорид	C8 H17 Cl N O4 P		2.92	1.74	АнтиХЭ
75	N-(изо-бутоксиметил)фосфорилокси) этаннимидонилхлорид	C7 H15 Cl N O4 P		2.76	1.25	АнтиХЭ
76	N-(ди-втор-бутоксифосфорилокси) этаннимидонилхлорид	C10 H21 Cl N O4 P		2.30	2.40	АнтиХЭ
77	N-(дипентокси)фосфорилокси) этаннимидонилхлорид	C12 H25 Cl N O4 P		нет данных	3.32	АнтиХЭ
78	N-(дигексокси)фосфорилокси) этаннимидонилхлорид	C14 H29 Cl N O4 P		нет данных	4.33	АнтиХЭ
79	N-(ди-изобутоксифосфорилокси) этаннимидонилхлорид	C10 H21 Cl N O4 P		2.06	2.63	АнтиХЭ
80	Алдикарб	C7 H14 N2 O2 S	116-06-3	2.90	1.31	Специфический
81	Ардван	C14 H21 Cl F2 N2 O2		нет данных	3.22	Специфический
82	Гептахлор	C10 H5 Cl7	76-44-8	0.95	5.32	Специфический
83	Беметрид	C8 H13 N O2	64-65-3	-0.34	0.06	Специфический
84	4-Amino-p-Butyric Acid	C4 H9 N O2	56-12-2	-0.97	-4.68	Специфический
85	Пентаметилентетразол	C6 H10 N4	54-95-5	0.29	1.04	Специфический
86	Лобелин	C22 H27 N O2	90-69-7	3.49	2.85	Специфический
87	Норборнан	C7 H12	279-23-2	1.19	2.46	Специфический
88	Пикротоксин	C15 H18 O7	124-87-8	1.46	-1.10	Специфический
89	Стрихнин	C21 H22 N2 O2	57-24-9	1.35	1.01	Специфический

**Избранные труды доктора химических наук, профессора О.А. Раевского
в области молекулярного компьютерного дизайна**

Продолжение таблицы 1

	Наименование	Брутто-формула	CAS	log(1/LC ₅₀)	logP Slipper	Тип действия
90	Тиосемикарбазид	CН5 N3 S	79-19-6	0.72	-1.56	Специфический
91	t-Бутил-бисциклофосфат	C8 H15 O4 P		1.99	1.19	Специфический
92	Фенилсилатран	C12 H17 N O3 Si	2097-19-0	0.84	нет параметр.	Специфический
93	Аллетрин	C19 H26 O3	584-79-2	1.85	4.96	Специфический
94	Дельтаметрин	C22 H19 Br2 N O3	52918-63-5	1.02	6.40	Специфический
95	Каделтрин	C28 H26 O5 S	58769-20-3	1.28	7.19	Специфический
96	Перметрин	C21 H20 Cl2 O3	52341-33-0	3.89	6.47	Специфический
97	Ресметрин	C22 H26 O3	10453-86-8	1.67	6.28	Специфический
98	Фенвалерат	C25 H22 Cl N O3	51630-58-1	4.54	6.55	Специфический
99	Фенотрин	C23 H26 O3	26002-80-2	3.84	6.53	Специфический
100	Фенпролагрин	C22 H23 N O3	39515-41-8	2.70	5.24	Специфический
101	Циперметрин	C22 H19 Cl2 N O3	52315-07-8	3.87	5.93	Специфический
102	Цифенотрин	C24 H25 N O3	39515-40-7	4.88	5.97	Специфический
103	Бензапирен	C20 H12	50-32-8	2.87	6.01	Специфический
104	Бензаграцен	C18 H12	56-55-3	3.00	5.27	Специфический
105	Бензобарбитал	C19 H16 N2 O4	744-80-9	-0.13	1.69	Специфический
106	Бармамил	C11 H18 N2 O3	57-43-2	-0.00	1.71	Специфический
107	Диметилсульфоксид	C2 H6 O S	67-68-5	0.27	-0.87	Специфический
108	Диазепам	C16 H13 Cl N2 O	439-14-5	1.43	2.56	Специфический
109	Кетамин	C13 H16 Cl N O	6740-88-1	1.22	1.82	Специфический
110	Кофеин	C8 H10 N4 O2	58-08-2	-0.95	-0.37	Специфический
111	Фенобарбитал	C12 H12 N2 O3	50-06-6	-0.46	0.92	Специфический
112	Барбитал	C8 H12 N2 O3	57-44-3	-0.81	0.29	Специфический
113	Феназепам	C15 H10 Br Cl N2 O	51753-57-2	1.35	3.49	Специфический

Окончание таблицы 1

114	Хинин	C20 H24 N2 O2	130-95-0	0.80	2.88	Специфический
115	Аллиловый спирт	C3 H6 O	107-18-6	-0.06	0.23	Специфический
116	Этилацетат	C4 H8 O2	141-78-6	-0.84	0.61	Специфический
117	Этилпропионат	C5 H10 O2	105-37-3	-0.35	1.04	Специфический
118	Аллиламин	C3 H7 N	107-11-9	0.27	-2.22	Специфический
119	Салицилальдегид	C7 H6 O2	90-02-8	1.32	1.84	Специфический

LC₅₀ – летальная концентрация вещества для 50% рычков *Daphnia magna* (инкубация 24 часа);

Общий тип токсического действия;

Полярный – действие по типу полярного наркоза;

АнтиХЭ – ингибиторы Холинэстераз;

Специфический – иной тип действия.

Табл. 2. Дискриминация веществ разного типа токсического действия друг от друга

	Тип действия	Верное распознавание № структурного соседа					Число ближайших соседей					Всего	
		1-й	2-й	3-й	4-й	5-й	1	1-3	1-5				
1	Общий	NN	19	21	17	15	22	21	19	24			
2		%	91,7%	79,2%	87,5%	70,8%	91,7%	87,5%	79,2%				
3	Иной тип	NN	86	89	84	85	86	88	91	95			
4		%	90,5%	94,7%	88,4%	89,5%	90,5%	92,6%	95,8%				
5	Полярный	NN	15	13	15	14	15	14	16	18			
6		%	83,3	72,2	83,3	77,7	83,3	77,8	88,9				
7	Иной тип	NN	94	94	94	88	94	94	93	101			
8		%	93,1%	93,1%	93,1%	87,1%	93,1%	93,1%	93,1%				
9	Общий	NN	24	21	20	19	24	22	22	24			

Тип действия		Верное распознавание № структурного соседа						Число ближайших соседей			Всего
10	%	100,00%	87,5%	83,3%	79,2%	75,00%	100,00%	91,7%	91,7%	91,7%	
11	NN	16	17	18	15	14	16	17	17	17	18
12	%	89,9%	94,4%	100,00%	83,3%	77,8%	88,9%	94,4%	94,4%	94,4%	
13	NN	30	28	28	24	24	30	29	26	26	37
14	%	81,1%	75,7%	75,7%	64,9%	64,9%	81,1%	78,4%	70,3%	70,3%	
15	NN	75	76	75	78	80	75	77	78	78	82
16	%	91,5%	92,7%	91,5%	95,1%	97,6%	91,5%	93,9%	95,1%	95,1%	
17	NN	26	23	20	21	23	26	19	20	20	40
18	%	65,0%	57,5%	50,0%	52,5%	57,5%	65,0%	47,5%	50,0%	50,0%	
19	NN	73	70	68	69	69	73	71	75	75	79
20	%	92,4%	88,6%	86,1%	87,3%	87,3%	92,4%	89,9%	94,9%	94,9%	

NN – число верно классифицированных веществ;

% – процент от числа соединений с данным типом действия;

№ структурного соседа – номер соединения в ранжированном, по уменьшению индекса Танимото, ряду;

Число ближайших соседей – номера соединений, из ранжированного ряда, с которыми производится сравнение

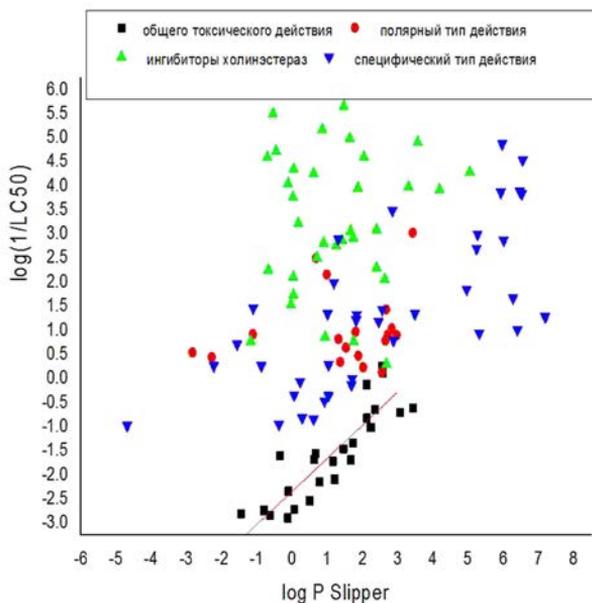


Рис. 1. Зависимость между токсичностью ($\log(1/LC_{50})$) и рассчитанной липофильностью ($\log P$ Slipper) для веществ с различным типом токсического действия на рачок *Daphnia magna*

ЛИТЕРАТУРА

1. Peters R.H. and Bernardi R.D.E. (Editors). DAPHNIA. Istituto Italiano Di Idriologia. 1987. v.45, Verbania, Pallanza.
2. Прозоровский В.Б. Фармакол. и токсикол. Т. 51. №4. 1978. 497–502.
3. Raevsky O.A., Grigoriev V.Yu, Trepalin S.V. HYBOT. Registration by Russian State Patent Agency, № 990090, 26.02.1999.
4. Raevsky O.A., Phys J. Org. Chem. Vol. 10. N5. 1997. 405–413.
5. Gerasimenko V.A., Trepalin S.V., Raevsky O.A. Molecular Modelling and Prediction of Bioactivity. eds. K. Gundertofe and F. Jorgensen, Kluwer Academic / Plenum Publ. 2000, pp. 423–424.
6. Raevsky O.A., Trepalin S.V., Trepalina E.P., Raevskaya O.E. MOLDIVS, Registration by Russian State Patent Agency, № 990093, 26.02.1999.
7. Raevsky O.A., Trepalin S.V., Trepalina E.P. Registration by Russian State Patent Agency, № 990089, 26.02.1999.
8. Lipnick L. In Practical Application of Quantitative Structure-Activity Relationships (QSAR) in Environmental Chemistry and Toxicology. W.

**Избранные труды доктора химических наук, профессора О.А. Раевского
в области молекулярного компьютерного дизайна**

Karcher and J. Devillers (Eds), Kluwer Academic Publishers, Dordrecht, 1990.
pp. 281–293.

9. *Raevsky O.A.* SAR & QSAR in Environmental Research. Vol. 12. 2001.
367–381.

10. *Raevsky O.A.* Minireview in Medicinal Chemistry. Vol. 4, N10, 2004.
1041–1052.

О.А. Раевский¹, В.Ю. Григорьев¹, О.В. Тихонова^{1,2}

РАЗВИТИЕ МОДЕЛЕЙ ВЗАИМОСВЯЗИ СТРУКТУРЫ И ТОКСИЧЕСКОГО ДЕЙСТВИЯ ХИМИЧЕСКИХ СОЕДИНЕНИЙ ПО ОТНОШЕНИЮ К GUPPY

*¹Институт физиологически активных веществ РАН,
Черноголовка, Московская область*

*²Научно-исследовательский институт биомедицинской химии
имени В.Н. Ореховича, Москва*

Химические соединения, включая лекарства, обладают различными побочными эффектами, в том числе токсичностью по отношению к живым организмам биосферы. Поэтому важно установить и понять взаимосвязь структуры и токсичности химических соединений для того, чтобы предсказать такие свойства для новых веществ.

Имеется по крайней мере две недавние важные публикации, в которых рассматриваются подходы для установления количественной взаимосвязи между структурой и активностью (КССА) для предсказания токсичности химических соединений по отношению к водным организмам. Это прежде всего система классификации химических соединений на четыре группы, основанная на наборе правил [1]. Используя такой подход, можно рассчитать токсичность химических соединений, принадлежащих к одному из четырех классов, на основе коэффициента распределения в системе октанол – вода ($\log P$).

Другой подход использует большой набор CODESSA дескрипторов (включая конститутивные, топологические, квантовохимические, физико-химические) для построения регрессионных соотношений между ними и токсичностью химических соединений [2].

В публикации [3] нами использована обучающая выборка [2] по токсичности химических соединений четырех типов действия: общего токсического (неполярный наркоз), полярного наркоза, реактивного и специфического типов действия по отношению к *Guppy* для построения КССА-моделей между дескрипторами, рассчитываемыми программой HYBOT[4], и токсичностью этих соединений.

В настоящей работе представлены существенно расширенные стабильные предсказательные классификационные модели для соединений, обладающих различными механизмами токсичности, и количественные КССА модели для рядов соединений, обладающих одним и тем же механизмом действия.

Экспериментальная часть

Данные по токсичности 293 химических соединений были взяты из работы [2] и выражены в виде $\log(1/LC_{50})(LC_{50}, \text{ ммоль/л}, 96 \text{ ч})$. Для описания молекулярной структуры был рассчитан ряд из 95 электронных, топологических и физико-химических дескрипторов с помощью компьютерных программ NYBOT, DNESTR, SLIPPER, созданных в отделе компьютерного молекулярного дизайна ИФАВ РАН [4, 5], и программного комплекса SYBYL[6].

Классификация химических соединений по механизму действия проведена с использованием дискриминантного анализа с применением как параметрических, так и непараметрических методов. В качестве параметрического использован метод линейного дискриминантного анализа (ЛДА), а в качестве непараметрического – дискриминантный анализ на основе метода «ближайшего соседа» (БС) с применением программы DNESTR. Классификация соединений проведена также с помощью метода «структурного соседа» (СС) на основе меры сходства молекул – индексов Танимото (ТС), рассчитанных с помощью программы MOLDIVS[7]. Отнесение соединений к разным типам действия с использованием СС производилось на основе упорядоченных по убыванию ТС пяти ближайших структурно сходных соединений. В случае использования одного соседа (1, 2, 3, 4 или 5) классифицируемому соединению приписывался тот же тип действия, что и у структурно близкого соединения. В случае нескольких соседей (1–3 или 1–5) правильным распознаванием считалась ситуация, когда большая часть соседей, а именно 2 из 3 и больше, 3 из 5 и больше, имели рассматриваемый тип действия.

Регрессионный анализ с применением стандартных процедур проводили как для ограниченных комбинаций дескрипторов, так и для всех возможных сочетаний из одного, двух и трех независимых переменных. В качестве критерия линейной независимости дескрипторов использовали коэффициент линейной корреляции (R). Переменные считались линейно независимыми при $R < 0,6$. Качество регрессионных уравнений оценивалось на основе числа соединений (N), коэффициента детерминации (R²), стандартного отклонения (SD) и критерия Фишера (F). Погрешности параметров уравнений оценивались в виде стандартных ошибок.

Результаты и их обсуждение

Поскольку примерно 60% веществ, загрязняющих окружающую среду, действуют по механизму общего токсического действия (неполярный нар-

коз), на первом этапе работы была поставлена задача дискриминации соединений общего токсического действия по отношению ко всем остальным. Таблица 1 содержит результаты классификации химических соединений на две группы. Первые 9 строчек таблицы относятся к классификации соединений общего токсического действия по отношению ко всем остальным с помощью методов ЛДА, БС и СС. Использование ЛДА позволило правильно классифицировать только 53 соединения этого типа из 90. Несколько лучшие результаты дало применение метода БС (71 соединение из 90 было правильно классифицировано). В случае же использования метода СС правильная идентификация соединений, проявляющих как общий токсический эффект, так и другие типы действия, оказалась на достаточно высоком уровне. Так, при использовании в качестве ближайшего структурно сходного соединения с максимальным ТС (1 сосед), 84 из 90 соединений общего токсического действия и 184 из 203 соединений иного типа действия были классифицированы правильно. При использовании в качестве структурно родственных соединений с меньшими значениями ТС процент правильного распознавания уменьшается. Так, в случае второго структурного соседа число правильно распознанных соединений с общим типом токсического действия равно 79 из 90, в случае третьего структурного соседа – 80 из 90, четвертого – 76 из 90 и пятого – 71 из 90. Уменьшается при этом и число правильно распознанных соединений других типов действия: для второго соседа – 180 из 203, третьего – 185 из 203, четвертого – 181 из 203 и пятого – 175 из 203. А в случае использования информации о трех или пяти соседях качество распознавания улучшается за счет использования статистического подхода. Таким образом, применение метода СС обеспечило правильное распознавание соединений общего токсического действия на уровне 95%, а соединений иного типа действия – на уровне 90 %.

Табл. 1. Результаты классификации химических соединений на две группы

Тип действия		Классификация		Распознавание, %			
NN..... Метод		Сосед					
	Группа 1	Группа 2		Группа 1	Группа 2	Группа 1	Группа 2
1	Неполярный наркоз	Остальное	ЛДА	53/37	169/34	58,9	83,3
2	Неполярный наркоз	Остальное	БС	3 71/19	183/20	78,9	90,1
3	Неполярный наркоз	Остальное	СС	1 84/6	184/19	93,3	90,6
4	Неполярный наркоз	Остальное	СС	2 79/11	180/23	87,8	88,7
5	Неполярный наркоз	Остальное	СС	3 80/10	185/18	88,9	91,1

**Избранные труды доктора химических наук, профессора О.А. Раевского
в области молекулярного компьютерного дизайна**

Продолжение таблицы 1

Тип действия		Классификация			Распознавание, %		
NN.....		Метод			Сосед		
	Группа 1	Группа 2		Группа 1	Группа 2	Группа 1	Группа 2
6	Неполярный наркоз	Остальное	СС 4	76/14	181 /22	84,4	89,2
7	Неполярный наркоз	Остальное	СС 5	71 /19	175/28	78,9	86,2
8	Неполярный наркоз	Остальное	СС 1-3	86/4	182/21	95,6	89,7
9	Неполярный наркоз	Остальное	СС 1-5	86/4	184/19	95,6	90,6
10	Полярный наркоз	Остальное	ЛДА	101/20	153/19	83,5	89,0
11	Полярный наркоз	Остальное	БС 3	114/7	157/15	94,2	91,3
12	Полярный наркоз	Остальное	СС 1	113/8	160/12	93,4	93,0
13	Полярный наркоз	Остальное	СС 2	111/10	152/20	91,7	88,4
14	Полярный наркоз	Остальное	СС 3	115/6	157/15	95,0	91,3
15	Полярный наркоз	Остальное	СС 4	108/13	149/23	89,3	86,6
16	Полярный наркоз	Остальное	СС 5	109 /12	151/21	90,1	87,8
17	Полярный наркоз	Остальное	СС 1-3	119/2	159/13	98,4	92,4
18	Полярный наркоз	Остальное	СС 1-5	119/2	155/17	98,4	90,1
19	Реактивный	Остальное	ЛДА	6/45	239/3	11,8	98,8
20	Реактивный	Остальное	БС 3	26/25	228/14	51,0	94,2
21	Реактивный	Остальное	СС 1	37/14	235/7	72,6	97,1
22	Реактивный	Остальное	СС 2	30/21	229/13	58,8	94,6
23	Реактивный	Остальное	СС 3	33/18	233/9	64,7	96,3
24	Реактивный	Остальное	СС 4	31/20	230/12	60,8	95,0
25	Реактивный	Остальное	СС 5	26/25	223/19	51,0	92,2
26	Реактивный	Остальное	СС 1-3	32/19	239/3	62,8	98,8
27	Реактивный	Остальное	СС 1-5	31/20	240/2	60,8	99,2
28	Специфич-ный	Остальное	ЛДА	22/9	262/0	71,0	100,0
29	Специфич-ный	Остальное	БС 3	19 /12	262/0	61,3	100,0

Окончание таблицы 1

Тип действия		Классификация				Распознавание, %		
NN..... Метод		Сосед						
	Группа 1	Группа 2			Группа 1	Группа 2	Группа 1	Группа 2
30	Специфич- ный	Остальное	СС	1	19 /12	260/2	61,3	99,2
31	Специфич- ный	Остальное	СС	2	15/16	260/2	48,4	99,2
32	Специфич- ный	Остальное	СС	3	17 /14	256/6	54,8	97,7
33	Специфич- ный	Остальное	СС	4	15/16	256/6	48,4	97,7
34	Специфич- ный	Остальное	СС	5	14 /17	257/5	45,2	98,1
35	Специфич- ный	Остальное	СС	1-3	16/15	262/0	51,6	100,0
36	Специфич- ный	Остальное	СС	1-5	16/15	262/0	51,6	100,0

Примечание. В колонке «Классификация» первое число означает количество правильно распознанных молекул, второе число – количество неправильно распознанных молекул.

После построения устойчивой модели, дискриминирующей соединения общего токсического действия по отношению к соединениям иных типов действия, целесообразно поставить вопрос об установлении количественной взаимосвязи между токсичностью $\log(1/LC_{50})$ и дескрипторами, описывающими структуру соединений этого типа действия. В таблице 2 (1–3 строки) приведены регрессионные уравнения, полученные на основе экспериментальной липофильности $\log P_{\text{эксп}}$ (модель 1), расчетных значений липофильности $\log P_{\text{расч}}$ и композитного дескриптора $\sum(E_d/\alpha)$, связанного с энтальпийным протонодонорным фактором (E_d) и поляризуемостью молекулы (α) (модель 2), а также свободноэнергетического протонакцепторного фактора (C_d), индекса валентной связности первого порядка ($\chi 1v$) и композитного дескриптора, полученного на основе энергий высшей занятой молекулярной орбитали (ВЗМО) и нижней свободной молекулярной орбитали (НСМО) (модель 3). Эти модели были получены путем перебора всех возможных комбинаций 95 дескрипторов с последующей селекцией на основе коэффициента корреляции. Статистические характеристики указанных трех уравнений очень хорошие (R^2 на уровне 0,95–0,97, а стандартное отклонение на уровне 0,27–0,33, что соответствует уровню экспериментальной ошибки определения токсичности). Тем не менее вышеуказанные модели носят формальный характер и не имеют ясной физико-химической интерпретации. Так, рассчитанное значение коэффициента распределения веще-

ства в системе октанол – вода; E_d – энтальпийный протонодонорный фактор; α – поляризуемость молекулы; C_a – свободноэнергетический протоноакцепторный фактор; $chl v$ – индекс валентной связности первого порядка; $Hn = (V3MO - H3MO)/2$, $V3MO$ – высшая занятая молекулярная орбиталь, $H3MO$ – низшая свободная молекулярная орбиталь; ESC – коэффициент вариации электроотрицательности; $Nv1$ – число вершин со степенью 1; ESM – среднее значение электроотрицательности; q^+ – положительный атомный заряд; Ca (макс.) – максимальное значение свободноэнергетического протоноакцепторного фактора; MW – молекулярный вес; $CICO$ – комплементарное информационное содержание 0 порядка; IDW_{cp} – средний информационный индекс распределения расстояний в молекуле в соответствии с их величиной; $abs(q)$ – абсолютное значение атомного заряда; E_a (макс.) – максимальное значение энтальпийного протоноакцепторного фактора, например, липофильность является композитным дескриптором, связанным как с размером молекулы, так и способностью молекул к образованию водородных связей [8], а топологические и квантовохимические дескрипторы трудно интерпретировать в рамках межмолекулярных взаимодействий, связанных с токсичностью соединений.

По этой причине были также исследованы регрессионные уравнения, связывающие токсичность и физико-химические дескрипторы, рассчитанные с помощью программы HYBOT. Результаты этой работы приведены в таблице 3 (строки 1–2). Очевидно, что уравнение, представленное в строке 1 и включающее в себя два физико-химических дескриптора (α и $\sum C_a$), имеет ясный физический смысл и одновременно практически такие же статистические характеристики, что и вышеуказанные модели. Согласно этому уравнению токсичность химических веществ, действующих по отношению к *Girru* по механизму неполярного наркоза, является результатом двух противоположно действующих вкладов: молекулярной поляризуемости, рост которой приводит к увеличению токсичности, и протоноакцепторной способности, увеличение которой приводит к уменьшению токсичности вещества. Таким образом, модель 1 (таблица 2), построенная на основе двух вышеуказанных дескрипторов, дает возможность не только количественно предсказывать меру токсичности, но и оптимизировать токсичность новых химических соединений на основе учета вкладов стерических взаимодействий и водородной связи.

На следующем этапе работы проведена классификация соединений, действующих по механизму полярного наркоза, по отношению к соединениям, действующим по другим механизмам, и созданы количественные модели токсичности соединений с полярным механизмом действия по отношению к *Girru*. В таблице 1 представлены результаты дискриминации соединений, действующих по механизму полярного наркоза, по отношению ко всем остальным на основе различных подходов. Из представленных данных очевидно, что все три использованных подхода (ЛДА, БС, СС) позволяют

Табл. 2. Параметры и статистика регрессионных уравнений, полученных путем перебора всех возможных комбинаций дескрипторов

NN	Тип действия	Уравнение	N	R ²	SD	F
1	Неполярный наркоз	$\log(1/LC_{50}) = -1,71(0,06) + 0,85(0,02)\log P_{\text{эксип}}$	90	0,948	0,33	1607
2	Неполярный наркоз	$\log(1/LC_{50}) = -1,58(0,07) + 0,86(0,02)\log P_{\text{раств}} + 0,38(0,09)\sum(E_d/\alpha)$	90	0,957	0,31	958
3	Неполярный наркоз	$\log(1/LC_{50}) = 0,21(0,21) - 0,74(0,03)\sum_a + 0,97(0,03)\text{chl}iv + 0,33(0,03)Hn$	90	0,967	0,27	851
4	Полярный наркоз	$\log(1/LC_{50}) = -0,88(0,07) + 0,76(0,03)\log P_{\text{жп}}$	121	0,880	0,35	871
5	Полярный наркоз	$\log(1/LC_{50}) = 0,64(0,37) + 0,73(0,02)\log P_{\text{эксип}} - 5,29(1,26)ESC$	121	0,895	0,33	506
6	Полярный наркоз	$\log(1/LC_{50}) = -4,14(0,64) - 0,090(0,028)Nv1 + 0,94(0,18)ESM + 0,76(0,02)\log P_{\text{эксип}}$	121	0,904	0,32	366
7	Реактивный	$\log(1/LC_{50}) = 3,20(0,23) - 44,1(5,9)\sum(q+/\alpha)$	51	0,530	0,63	55,2
8	Реактивный	$\log(1/LC_{50}) = 1,79(0,32) - 0,90(0,16)C_a(\text{макс}) + 0,0073(0,0012)MW$	51	0,700	0,51	56,0
9	Реактивный	$\log(1/LC_{50}) = 1,20(0,26) - 1,18(0,14)CICO + 0,65(0,06)IDW_{\text{ср}} + 0,46(0,06)\log P_{\text{раств}}$	51	0,765	0,45	50,9
10	Специфичный	$\log(1/LC_{50}) = 0,07(0,46) + 0,12(0,02)\alpha$	31	0,582	0,88	40,3
11	Специфичный	$\log(1/LC_{50}) = 0,46(0,42) + 0,16(0,02)\alpha - 0,63(0,19)E(\text{abs}(q))$	31	0,697	0,76	32,2
12	Специфичный	$\log(1/LC_{50}) = 1,75(0,65) + 1,52(0,24)Ea(\text{макс}) - 13,8(1,9)\sum_a + 0,35(0,09)Nv1$	31	0,771	0,67	30,4

Примечание. $\log P_{\text{эксип}}$ – экспериментальное значение коэффициента распределения вещества в системе октанол – вода; $\log P_{\text{раств}}$ – рассчитанное значение коэффициента распределения вещества в системе октанол – вода; E_d – энталпийный протодонорный фактор; α – поляризуемость молекулы; C_a – свободноэнергетический протонакцепторный фактор; $\text{chl}iv$ – индекс валентной связности первого порядка; $Hn = (V3MO - HSMO)/2$, $V3MO$ – высшая занятая молекулярная орбиталь, $HSMO$ – низшая свободная молекулярная орбиталь; ESC – коэффициент вариации электроотрицательности; $Nv1$ – число вершин со степенью 1; ESM – среднее значение электроотрицательности; q^+ – положительный атомный заряд; $C_a(\text{макс.})$ – максимальное значение свободноэнергетического протонакцепторного фактора; MW – молекулярный вес; $CICO$ – комплементарное информационное содержание 0 порядка; $IDW_{\text{ср}}$ – средний информационный индекс распределения расстояний в молекуле в соответствии с их величиной; $\text{abs}(q)$ – абсолютное значение атомного заряда; $E_a(\text{макс.})$ – максимальное значение энталпийного протонакцепторного фактора.

достаточно четко идентифицировать соединения с механизмом полярного наркоза и соединений с иными механизмами действия достаточно четко. Наилучшие результаты классификации получены при использовании метода СС с учетом трех соседей. В этом случае 119 из 121 (98,4 %) соединений с механизмом полярного наркоза классифицируются правильно и 159 из 172 (92,4 %) соединений иного механизма действия также классифицируются правильно.

Что касается количественных соотношений между токсичностью соединений с полярным механизмом действия и дескрипторами молекулярной структуры, то можно отметить, что в случае перебора всех возможных комбинаций дескрипторов удается построить одно-, двух- и трехпараметровые модели с достаточно хорошими статистическими характеристиками. Однако эти формализованные модели (строчки 4–6 таблицы 2) содержат экспериментально определяемый параметр ($\log P_{\text{эксп}}$), который имеет ограниченную область использования, поэтому были использованы дескрипторы, рассчитываемые с помощью программы НУВОТ. Полученная модель из трех параметров (строчка 4, таблица 3) хотя и немного уступает соответствующей формализованной модели по статистическим параметрам, но зато имеет ясный физический смысл.

Ввиду обширной дискуссии в литературе о том, существует ли различие между механизмами общего токсического действия и полярного наркоза [9–11], было проведено сопоставление полученных регрессионных уравнений для соединений указанных выше типов действия:

$$\log(1/LC_{50}) = -1,06 + 0,20\alpha - 0,722C_a - 0,102C_d \text{ (полярный наркоз);}$$

$$\log(1/LC_{50}) = -2,09 + 0,26\alpha - 0,81EC_a \text{ (неполярный наркоз);}$$

$\Delta\log(1/LC_{50}) = 1,03 - 0,06\Delta\alpha + 0,09\Delta\sum C_a - 0,102\sum\Delta C_d$ (разность между полярным и неполярным наркозом).

Любопытно, что разность уравнений для полярного и неполярного наркоза практически совпадает с уравнением растворимости соединений в октаноле [12]:

$$\log S_0 = 1,06 - 0,063\alpha + 0,03\sum C_a - 0,14\sum C_d.$$

Таким образом, можно утверждать, что для соединений с механизмом полярного наркоза характерен существенный вклад общего токсического действия, а также дополнительный вклад, связанный с движением молекул в мембранную фазу.

КССА-модели, полученные для соединений, проявляющих реактивный и специфический типы действия, имеют гораздо более скромные статистические характеристики по сравнению с рассмотренными выше моделями

Табл. 3. Параметры и статистика регрессионных уравнений, полученных в рамках физико-химической модели на основе дескрипторов, рассчитанных с помощью программы НУВОТ

NN	Тип действия	Уравнение	N	R ²	SD	F
1	Неполярный наркоз	$\log(1/LC_{50}) = -2,09(0,11) + 0,26(0,01)\alpha - 0,81(0,03)\sum C_a$	90	0,956	0,31	940
2	Неполярный наркоз	$\log(1/LC_{50}) = -2,09(0,11) + 0,26(0,01)\alpha - 0,80(0,04)\sum C_a + 0,02(0,05)\sum C_d$	90	0,956	0,31	621
3	Полярный наркоз	$\log(1/LC_{50}) = -0,87(0,17) + 0,20(0,01)\alpha - 0,68(0,05)\sum C_a$	121	0,859	0,38	359
4	Полярный наркоз	$\log(1/LC_{50}) = -1,06(0,17) + 0,20(0,01)\alpha - 0,72(0,05)\sum C_a - 0,10(0,03)\sum C_d$	121	0,872	0,36	265
5	Реактивный	$\log(1/LC_{50}) = 0,60(0,38) + 0,11(0,02)\alpha - 0,29(0,16)\sum C_a$	51	0,428	0,70	18,0
6	Реактивный	$\log(1/LC_{50}) = 0,55(0,38) + 0,11(0,02)\alpha - 0,29(0,16)\sum C_a + 0,08(0,10)\sum C_d$	51	0,437	0,70	12,1
7	Специфичный	$\log(1/LC_{50}) = 0,39(0,51) + 0,13(0,02)\alpha - 0,18(0,13)\sum C_a$	31	0,607	0,87	21,6
8	Специфичный	$\log(1/LC_{50}) = 0,52(0,62) + 0,12(0,02)\alpha - 0,17(0,13)\sum C_a + 0,04(0,11)\sum C_d$	31	0,609	0,88	14,0

неполярного и полярного наркоза. Дискриминация соединений с реактивным типом действия по отношению ко всем остальным механизмам действия привела к удовлетворительным результатам только в случае использования одного соседа в методе СС (в этом случае было правильно классифицировано 37 из 51 соединения, что составило 72,6%, и 235 из 342, или 97,1%) соединений с иными типами действия). Другие использованные методы классификации дают гораздо худшие результаты. Формализованная модель из трех параметров для соединений с реактивным механизмом действия имеет неплохие статистические параметры (таблица 2, строка 9), а модель, основанная на НУВОТ дескрипторах, оказалась гораздо хуже.

В случае классификации соединений специфического механизма действия по отношению ко всем остальным наилучшие результаты достигнуты с помощью метода ЛДА: 22 из 31 соединения специфического механизма действия (71,0%) и все 262 (100%) соединения других типов действия были классифицированы правильно.

Что касается количественных соотношений, формализованная трехпараметровая модель (таблица 2, строка 12) имеет достаточно хорошие статистические характеристики. Однако эта модель, как и соответствующая модель для реактивного типа действия содержит кроме физико-химических дескрипторов и иные дескрипторы и не имеет ясного физического смысла.

Таким образом, в данной работе удалось построить четкие классификационные модели для дискриминации соединений общего токсического действия по отношению ко всем остальным соединениям с другими типами действия и модели для дискриминации соединений полярного типа действия по отношению ко всем остальным. Классификационные модели для дискриминации соединений с реактивным и специфическим типом действия уступают по качеству. Что касается количественных соотношений между токсичностью соединений по отношению к *Guppy* и дескрипторами, то можно отметить хорошую двухпараметровую модель, основанную на молекулярной поляризуемости и протоноакцепторной способности молекул, а также выявленную разницу между механизмом полярного типа действия и механизмом общего токсического действия. Количественные модели для описания токсичности соединений реактивного и специфического типа действия авторы этой работы рассматривают как формализованные и предварительные. Дальнейшее улучшение таких моделей возможно при более углубленном описании структуры с привлечением не только двухмерных, но и трехмерных дескрипторов.

Работа выполнена при финансовой поддержке МНТЦ (проект 0888).

ЛИТЕРАТУРА

1. *Verhaar H.J.M., van Leeuwen C.J. and Hermens J.L.M.* Chemosphere. 1992. Vol. 25(4). P. 471–491.
2. *Katritzky A.R., Tatham D.B. and Maran U.* J. Chem. Inf. Comput. Sci. 2001. Vol. 41(5). P. 1162–1176.
3. *O.A. Raevsky and J.C. Dearden* // SAR and QSAR in Environmental Research. 2004. Vol. 15(5–6). P. 433–448.
4. *Раевский О.А., Скворцов В.С., Григорьев В.Ю., Трепалин С. В.* НУВОТ. Свидетельство РОСПАТЕНТа об официальной регистрации программы для ЭВМ №2002610496. 04.04.2002.
5. *Раевский О.А., Трепалин С.В., Трепалина Е.П.* SLIPPER. Свидетельство РОСПАТЕНТа об официальной регистрации программы для ЭВМ № 990089. 26.02.1999.
6. <http://www.tripos.com/>, Sybyl 6.7, Tripos Inc., 1699 South Hanley Road, St Louis, Missouri, 63144, USA.
7. *Раевский О.А., Герасименко В.А., Трепалин С.В.* MOL-DIVS. Свидетельство РОСПАТЕНТа об официальной регистрации программы для ЭВМ № 990093. 26.02.1999.
8. *Раевский О.А., Григорьев В.Ю.* Хим.-фарм. журн. 1999. Т. 33(5). С. 46–49.
9. *Roberts D.W. and Costello J.F.* QSAR Comb. Sci. 2003. Vol. 22. p.226–233.
10. *Vaes W.H.J., Ramos E.U., Verhaar H.J.M. and Hermens J. L. M.* Environ. Toxicol. Chem. 1998. Vol. 17. p. 1380–1384.
11. *Dearden J.C., Cronin M.T.D., Zhao Y.-H. and Raevsky O. A.* Quant. Struct.-Act. Relat. 2000. Vol. 19. p. 3–9.
12. *Raevsky O.A., Tonkopii V.D., Dearden J.C. and Schaper K.-J.* Posters of the 12th International Workshop on Quantitative Structure-Activity Relationships in Environmental Toxicology, Lyon, 2006, p. 114.

В.Ю. Григорьев, О.А. Раевский

ФРАКТАЛЬНАЯ РАЗМЕРНОСТЬ ГИСТОГРАММ МЕЖАТОМНЫХ РАССТОЯНИЙ – НОВЫЙ 3D-ДЕСКРИПТОР МОЛЕКУЛЯРНОЙ СТРУКТУРЫ

*Институт физиологически активных веществ РАН,
Черноголовка, Московская область*

Введение

В последние десятилетия при исследовании различных объектов и процессов широкое распространение получила фрактальная геометрия [1, 2]. Концепция фракталов и связанное с ней понятие о фрактальной размерности в настоящее время успешно используются в различных областях естествознания [3–7]. В частности, при изучении свойств нанообъектов исследуются взаимоотношения между фрактальной размерностью и свойствами материалов [8, 9]. Существование таких связей, очевидно, обусловлено наличием фрактальных свойств у молекулярных кластеров наноразмера. При этом следует констатировать отсутствие аналогичных работ в области изучения взаимоотношений «структура – свойство» и «структура – активность», по крайней мере, при использовании обычных «малых» молекул. Вероятно, это связано с тем, что такие молекулы не обладают фрактальными свойствами. Это же отмечается в работе [10], которая является одной из немногих, где использованы фрактальные идеи для изучения связи «структура – активность». Следует отметить, что отсутствие инвариантности при изменении масштаба (самоподобия) геометрических структур молекул, выраженных, например, в виде декартовых координат атомов, не исключает присутствия самоподобия при других способах представления пространственной структуры.

Целью настоящей работы является оценка фрактальной размерности геометрической структуры молекул, выраженной в виде гистограммы межатомных расстояний.

Структуру молекул рассчитывали с использованием полуэмпирического квантовохимического метода AM1, реализованного в программном комплексе HyperChem [11]. В качестве стартовых молекул при проведении полной

оптимизации служили генерируемые программой трехмерные структуры, полученные из соответствующих двухмерных структур. Точечные группы симметрии молекул устанавливались с помощью информационной системы KnowItAll [12] и HyperChem [11]. Для каждого соединения рассчитывалось 3224 дескрипторов на основе компьютерной программы DRAGON [13]. Величины плотности алканов взяты из работы [14].

Гистограммы (дискретные функции распределения) межатомных расстояний 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, алгоритм которой приведен в работе [15], путем покрытия оцифрованного изображения клетками переменной величины и подсчетом числа клеток, которые пересекают закрашенные области гистограмм. Локальная фрактальная размерность рассчитывалась по уравнению:

$$\log N(L) = \log C - D \log L, \quad (1)$$

где:

$N(L)$ – минимальное число клеток со стороной L , необходимых для покрытия фрактала,

D – «клеточная» фрактальная размерность (со знаком –),

C – константа.

Для определения величины D , в соответствии с уравнением (1), нужно провести два измерения при двух различных величинах L . Однако для более надежной оценки фрактальной размерности необходимо проводить большее число измерений. С этой целью был использован набор клеток с размерами сторон $L = 1, 2, \dots, L_{\max}$. Для каждой молекулы величина L_{\max} определялась путем деления на 10 минимального значения, выбранного из двух чисел: $(R_{\max} - R_{\min})/\Delta R$ и $f_{\max}/\Delta f$.

При расчете D необходимо учитывать то, что гистограммы имеют различные коэффициенты подобия по различным осям, так как размерности осей абсцисс и ординат имеют разный физический смысл [2]. Поэтому для сравнительного фрактального анализа молекул «клеточным» методом необходимо зафиксировать стандартную модель с начальной минимальной клеткой,

размеры которой должны быть малы по сравнению с размахом гистограммы по обеим осям. Такой минимальной клетке соответствует значение $L = 1$ в уравнении (1). Так как минимальная величина межатомного расстояния в молекуле составляет порядка 1 \AA , а точность расчета длин межатомных расстояний, полученная с помощью полуэмпирических квантовохимических методов, составляет $0,01 \text{ \AA}$ [16], эта величина была выбрана в качестве минимально возможного размера клетки ΔR по оси абсцисс. В связи с тем, что минимальная, отличная от нуля, величина бара на оси ординат составляет 1, в качестве разумного размера клетки Δf по оси ординат было установлено значение 0,01. Размер оцифрованной квадратной фрактальной матрицы определялся на основе максимума из двух величин: $R_{\max}/0,01$ и $f_{\max}/0,01$.

Регрессионный анализ выполнен с помощью модифицированной авторами компьютерной программы SVD [17]. В качестве статистических характеристик уравнений использованы: n – число точек, r^2 – квадрат коэффициента линейной корреляции, s – стандартное отклонение, F – критерий Фишера, q^2 – квадрат коэффициента линейной корреляции в условиях скользящего контроля с выбором по одному.

В качестве примера на рисунках 1 и 2 представлены результаты расчета фрактальной размерности гистограммы межатомных расстояний для молекулы метанола.

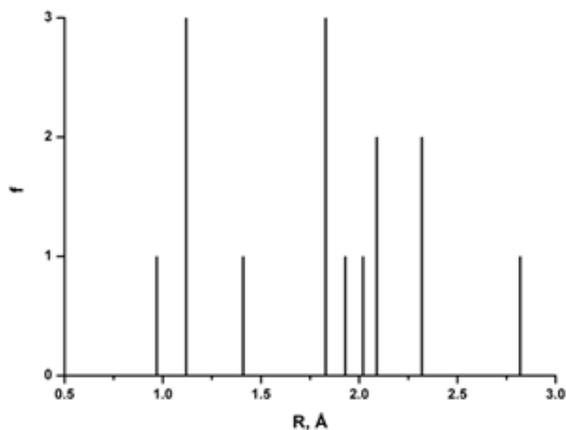


Рис. 1. Гистограмма межатомных расстояний в молекуле метанола, рассчитанная с разрешением $0,01 \text{ \AA}$

Эта молекула содержит шесть атомов, поэтому общее число межатомных расстояний равно $6 \cdot (6 - 1) / 2 = 15$, чему соответствует сумма всех баров на рис.1. Для оцифровки гистограммы необходима фрактальная матрица размером 300×300 пикселей ($3 / 0,01 = 300$). В результате работы «клеточно-го» алгоритма были получены данные, представленные на рисунке 2. Они описываются прямой линией с хорошими статистическими показателями. Полученное значение фрактальной размерности $D = 1,06$.

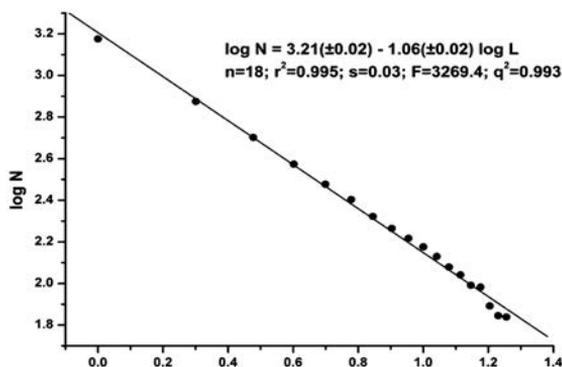


Рис. 2. Зависимость $\log N$ от $\log L$ (уравнение (1)) для молекулы метанола

В таблице 1 представлены результаты расчетов для 73 органических молекул, относящихся к шести 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, а максимальная – 1,58. Для сравнения: измерения фрактальной размерности, проведенные по описанной выше методике, дают для отрезка прямой значение $D = 1$, а для квадрата $D = 2$, что соответствует их топологическим размерностям.

Данные, представленные на рисунке 3, свидетельствуют 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 (№ 21 в таблице 1), 2-Метил-1-пропанол (№ 22), 2-Бутанол (№ 23) и 2-Метил-2-пропанол (№ 24) имеют по 15 атомов, но различаются по фрактальной размерности. Величина D при этом меняется от 1,20 до 1,38.

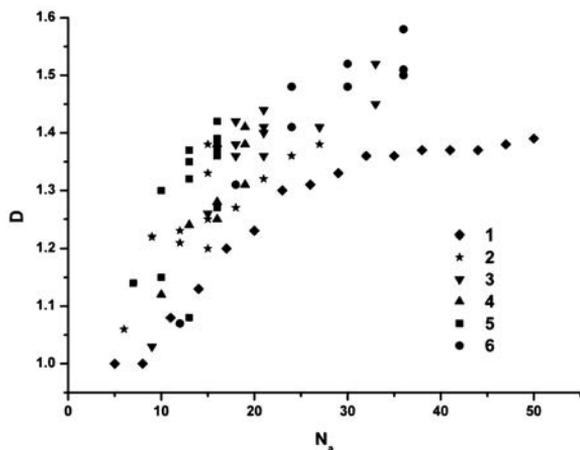


Рис. 3. Зависимость фрактальной размерности от числа атомов для различных молекул (1 – алканы, 2 – спирты, 3 – эфиры, 4 – кетоны, 5 – амины, 6 – ароматические соединения)

Наблюдающееся явление может быть объяснено, если принять во внимание точечные группы симметрии молекул. Действительно, соединения № 21 и № 24 относятся к группе C_s и имеют $D = 1,25$ и $1,20$ соответственно. Напротив, соединения № 22 и № 23 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 . Для 1-Бутанол $N_b = 49$, для 2-Метил-2-пропанол – 36, для 2-Метил-1-пропанол – 68 и для 2-Бутанол – 71. Очевидно, что чем большим числом элементов симметрии обладает молекула, тем больше будет число совпадающих межатомных расстояний и тем меньше будет представлено в гистограммах различных баров. Как показало, однако, дальнейшее исследование, более лучшей количественной мерой для оценки влияния числа атомов и точечной группы симметрии в молекуле является нормированная величина, получаемая путем деления числа баров на число атомов: N_b/N_a .

Регрессионный анализ зависимостей для отдельных точечных групп симметрии привел к следующим результатам:

$$C_{2v}: \quad D = 0,97(\pm 0,03) + 0,81(\pm 0,06) \log(N_b/N_a) \quad (2)$$

$n = 14; r^2 = 0,935; s = 0,04; F = 172,4; q^2 = 0,889;$

$$C_s: \quad D = 0,91(\pm 0,03) + 0,70(\pm 0,06) \log(N_b/N_a) \quad (3)$$

$n = 23; r^2 = 0,851; s = 0,04; F = 119,8; q^2 = 0,810;$

$$C_1: \quad D = 0,91(\pm 0,06) + 0,63(\pm 0,08) \log(N_b/N_a) \quad (4)$$

$n = 18; r^2 = 0,785; s = 0,03; F = 58,5; q^2 = 0,682;$

Уравнения (2–4) 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 .

Чтобы определить место фрактальной размерности гистограмм межатомных расстояний среди других дескрипторов, нами были проанализированы абсолютные значения парных коэффициентов линейной корреляции 3224 дескрипторов с величиной D для 73 изученных соединений. Из указанного набора дескрипторов предварительно были удалены малоинформативные переменные, которые имели свыше 95% постоянных значений. В результате такого скрининга осталось 1363 дескриптора. На рисунке 4

приведена гистограмма коэффициентов корреляции, рассчитанная с шагом 0,05. При этом минимальное значение $|r|$ равно 0,002, максимальное составляет 0,859, а средняя величина равна 0,332.

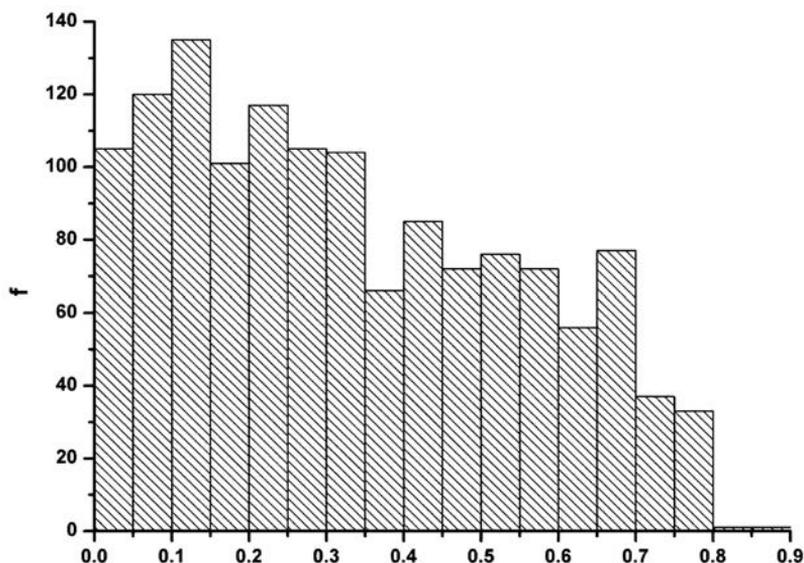


Рис. 4. Гистограмма распределения коэффициентов корреляции фрактальной размерности с 1363 дескрипторами

В работе [18] предложена классификация молекулярных дескрипторов в зависимости от их коэффициентов корреляции $|r|$ по отношению к газохроматографическому времени удерживания. В соответствии с этими данными, дескрипторы с $|r| \geq 0,99$ являются фундаментальными, $0,99 > |r| \geq 0,80$ – важными, $0,80 > |r| \geq 0,50$ – вероятными, $0,50 > |r|$ – специфичными.

Предложенную классификацию можно использовать не только для характеристики дескрипторов по отношению к рассматриваемому свойству, но и рассматриваемого свойства, в частности, фрактальной размерности, по отношению к дескрипторам. Действительно, из данных, представленных на рис. 4, следует, что из 1363 дескрипторов 1010 имеют значение $|r| < 0,5$, 351 находятся в пределах $0,80 > |r| \geq 0,50$ и 2 дескриптора имеют $0,99 > |r| \geq 0,80$. Интересно отметить, что два последних дескриптора представляют собой топологические индексы: структурное информационное содержание 1 и 2 порядка. Доля «специфичности» во фрактальной размерности по отношению к 1363 дескрипторам составляет $100 \cdot 1010 / 1363 = 74,1\%$, доля «вероятности» $100 \cdot 351 / 1363 = 25,8\%$, а оставшаяся доля «важности» – 0,1%. То есть фрактальную размерность можно рассматривать в качестве весьма «специфического» дескриптора.

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 (таблица 2).

$$d_4^{20} (\text{г/см}^3) = -0,21(\pm 0,05) + 0,70(\pm 0,03) D \quad (5)$$

$$n = 18; r^2 = 0,965; s = 0,01; F = 441,6; q^2 = 0,956.$$

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. Таким образом, фрактальная размерность может быть рекомендована к использованию в качестве дескриптора при исследовании связи «структура – свойство» и «структура – активность».

В заключение следует отметить, что предлагаемый дескриптор вполне соответствует основным требованиям, предъявляемым к новым дескрипторам [19]. Он является инвариантным по отношению к обозначению и нумерации молекулярных атомов. Величина D не зависит от молекулярного вращения и трансляции, так как в основе ее расчета лежит инвариант: межатомное расстояние. Приведенное описание расчета фрактальной размерности гистограммы межатомных расстояний не является двусмысленным. И, наконец, диапазон изменения D от 1 до 2 является удобным с практической точки зрения.

Табл. 1. Точечные группы симметрии (S), число атомов (N_a), число баров в гистограмме межатомных расстояний (N_b), фрактальная размерность (D) молекул и статистические характеристики регрессионных уравнений

№	Молекула	S	N_a	N_b	D	n	r^2	s	F	q^2
1	Метан	T_d	5	2	1,00(±0,01)	7	0,999	0,01	358483	0,999
2	Этан	D_{3d}	8	6	1,00(±0,01)	19	0,999	0,01	307988	0,999
3	Пропан	C_{2v}	11	16	1,08(±0,01)	32	0,998	0,02	17564	0,998
4	Бутан	C_{2h}	14	23	1,13(±0,01)	44	0,997	0,02	16044	0,997
5	Пентан	C_{2v}	17	32	1,20(±0,01)	57	0,992	0,04	6951	0,991
6	Гексан	C_{2h}	20	43	1,23(±0,01)	69	0,992	0,04	7938	0,991
7	Гептан	C_{2v}	23	54	1,30(±0,01)	82	0,991	0,05	9326	0,990
8	Октан	C_{2h}	26	64	1,31(±0,01)	94	0,993	0,05	12260	0,992
9	Нонан	C_{2v}	29	76	1,33(±0,01)	107	0,991	0,05	11748	0,990
10	Декан	C_{2h}	32	93	1,36(±0,01)	119	0,992	0,05	15083	0,992
11	Ундекан	C_{2v}	35	96	1,36(±0,01)	132	0,992	0,05	16576	0,991

Продолжение таблицы 1

№	Молекула	S	Na	Nb	D	n	r ²	s	F	q ²
12	Додекан	C _{2h}	38	106	1,37(±0,01)	144	0,993	0,05	19326	0,992
13	Тридекан	C _{2v}	41	116	1,37(±0,01)	157	0,993	0,05	22858	0,993
14	Тетрадекан	C _{2h}	44	132	1,37(±0,01)	169	0,993	0,05	23095	0,992
15	Пентадекан	C _{2v}	47	139	1,38(±0,01)	182	0,992	0,05	22773	0,992
16	Гексадекан	C _{2h}	50	153	1,39(±0,01)	194	0,992	0,05	23793	0,991
17	Метанол	C _s	6	10	1,06(±0,02)	18	0,995	0,03	3269	0,993
18	Этанол	C _s	9	23	1,22(±0,02)	30	0,993	0,04	3777	0,991
19	1-Пропанол	C _s	12	33	1,23(±0,01)	43	0,996	0,03	10460	0,995
20	2-Пропанол	C ₁	12	39	1,21(±0,01)	33	0,997	0,02	12298	0,994
21	1-Бутанол	C _s	15	49	1,25(±0,01)	55	0,995	0,03	10609	0,994
22	2-Мтил-1-пропанол	C ₁	15	68	1,33(±0,01)	43	0,998	0,02	20631	0,998
23	2-Бутанол	C ₁	15	71	1,38(±0,01)	46	0,996	0,03	10823	0,995
24	2-Метил-2-пропанол	C _s	15	36	1,20(±0,01)	33	0,999	0,01	32528	0,999
25	1-Пентанол	C _s	18	60	1,27(±0,01)	68	0,995	0,04	12865	0,994
26	1-Гексанол	C _s	21	80	1,32(±0,01)	80	0,993	0,05	10518	0,992
27	1-Гептанол	C _s	24	98	1,36(±0,01)	93	0,992	0,05	11146	0,991
28	1-Октанол	C _s	27	112	1,38(±0,01)	105	0,993	0,05	13949	0,992
29	Диметиловый эфир	C _{2v}	9	13	1,03(±0,01)	29	0,998	0,01	17938	0,998
30	Диэтиловый эфир	C _{2v}	15	34	1,26(±0,02)	55	0,993	0,04	7545	0,992
31	Бутилметило- вый эфир	C _s	18	70	1,38(±0,01)	67	0,994	0,04	10353	0,993
32	втор-Бутил- метиловый эфир	C ₁	18	105	1,42(±0,01)	48	0,998	0,03	21181	0,997
33	трет-Бутил- метиловый эфир	C _s	18	65	1,36(±0,01)	42	0,998	0,03	14145	0,997
34	Дипропило- вый эфир	C _{2v}	21	59	1,36(±0,01)	79	0,992	0,05	9060	0,990
35	Диизопропи- ловый эфир	C ₂	21	83	1,41(±0,01)	55	0,995	0,04	11518	0,994

№	Молекула	S	Na	Nb	D	n	r ²	s	F	q ²
36	трет-Амил-метиловый эфир	C ₁	21	138	1,44(±0,01)	47	0,998	0,02	25553	0,998
37	Бутилэтиловый эфир	C _s	21	86	1,40(±0,01)	80	0,995	0,04	16483	0,995
38	Дибутиловый эфир	C _{2v}	27	86	1,41(±0,01)	104	0,994	0,05	16003	0,993
39	Дипентиловый эфир	C _{2v}	33	125	1,45(±0,01)	129	0,992	0,05	15857	0,991
40	Диизопентиловый эфир	C ₂	33	202	1,52(±0,01)	104	0,997	0,03	33317	0,996
41	Ацетон	C _s	10	20	1,12(±0,01)	30	0,997	0,02	10729	0,997
42	2-Бутанон	C _s	13	40	1,24(±0,01)	42	0,998	0,02	18576	0,998
43	3-Метил-2-бутанон	C ₁	16	78	1,38(±0,01)	43	0,997	0,03	14187	0,996
44	2-Пентанон	C _s	16	58	1,28(±0,01)	57	0,995	0,04	11109	0,995
45	3-Пентанон	C _s	16	35	1,25(±0,01)	57	0,993	0,04	7646	0,992
46	2-Гексанон	C _s	19	67	1,31(±0,01)	69	0,995	0,04	14486	0,995
47	3-Гексанон	C _s	19	67	1,31(±0,01)	69	0,994	0,04	10935	0,993
48	2-Метил-3-пентанон	C ₁	19	110	1,38(±0,01)	57	0,996	0,03	15628	0,996
49	4-Метил-2-пентанон	C ₁	19	116	1,41(±0,01)	57	0,998	0,03	23218	0,997
50	Метиламин	C _s	7	12	1,14(±0,03)	19	0,990	0,04	1648	0,986
51	Диметиламин	C _s	10	22	1,15(±0,01)	31	0,996	0,03	6742	0,994
52	Этиламин	C ₁	10	35	1,30(±0,01)	31	0,997	0,03	9338	0,995
53	N-Этилмети-ламин	C ₁	13	56	1,37(±0,01)	44	0,996	0,03	10269	0,995
54	Изопропила-мин	C ₁	13	53	1,32(±0,01)	33	0,997	0,03	10903	0,996
55	Пропиламин	C ₁	13	55	1,35(±0,01)	40	0,996	0,03	9001	0,995
56	Триметила-мин	C _{3v}	13	15	1,08(±0,01)	30	0,996	0,03	6699	0,995
57	Бутиламин	C ₁	16	76	1,39(±0,01)	56	0,994	0,04	9355	0,993

Окончание таблицы 1

№	Молекула	S	Na	Nb	D	n	r ²	s	F	q ²
58	втор-Бутиламин	C ₁	16	80	1,42(±0,01)	46	0,998	0,03	20548	0,997
59	трет-Бутиламин	C _s	16	43	1,27±0,01)	33	0,998	0,02	14740	0,998
60	Диэтиламин	C _s	16	54	1,39(±0,02)	57	0,993	0,05	7333	0,991
61	N,N-Диметилэтиламин	C ₁	16	81	1,36(±0,01)	43	0,997	0,03	15907	0,997
62	Изобутиламин	C ₁	16	80	1,37(±0,01)	44	0,998	0,02	21732	0,998
63	N-Метилизопропиламин	C ₁	16	87	1,38(±0,01)	44	0,998	0,02	26579	0,998
64	N-Метилпропиламин	C ₁	16	78	1,36(±0,01)	57	0,996	0,03	14017	0,996
65	Бензол	D _{6h}	12	10	1,07(±0,01)	38	0,995	0,03	6572	0,994
66	Нафталин	D _{2h}	18	39	1,31(±0,01)	61	0,993	0,04	8584	0,992
67	Антрацен	D _{2h}	24	66	1,41(±0,01)	84	0,992	0,05	10382	0,991
68	Фенантрен	C _{2v}	24	113	1,48(±0,01)	81	0,996	0,04	18046	0,995
69	2,3-Бензантрацен	D _{2h}	30	99	1,48(±0,01)	108	0,993	0,05	14368	0,992
70	1,2-Бензантрацен	C _s	30	252	1,52(±0,01)	104	0,997	0,03	39943	0,997
71	Пентацен	D _{2h}	36	133	1,50(±0,01)	132	0,994	0,05	22053	0,993
72	1,2:3,4-Дибензантрацен	C _{2v}	36	199	1,51(±0,01)	103	0,997	0,04	29468	0,996
73	1,2:5,6-Дибензантрацен	C _{2h}	36	235	1,58(±0,01)	124	0,996	0,04	30875	0,996

Табл. 2. Величины фрактальной размерности (D) и плотности (d_4^{20}) в гомологическом ряду алканов

№	Алкан	D	d_4^{20} , г/см ³
1	Пентан	1,20	0,6262
2	Гексан	1,23	0,6603
3	Гептан	1,30	0,6838

Окончание таблицы 2

4	Октан	1,31	0,7025
5	Нонан	1,33	0,7176
6	Декан	1,36	0,7300
7	Ундекан	1,36	0,7402
8	Додекан	1,37	0,7487
9	Тридекан	1,37	0,7564
10	Тетрадекан	1,37	0,7628
11	Пентадекан	1,38	0,7685
12	Гексадекан	1,39	0,7730
13	Гептадекан	1,39	0,7780
14	Октадекан	1,41	0,7820
15	Нонадекан	1,42	0,7855
16	Эйкозан	1,42	0,7886
17	Пентакозан	1,46	0,8012
18	Триаконтан	1,47	0,8097

ЛИТЕРАТУРА

1. *Мандельброт Б.* Фрактальная геометрия природы. М.: Институт компьютерных исследований, 2002. 656 с.
2. *Федер Е.* Фракталы. М.: Мир, 1991. 254 с.
3. *Risovic D., Mahovic Poljacek S., Gojo M.* Applied Surface Science. 2009. Vol. 255. N. 7. P. 4283–4288.
4. *Eftekhari A.* Physica B: Condensed Matter. 2007. Vol. 387. N. 1–2. P. 92–97.
5. *Ghanbari Kh., Mousavi M.F., Shamsipur M., Rahmanifar M.S., Heli H.* Synthetic Metals. 2006. Vol. 156. N. 14–15. P. 911–916.
6. *Koňas P., Buchar J., Severa L.* European Journal of Mechanics – A/Solids. 2009. Vol. 28. N. 3. P. 545–550.
7. *Chang-Yong Lee.* Physical Review E. 2006. Vol. 73. 042901.
8. *Mihryan A., Strømme M.* Surface Science. 2007. Vol. 601. P. 315–319.
9. *Ono Y., Mayama H., Furó I., Sagidullin A.I., Matsushima K., Ura H., Uchiyama T., Tsujii K.* Journal of Colloid and Interface Science. 2009. Vol. 336. N. 1. P. 215–225.
10. *Artemenko A.G., Kovdienko N.A., Kuz'min V.E. et al.* Experimental Oncology. 2002. Vol. 24. P. 123–127.

11. <http://www.hyper.com>
12. <http://www.knowitall.com>
13. <http://www.talete.mi.it>
14. Свойства органических соединений. Справочник / Под ред. А.А. Потекина. Л.: Химия, 1984. 520 с.
15. Кроновер Р.М. Фракталы и хаос в динамических системах. Основы теории. М.: Постмаркет, 2000. С. 137.
16. Блатов В.А., Шевченко А.П., Пересыпкина Е.В. Полуэмпирические расчетные методы квантовой химии: Учебное пособие. Самара: Универс-групп, 2005. 32 с.
17. Форсайт Дж., Малькольм М., Моулер К. Машинные методы математических вычислений. М.: Мир, 1980. 280 с.
18. Dimov N., Osman A., Mekenyan O., Papazova D. *Analytica Chimica Acta*. 1994. Vol. 298. N. 3. P. 303–317.
19. <http://www.moleculardescriptors.eu>

О.А. Раевский, Е.А. Модина, О.Е. Раевская

КОМПЬЮТЕРНЫЕ МОДЕЛИ ВЗАИМОСВЯЗИ СТРУКТУРЫ ОРГАНИЧЕСКИХ СОЕДИНЕНИЙ И ИХ ИНГАЛЯЦИОННОЙ ТОКСИЧНОСТИ

*Институт физиологически активных веществ РАН,
Черноголовка, Московская область*

Современное состояние окружающей среды требует особого внимания к токсичности химических соединений (ХС) при ингаляционном пути введения, которые способны вызывать как структурно-функциональные нарушения органов дыхания, так и тяжелейшие расстройства всего организма. Поступление химических соединений через дыхательную систему относится к наиболее быстрому пути введения ХС в организм. Это объясняется как очень большой поверхностью легочных альвеол (по подсчетам она составляет 100–120 м²), так и непрерывным током крови по легочным капиллярам.

В настоящее время проблеме безопасности химических веществ уделяется особое внимание. Так, например, Евросоюз требует полную токсикологическую информацию обо всех химических веществах, производящихся или импортируемых в существенных объемах [1]. Однако проведение всесторонних испытаний разнообразных видов токсичности, в том числе ингаляционной, требует использования большого количества экспериментальных животных и сопровождается значительными финансовыми затратами. Поэтому такой всесторонней информации нет даже для многих химических соединений, широко используемых в промышленности и сельском хозяйстве. В качестве разумных альтернатив тестированию *in vivo* предлагаются компьютерные модели количественной взаимосвязи структура – активность (КССА), формирование категорийных правил, экспертных систем и методов *in vitro* [1]. Однако к настоящему моменту нет каких-либо устойчивых прогностических КССА моделей ингаляционной токсичности. Состояние дел в этой области описано в недавно опубликованном обзоре сотрудников Химического Бюро Евросоюза [2]. Поэтому целью настоящего исследования стала разработка компьютерных моделей взаимосвязи структуры орга-

нических соединений и их ингаляционной токсичности. В статье представлены первые результаты КССА-моделирования.

Экспериментальная часть

Использованы данные по ингаляционной токсичности 403 органических соединений по отношению к крысам из RTECS (Registry of Toxic Effects of Chemical Substances) [3]. В качестве количественной меры токсичности использовались величины $\log(1/LC_{50})$, где концентрации веществ, оказывающие летальное воздействие на 50% испытываемых крыс при 4 ч экспозиции, выражены в ммоль/м³.

Расчет физико-химических дескрипторов осуществляли с использованием компьютерной программы HUBOT [4]. Оценку структурного сходства осуществляли с помощью программы MOLDIVS [5].

Результаты и их обсуждение

Базовая неспецифическая ингаляционная токсичность

Традиционно считается, что аккумуляция химических соединений в липидных мембранах клеток играет ключевую роль в общей неспецифической токсичности. В токсикологии по отношению к водным организмам это явление именуется как (летальный) наркоз. Проведенное изучение на крысах подтвердило предположение, что неспецифический острый летальный наркоз характерен и для млекопитающих и прямо связан с величиной распределения газообразных химических соединений в липидной компоненте мозга [6]. Более того, было установлено, что для соединений с общей неспецифической токсичностью можно предполагать общий механизм токсичности для водных и теплокровных организмов [6, 7]. Verhaag в своей классификации химических соединений по типу действия (Mode of Action, MOA) отнес вещества, которые не взаимодействуют со специфическими рецепторами в организме, к инертным неспецифическим [8]. Обычно к таким соединениям относят углеводороды, спирты, кетоны, алкилбензолы. Для построения КССА-моделей токсичности этих соединений используют экспериментальные значения коэффициентов распределения в системах «октанол – вода» и «вода – воздух» [9].

Недавно мы провели детальное КССА-исследование токсичности органических соединений по отношению к гуппи, гольяну, радужной форели [10, 11] и *Daphnia Magna* [12]. При этом КССА-модели для соединений с неспецифическим типом токсичности, основанные на экспериментальных параметрах распределения, были успешно заменены стабильными прогностическими моделями на основе физико-химических дескрипторов, рассчитываемых программой HUBOT. Аналогичный подход было решено применить и для количественного описания базовой неспецифической ингаляционной токсичности.

В таблице 1 приведены данные о токсичности и физико-химических параметрах всех соединений из базы RTECS, которые обладают базовым неспецифическим эффектом (углеводороды, этиленовые соединения, алкилбензолы, спирты и кетоны). Интервал значений токсичности $\log(1/LC_{50})$ для этих соединений достаточно широк (от $-4,05$ до $-1,46$).

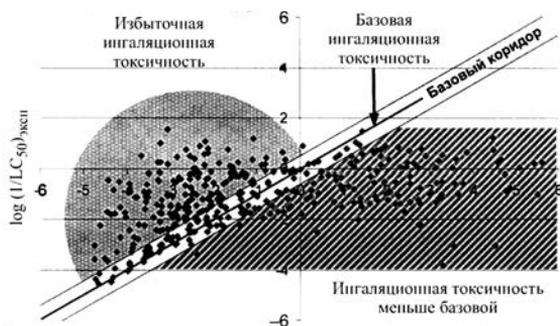
Табл. 1. Базовая ингаляционная токсичность химических соединений

N	Наименование	CAS*	$\log O_{KT/ВОЗД}$	α	ЕСa	ЕСd	$\log(1/LC_{50})_{эксп}$
1	нонан	111-84-2	3,86	17,289	0	0	-2,12
2	1-пентен	3887-49-8	1,93	9,757	0,22	0	-3,40
3	2-метилбутан; изопентан	78-78-4		9,949	0	0	-3,59
4	пентан	109-66-0	1,95	9,949	0	0	-3,70
5	метилбензол; толуол	108-88-3	3,31	12,269	0,48	0	-2,73
6	1-метилэтил- бензол	3048-64-4	3,98	15,939	0,68	0	-2,51
7	циклогексанон	108-94-1	4,24	11,164	2,06	0	-2,51
8	1,2,4-триметил- бензол	95-63-6	4,16	15,939	0,59	0	-2,18
9	1,3,5-триметил- бензол	108-67-8	4,25	15,939	0,82	0	-2,30
10	2-метил-1- пентен	763-29-1		11,592	0,12	0	-3,04
11	декагидронаф- тилин	91-17-8		17,576	0	0	-1,46
12	2-метил-2- пентен	625-27-4		11,592	0,21	0	-3,01
13	1-гексен	592-41-6	2,41	11,592	0,20	0	-3,12
14	гексан	110-54-3	2,44	11,784	0	0	-3,29
15	2-бутанол	78-92-2	3,83	8,751	1,62	-1,13	-2,82
16	1-бутанол	71-36-3	4,31	8,751	1,60	-1,32	-2,51
17	2-метокси-2- метилпропан	1634-04-4	2,58	10,586	1,58	0	-2,67
18	2-метил-1- пропанол; изобутанол	78-83-1	3,93	8,751	1,57	-1,15	-2,41
19	бутан	106-97-8	1,53	8,114	0	0	-4,05
20	1-этинил-1- циклогексен	2622-08-4		14,296	0,34	0	-1,99
21	1,4-диметил- бензол	106-42-3	3,79	14,104	0,58	0	-2,27

Окончание таблицы 1

N	Наименование	CAS*	$\log K_{\text{окт./возд}}$	α	ЕСа	ЕСd	$\log(1/LC_{50})_{\text{эксп}}$
22	2-гексанон	591-78-6	3,68	11,938	1,90	0	-2,51
23	2-бутен	1514-76-23		7,922	0,20	0	-3,79
24	2-метил-1-пропен	590-18-1		7,922	0,20	0	-3,99
25	метанол	67-56-1	2,87	3,246	1,63	-1,50	-3,42
26	октан	111-65-9	3,40	15,454	0	0	-3,01
27	гептан	142-82-5	2,95	13,619	0	0	-3,01
28	этинилбензол; стирол	100-42-5	3,86	13,912	1,16	0	-2,05

CAS – регистрационный номер химического соединения;
 $\log K_{\text{окт./возд}}$ – константа распределения в системе 1-октанол – воздух.



$\log(1/LC_{50})_{\text{теор}}$

Базовая ингаляционная токсичность (крысы, экспозиция 4 ч)

Сравнение экспериментальных (ось ординат) и расчетных (ось абсцисс) значений ингаляционной токсичности.

Для построения КССА-моделей базовой неспецифической токсичности были прежде всего использованы константы распределения химических веществ в системах «октанол – вода» ($\log K_{\text{окт./вода}}$), «октанол – воздух» ($\log K_{\text{окт./воздух}}$) и «вода – воздух» ($\log K_{\text{вода./воздух}}$). В случае неспецифической токсичности для водных организмов хорошие однопараметровые КССА-модели получаются при использовании $\log K_{\text{окт./вода}}$ [9, 10]. Однако использование этого параметра для описания неспецифической ингаляционной токсичности не дало какой-либо значимой корреляции. Такой же неудовлетворительный результат был получен и при использовании $\log K_{\text{вода./воздух}}$. При использовании же $\log K_{\text{окт./воздух}}$ была получена удовлетворительная корреляция следующего вида:

$$\log(1/LC_{50}) = -4,64(\pm 0,23) + 0,56(\pm 0,07)\log_{\text{окт./воздух}} \quad (1)$$

$n = 24, R = 0,884, SD = 0,26, F = 67,96,$

где n – число рассматриваемых соединений; R – коэффициент корреляции, SD – стандартное отклонение, F – критерий Фишера. Здесь и далее в скоб-

ках указаны ошибки оценки параметров в пределах 95% доверительного интервала.

Обнаруженная корреляция базовой неспецифической ингаляционной токсичности с константой распределения химических соединений в системе «октанол – воздух» представляется естественной: токсичность веществ у водных организмов (средой обитания которых является вода) зависит от $\log K_{\text{окт/вода}}$, а ингаляционная токсичность у теплокровных (средой обитания которых является воздух) – от $\log K_{\text{окт/воздух}}$.

Константы распределения химических соединений в различных средах хотя и являются полезными при обсуждении транспорта веществ в организмах, имеют существенный недостаток – для их использования в КССА необходимо синтезировать соединение и измерить соответствующую константу. Таких измеренных констант не так уж много в литературе (и измерения часто проводятся при различных условиях). По этой причине гораздо целесообразнее использовать в КССА-моделях расчетные параметры. Ранее нами для моделей водной токсичности были успешно использованы дескрипторы программы HUVOT [10–13]. В данной работе использование этих дескрипторов для базовой неспецифической ингаляционной токсичности позволило построить следующее корреляционное уравнение:

$$\begin{aligned} \log(1/LC_{50}) = & -5,55(\pm 0,24) + 0,201(\pm 0,018)\alpha + \\ & + 0,40(\pm 0,09)\sum Ca(o) - 0,50(\pm 0,16)\sum Cd, \\ N = 28, R = 0,927, SD = 0,26, F = 48,67, \end{aligned} \quad (2)$$

где α – молекулярная поляризуемость, характеризующая стерические взаимодействия, $Ca(o)$ – overall scale) и Cd – водородносвязанные акцепторный и донорный факторы, характеризующие прочность водородных связей в рассматриваемой среде.

Уравнение (2) имеет хорошие статистические характеристики и позволяет рассчитывать общую неспецифическую ингаляционную токсичность с точностью, близкой к экспериментальной ошибке для всех органических соединений. Более того, уравнение (2) позволяет количественно оценивать вклады различных межмолекулярных взаимодействий в неспецифическую токсичность и свидетельствует о том, что и стерические взаимодействия, и водородные связи повышают базовую неспецифическую токсичность. Преимущества уравнения (2) по сравнению с уравнением (1) очевидны.

Сопоставление ингаляционной токсичности всех рассматриваемых соединений по отношению к базовой неспецифической ингаляционной токсичности

Знание экспериментальной величины ингаляционной токсичности для любого органического соединения и рассчитанного для этого соединения по уравнению (2) вклада общей неспецифической ингаляционной токсичности

позволяет произвести оценку дополнительного вклада в токсичность за счет различных взаимодействий вещества в организме. На рисунке приведено расположение всех изученных соединений на плоскости $\log(1/LC_{50})_{эксп}/\log(LC_{50})$ теор рассчитанное с использованием уравнения (2). Линия, соответствующая уравнению (2), проходит через начало координат и все парные одинаковые значения по осям X и Y. Учитывая возможные ошибки измерений и расчетов, можно выделить коридор базовой неспецифической токсичности ($\pm 0,50$). В этот коридор помимо 28 соединений с априорной базовой неспецифической токсичностью попали еще 82 соединения. Все остальные соединения располагаются в 2 областях: выше базового коридора (166 веществ) и ниже базового коридора (127 веществ). К соединениям, проявляющим «избыточную» ингаляционную токсичность, можно отнести ряд фосфорорганических соединений, изоцианатов, фторорганических производных, оксиранов, хлорированных кетонов и сложных эфиров, нитрилов. Очевидно, эти соединения способны проявлять специфическую ингаляционную токсичность в дополнение к общей неспецифической. У целого ряда нитроароматических и гетероциклических соединений, эфиров, аминов экспериментальная ингаляционная токсичность оказалась существенно ниже по сравнению с рассчитанной общей неспецифической токсичностью. Разумное объяснение этому – быстрый гидролиз и метаболизм. Очевидно, в этом случае исходные молекулы подвергаются распаду на более простые соединения с меньшей токсичностью.

В коридор базовой неспецифической токсичности попали 82 соединения, структура которых существенно отличается от инертных соединений. Учитывая наличие среди изученных соединений целой группы веществ с меньшей экспериментальной токсичностью, по сравнению с рассчитанной общей неспецифической токсичностью, можно полагать, что для этих 82 соединений наблюдается примерная компенсация эффектов исходной избыточной токсичности и метаболизма.

Классификационные и регрессионные КССА-модели ингаляционной токсичности

Существенное влияние метаболизма на ингаляционную токсичность не позволяет применить к этому свойству классификацию химических соединений по типу действия, предложенную Verhaar [8]. Вероятно, в этом случае целесообразно опираться на распределение соединений, приведенное на рисунке и обсуждать четыре класса соединений.

Класс 1 – инертные соединения, обеспечивающие базовую ингаляционную токсичность.

Класс 2 – соединения, избыточная токсичность которых за счет специфического воздействия компенсируется гидролизом и (или) метаболизмом. Токсичность таких соединений, хотя и соответствует рассчитанной для базовой, является результатом указанных противоположных эффектов.

Класс 3 – соединения с избыточной токсичностью по отношению к базовой за счет специфического действия на биомишени (при этом эффект от данного специфического действия преобладает над частичной потерей токсичности за счет гидролиза и метаболизма).

Класс 4 – соединения, существенно подверженные гидролизу и метаболизму, причем токсичность у продуктов этих реакций низка.

Очевидно, что получить в таких условиях какие-либо удовлетворительные КССА-модели для всех изученных 403 соединений невозможно. В данной работе мы попытались построить классификационные и регрессионные модели на основе структурного сходства и дескрипторов программы НУВОТ. Для классификации использовалась информация о принадлежности ближайшего соседа к одному из указанных классов. А величина токсичности рассчитывалась на основе следующей модификации уравнения (2):

$$\log(1/LC_{50}(i)) = \log(1/LC_{50}(ccc)) + 0.201(\alpha_i - \alpha_{ccc}) + 0,40(\sum Ca(o)_i - \sum Ca(o)_{ccc}) - 0.50(\sum Cd_i - \sum Cd_{ccc}), \quad (3)$$

где i – рассматриваемое соединение, ccc – структурно сходное соединение.

При таком подходе предполагается, что рассматриваемое и структурно сходное соединения обладают сходными эффектами (специфической токсичности, метаболизма), а разница в их токсичности связана только с разностью величины вклада общей неспецифической токсичности за счет различий их физико-химических параметров.

Разумеется, что использование уравнения (3) целесообразно только для соединений, имеющих в рассматриваемом массиве другие соединения с хорошими значениями индексов структурного сходства.

Таблица 2 содержит результаты применения указанного подхода к изученным соединениям (N – число соединений). Из 403 соединений только 135 имеют сходные структуры со значением коэффициента Танимото (T_c) 0,50 и больше. В этом подмассиве удалось правильно классифицировать 75% соединений 1, 3 и 4 классов. В подмассиве с $T_c > 0,60$ (75 соединений) существенно улучшилось распознавание соединений первого класса. А в подмассиве с $T_c > 0,70$ (43 соединения) улучшилось и распознавание соединений третьего класса. Что касается регрессионных соотношений, то использование уравнения (3) привело к хорошему количественному описанию ингаляционной токсичности соединений первого класса при $T_c > 0,50$; второго и третьего классов при $T_c > 0,70$. Можно полагать, что дальнейшее расширение массива данных по ингаляционной токсичности приведет к появлению все большего числа структурно сходных соединений и даст возможность построить устойчивые прогностические КССА-модели ингаляционной токсичности на основе физико-химических дескрипторов и структурного сходства.

Табл. 2. Расчет $\log(1/LC_{50})$ и степени правильного распознавания классов для ингаляционной токсичности соединений у крыс на основе уравнения (3)

Tc	N	Количество соединений в классе	Степень правильного распознавания класса первого структурно-сходного соседа (1сс), %	$\log(LC_{50})_{\text{ксп}} = \alpha_0 + \alpha_1 \cdot \log(1/LC_{50})_{\text{ср}}$					
				α_0	α_1	R ²	R	SD	F
> 0,50	135	18 класс	72,22	-1,06	0,68	0,718	0,847	0,35	40,66
		32 класс	34,38	-0,80	0,54	0,569	0,754	0,82	39,66
		52 класс	75,00	-0,18	0,61	0,500	0,707	0,72	49,91
		33 класс	75,76	-0,46	0,14	0,031	0,176	1,07	0,99
> 0,60	75	13 класс	92,31	-0,10	0,97	0,930	0,964	0,17	145,25
		13 класс	23,08	-0,58	0,54	0,615	0,784	0,89	17,59
		28 класс	75,00	-0,05	0,73	0,623	0,789	0,63	42,96
		21 класс	66,67	-0,34	0,11	0,017	0,130	1,23	0,33
> 0,70	43	13 класс	92,31	-0,10	0,97	0,930	0,964	0,17	145,25
		4 класс	0	0,67	1,50	0,980	0,990	0,25	99,09
		18 класс	88,89	-0,08	0,82	0,709	0,842	0,62	39,04
		8 класс	62,50	-0,99	0,24	0,038	0,195	1,38	0,20

Работа выполнена при финансовой поддержке МНПЦ (проект № 3777).

ЛИТЕРАТУРА

1. Commission of the European Communities. Regulation (EC) No. 1907 / 2006 of the European Parliament and of the Council of 18 December 2006 concerning the registration, Evaluation, Authorisation and Restriction of Chemicals (REACH).
2. Lessigiarska, T. Netzeva and A.P. Worth. QSAR Comb. Sci. 2008. Vol. 27(1). P. 41–48.
3. RTECS (Registry of Toxic Effects of Chemical Substances), Cincinnati, OH, NIOSH (2009); [<http://www.cdc.gov/niosh/rtecs/>].
4. Раевский О.А., Григорьев В.Ю., Трепалин С.В. HYBOT (Hydrogen Bond Thermodynamics) program package, Патент России № 990090 (1999).
5. Raevsky O.A., Gerasimenko V.A., Trepalin S.V. MOLDIV (MOlecular DI-Versity & Similarity), Патент России № 990093.1999.
6. De Jough J., Verhaar H.J.M., Hermens J.L.M. Toxicol. Sci. 1998. Vol. 45(1). P. 26–32.
7. McCarty L.S., Mackay D. Envir. Sci. Technol. 1993. Vol. 27(12). P. 1719–1728.
8. Verhaar H.J.M., van Leeuwen C.J., Hermens J.L.M. Chemosphere. 1992. Vol. 25(5). P. 471–491.
9. De Wolf W., Lieder P.H. and Walker J.D. QSAR & Comb. Sci. 2004. Vol. 23(7). P. 521–524.
10. Raevsky O.A., Grigor'ev V.Yu., Weber E. E, Dearden J.C. QSAR & Comb. Sci. 2008. Vol. 27(11/12). P. 1274–1281.
11. Raevsky O.A., Grigor'ev V.Yu., Dearden J.C. and Weber E.E. QSAR & Comb. Sci. 2009. Vol. 28(1). P. 1–12.
12. Раевский О.А., Раздольский А.Н., Тонкопий В.Д. и др. // Хим.-фарм. журн. 2008. Т. 42(6). С. 22–27.
13. Raevsky O.A. and Dearden J.C. SAR & QSAR in Envir. Res. 2004. Vol. 15(5, 6). P. 433–448.

В.Ю. Григорьев, О.А. Раевский

РЕКУРРЕНТНАЯ МОДЕЛЬ ОСТРОЙ ТОКСИЧНОСТИ В ГОМОЛОГИЧЕСКИХ РЯДАХ ОРГАНИЧЕСКИХ СОЕДИНЕНИЙ

Институт физиологически активных веществ РАН,
Черноголовка, Московская область

Гомологические ряды органических соединений находят применение, в частности, в качестве объектов при изучении механизмов токсичности [1, 2]. Это связано с тем, что гомологи являются структурно близкими соединениями, имеющими одинаковый тип действия. Известно, что в гомологических рядах молекул с ростом числа атомов углерода токсичность увеличивается до известного предела, после которого наступает насыщение или даже разворот [3, 4]. Это может быть связано с различными процессами, например, с уменьшением растворимости веществ или ростом степени ассоциации молекул.

Для описания зависимости токсичности химических соединений от их структуры используются как простые линейные, так и параболические модели [1, 5–7]. Очевидно, что линейные уравнения имеют ограниченную область применения для гомологических рядов молекул. Параболические модели выглядят более предпочтительными, но их надежность при экстраполяции данных вызывает сомнения. Не так давно была продемонстрирована эффективность использования рекуррентной модели для описания различных физико-химических свойств молекул в гомологических рядах [8].

Целью настоящей работы является исследование токсичности в гомологических рядах на основе рекуррентных уравнений.

Методы исследования и данные

В работе были использованы данные по снижению световой эмиссии бактерии *Vibrio fisheri* ($\log(1/EC_{50})$, ммоль/л, 5 мин.) [9], по ингибированию роста инфузории *Tetrahymena pyriformis* ($\log(1/IGC_{50})$, ммоль/л, 40 час.), [10, 11], по острой водной токсичности химических соединений ($\log(1/LC_{50})$, ммоль/л, 96 час.) по отношению к рыбам *Poecilia reticulata* [12] и *Pimephales promelas* [13], а также данные по острой токсичности соединений при внутривенном введении мышам и крысам ($\log(1/LD_{50})$, ммоль/кг) взятые из

базы данных SYMYX [14]. Всего было исследовано 85 химических соединений из семи гомологических рядов.

Регрессионный анализ выполнен с помощью модифицированной программы SVD [15]. Интерполяция данных проводилась на основе кубических сплайнов [15] или линейных регрессионных уравнений. В качестве статистических характеристик были использованы: n – число точек, r^2 – квадрат коэффициента линейной корреляции, s – стандартное отклонение, F – критерий Фишера, q^2 – квадрат коэффициента линейной корреляции в условиях скользящего контроля с выбором по одному.

Результаты и их обсуждение

В таблице 1 представлены данные по острой токсичности исследованных соединений, включая спирты (13 соединений), кетоны (11), сложные эфиры (10), амины (12), фенолы (10), альдегиды (12) и кислоты (17). Токсичность меняется в широких пределах: для водных организмов от $\log(1/LC_{50}) = -2,96$ до $\log(1/EC_{50}) = 3,69$. Для теплокровных организмов минимальное значение $\log(1/LD_{50})$ составляет $-2,17$, а максимальное $1,09$.

Зависимость токсичности от числа атомов углерода во всех исследованных гомологических рядах имеет прямолинейный характер или стремится к некоторому пределу. Поэтому для аппроксимации данных может быть использовано простое рекуррентное уравнение вида:

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

где T_n – токсичность текущего n -го члена гомологического ряда, T_{n-1} – токсичность предшествующего члена гомологического ряда, A и B – параметры.

Результаты соответствующих расчетов представлены в таблице 2. Полученные уравнения характеризуются высокими статистическими показателями. Например, в большинстве случаев, $r^2 > 0,9$. Часть представленных зависимостей имеет предельное значение токсичности ($B < 1$), другая часть не имеет предела ($B > 1$). При этом коэффициент A варьирует в пределах от $0,13$ до $0,67$.

Полезность математических моделей для описания токсичности, со статистической точки зрения, можно связать с качеством экстраполяции данных. С этой целью мы сопоставили три модели: рекуррентную (уравнение 1),

линейную:

$$T = a_0 + a_1 M_{CH_2} \quad (2)$$

и параболическую:

$$T = a_0 + a_1 M_{CH_2} + a_2 M_{CH_2}^2 \quad (3)$$

где T – токсичность; M_{CH_2} – число метиленовых групп в гомологическом ряду $R_1 - (CH_2)_m - R_2$; a_0, a_1, a_2 – параметры уравнений.

Для сравнения были выбраны данные по токсичности с учетом количества молекул в ряду и с максимальным представлением всех использованных типов химических соединений и биологических объектов. В результате было отобрано пять комбинаций гомологической ряд/биообъект: спирты и *Tetrahymena pyriformis*, кетоны и *Vibrio fisheri*, амины и *Pimephales promelas*, альдегиды и *Poecilia reticulata*, кислоты и мыши. Из каждого ряда отбрасывались последние три точки, а на основе оставшихся соединений рассчитывались соответствующие уравнения. Затем эти уравнения использовались для расчета токсичности отброшенных точек. В качестве статистики для сравнения результатов было использовано стандартное отклонение s для экспериментальных и рассчитанных значений токсичности. Всего для каждой модели было получено по пять значений s на основе 15 тестовых точек.

В результате было установлено, что для линейной модели среднее значение $s = 0,36$, для параболической модели $s = 0,49$ и для рекуррентной модели $s = 0,33$. Полученные результаты свидетельствуют о том, что рекуррентная модель обладает наилучшей способностью к экстраполяции среди рассмотренных трех уравнений.

Одним из эффективных подходов при исследовании связи «структура – токсичность» является использование классификации молекул по типу действия [1]. В литературе можно найти различные схемы классификации, при этом соединения одного типа могут называться по-разному, например, «неполярный наркоз» [12] и «наркоз I» [13]. Обобщая [16], все типы действия можно разбить на три класса. Первый класс – неполярная неспецифическая токсичность, второй класс – полярная неспецифическая токсичность и третий класс – специфическая токсичность. Учитывая это, рассматриваемые гомологические ряды спиртов, кетонов и сложных эфиров можно отнести к классу 1, амины и фенолы к классу 2, альдегиды и кислоты – к классу 3.

Использованная классификация основана на изучении токсичности химических соединений по отношению к водным организмам (экоотоксичность). Однако, с известной долей вероятности, она может быть использована и для теплокровных мышей и крыс. Основанием для такого вывода является факт существования корреляции токсичности между рассмотренными выше биологическими объектами [5].

Представленные в таблице 2 рекуррентные модели могут быть выражены в виде точек на плоскости А–В с учетом типа гомологического ряда и вида биологического объекта. На рисунке 1 изображены коэффициенты рекуррентных уравнений для различных типов химических соединений.

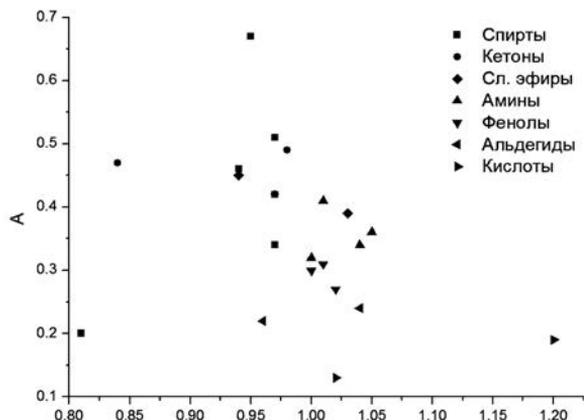
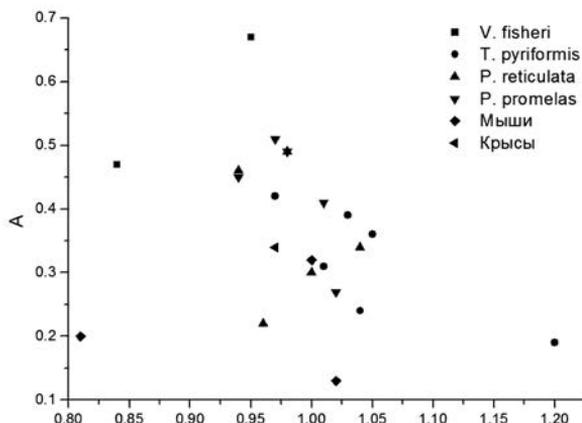


Рис. 1. Параметры рекуррентных уравнений (таблица 2) на плоскости А–В с учетом типа гомологического ряда

Этот рисунок содержит 23 точки, которые соответствуют уравнениям из таблицы 2. На нем отчетливо видны три кластера. В первый кластер входят спирты, кетоны и сложные эфиры. Во второй кластер – амины и фенолы. Третий кластер состоит из альдегидов и кислот. Таким образом, представленный метод графического анализа позволяет различать гомологические ряды между собой, относя их к веществам с неспецифическим или специфическим характером токсического действия.

Одной из важных дискутируемых тем в литературе является вопрос о сравнительной токсической чувствительности различных биологических объектов [5]. Одним из подходов для получения результатов является корреляционный анализ токсичности в одинаковых рядах соединений [17, 18]. Обычно в этом случае сходство биообъектов между собой оценивается с помощью коэффициента корреляции. В работе [19] использован другой подход, основанный на проекции данных из матрицы, образованной на основе токсичности одного и того же ряда соединений по отношению к различным биообъектам, на главные компоненты. Мы в своей работе анализировали токсичность различных гомологических рядов между различными биологическими объектами. При этом в качестве меры токсикологической близости биообъектов между собой использовали расстояние между точками на рисунке 2.



При этом в качестве меры токсикологической близости биообъектов между собой использовали расстояние между точками на рисунке 2.

Рис. 2. Параметры рекуррентных уравнений (таблица 2) на плоскости А–В с учетом вида биологического объекта

Обращает на себя внимание уникальность данных для бактерии *V. fisheri* ввиду отсутствия в ближайших окрестностях аналогичных моделей токсичности. Близость (№ 3, № 12 из таблицы 2) или даже совпадение (№ 9, № 10) параметров рекуррентных уравнений, описывающих токсичность химических соединений по отношению к рыбам *P. reticulata* и *P. promelas* не вызывает удивления: все-таки это очень сходные организмы. В работе [20] установлена токсикологическая близость *T. pyriformis* и *P. reticulata* по отношению к ряду химических соединений. Аналогично констатируется большое сходство регрессионных моделей, описывающих неспецифическую токсичность, по отношению к *T. pyriformis* и *P. promelas* [21]. Из наших данных также следует сходство моделей токсичности *T. pyriformis* и *P. reticulata* (№ 13, 14), а также *T. pyriformis* и *P. promelas* (№ 11, 15).

В случае мышей две модели не имеют ближайших соседей. А вот параметры третьей модели, описывающей токсичность в ряду аминов (№ 16), оказываются очень близкими к моделям № 17 и № 18. Первая из них связана с токсичностью фенолов по отношению к *T. pyriformis*, а вторая описывает токсические свойства фенолов по отношению к *P. reticulata*. Что касается крыс (№ 6), то ближайшим соседом для них оказываются мыши (№ 16).

Таким образом, в данной работе удалось показать, что в качестве модели для описания острой токсичности в гомологических рядах органических соединений по отношению к различным биологическим объектам может быть использовано простое рекуррентное уравнение. При этом было выявлено, что рекуррентная модель токсичности обладает лучшей способностью к экстраполяции данных по сравнению с линейной и параболической моделями. При изображении коэффициентов рекуррентных уравнений на соответствующей плоскости с учетом типов химических соединений удастся выявить неспецифический или специфический характер токсического действия гомологического ряда. Аналогичная визуализация с использованием биологических объектов позволяет оценить их токсикологическую близость между собой с помощью расстояния между ними.

Полученные результаты могут служить основой для создания в дальнейшем компьютерных программ, оценивающих острую токсичность новых химических соединений на базе подходов «структура – активность».

Работа выполнена при финансовой поддержке МНТЦ (проект 3777).

**Избранные труды доктора химических наук, профессора О.А. Раевского
в области молекулярного компьютерного дизайна**

Табл. 1. Острая токсичность органических соединений по отношению к различным биологическим объектам

№	Соединение	Биологический объект					
		V.f.	T.p.	P.r.	P.p.	M	K
1	Метанол		-2,67	-2,94	-2,96	-2,17	-1,82
2	Этанол	-2,89	-1,99	-2,44	-2,50	-1,63	-1,49
3	1-Пропанол	-2,22	-1,75	-1,95*	-1,88	-1,06	-0,99
4	1-Бутанол	-1,66	-1,43	-1,37	-1,37	-0,71	-0,62
5	1-Пентанол	-0,53	-1,05	-0,67*	-0,73	-0,32	-0,28
6	1-Гексанол	0,68	-0,38	0,02	0,02	0,00	
7	1-Гептанол	1,07	0,11	0,55*	0,53	0,19*	
8	1-Октанол	1,58	0,63	0,98	0,98	0,28	
9	1-Нонанол	1,87*	0,85	1,40	1,40		
10	1-Деканол	2,08	1,34	1,81	1,82		
11	1-Ундеканол	2,59*	1,95	2,21			
12	1-Додеканол	3,69	2,16	2,26			
13	1-Тридеканол		2,45				
14	Ацетон	-2,44	-2,15	-2,10	-2,09		
15	2-Бутанон	-1,68	-1,75	-1,65	-1,65		
16	2-Пентанон	-0,93*	-1,22	-1,13*	-1,16		
17	2-Гексанон	-0,23*	-1,34	-0,56*	-0,63		
18	2-Гептанон	0,38*	-0,49	0,01*	-0,06		
19	2-Октанон	0,86	-0,15	0,55	0,55		
20	2-Нонанон	1,18*	0,66	1,04*	0,97		
21	2-Деканон	1,30	0,58	1,43	1,51		
22	2-Ундеканон		1,53		2,06		
23	2-Додеканон		1,67		2,19		
24	2-Тридеканон		2,12		2,74		
25	Метилацетат		-1,60		-0,68		
26	Этилацетат	-1,77	-1,30		-0,42		
27	Пропилацетат		-1,24		0,23		
28	Бутилацетат	0,22	-0,49		0,81		
29	Амилацетат		-0,15*		1,25*		
30	Гексилацетат		-0,01		1,52		

Продолжение таблицы 1

№	Соединение	Биологический объект					
		V.f.	T.p.	P.r.	P.p.	M	K
31	Гептилацетат		0,44*				
32	Октилацетат		1,06				
33	Нонилацетат		1,59*				
34	Децилацетат		1,88				
35	Этиламин			-0,70			
36	Пропиламин		-0,71	-0,72	-0,72		
37	Бутиламин		-0,57	-0,56	-0,56	-0,43	
38	Амиламин		-0,48	-0,31	-0,31	-0,11*	
39	Гексиламин		-0,22	0,25	0,25	0,22*	
40	Гептиламин		0,21	0,72	0,72	0,54*	
41	Октиламин		0,35	1,40	1,40	0,86	
42	Нониламин		1,70	1,82	1,82		
43	Дециламин		2,06	2,18	2,18		
44	Ундециламин		2,33	2,91	2,91		
45	Додециламин			3,27	3,26		
46	Тридециламин			3,46	3,48		
47	Фенол		-0,35	0,45	0,46		
48	4-Метилфенол		-0,16	0,74	0,82		
49	4-Этилфенол		0,21	1,07	1,07		
50	4-Пропилфенол		0,64	1,09	1,09		
51	4-Бутилфенол		0,88*	1,47	1,62*		
52	4-Пентилфенол		1,20*	2,12	1,92*		
53	4-Гексилфенол		1,52*	2,24*	2,22*		
54	4-Гептилфенол		1,84*	2,55*	2,52*		
55	4-Октилфенол		2,16*	2,86*	2,82*		
56	4-Нонилфенол		2,47	3,20	3,20		
57	Метаналь			0,04			
58	Этаналь			0,10	0,12		
59	Пропаналь		-0,49	0,59			
60	Бутаналь		-0,38	0,72	0,69		
61	Пентаналь		-0,02	0,82	0,82		

№	Соединение	Биологический объект					
		V.f.	T.p.	P.r.	P.p.	M	K
62	Гексаналь		-0,17	1,01	0,76		
63	Гептаналь		0,00	1,11			
64	Октаналь		0,48	1,21			
65	Нонаналь		0,81	1,38*			
66	Деканаль		1,28	1,69			
67	Ундеканаль		1,69				
68	Додеканаль		1,76				
69	Уксусная кислота					-0,94	
70	Пропановая кислота		-0,51			-0,93	
71	Бутановая кислота		-0,57			-0,96	
72	Пentanовая кислота		-0,27			-1,10	
73	Гексановая кислота		-0,21			-1,17	
74	Гептановая кислота		-0,11			-0,96	
75	Октановая кислота		0,08			-0,62	
76	Нонановая кислота		0,35			-0,15	
77	Декановая кислота		0,51			0,13	
78	Ундекановая кислота		0,90			0,12	
79	Додекановая кислота					0,18	
80	Тридекановая кислота					0,22	
81	Тетрадекановая кислота					0,73	
82	Пентадекановая кислота					0,65	
83	Гексадекановая кислота					0,65	
84	Гептадекановая кислота					0,88	
85	Октадекановая кислота					1,09	

Примечание:

V. f. – *Vibrio fisheri* ($\log(1/EC_{50})$, ммоль/л);

T. p. – *Tetrahymena pyriformis* ($\log(1/IGC_{50})$, ммоль/л);

P. r. – *Poecilia reticulata* ($\log(1/LC_{50})$, ммоль/л);

P. p. – *Pimephales promelas* ($\log(1/LC_{50})$, ммоль/л);

M – мыши ($\log(1/LD_{50})$, ммоль/кг);

K – крысы ($\log(1/LD_{50})$, ммоль/кг);

* – интерполированные значения.

Табл. 2. Параметры и статистические характеристики рекуррентных уравнений

№	Соединения	Биообъект	A	B	N	r ²	s	F	q ²
1	Спирты	V. fisheri	0,67(±0,12)	0,95(±0,06)	10	0,966	0,37	226,1	0,946
2	Спирты	T. pyriformis	0,42(±0,05)	0,97(±0,03)	12	0,989	0,17	879,0	0,983
3	Спирты	P. reticulata	0,46(±0,04)	0,94(±0,03)	11	0,994	0,14	1395,5	0,989
4	Спирты	P. promelas	0,51(±0,04)	0,97(±0,03)	9	0,995	0,11	1470,5	0,992
5	Спирты	Мыши	0,20(±0,04)	0,81(±0,03)	7	0,992	0,07	634,3	0,982
6	Спирты	Крысы	0,34(±0,13)	0,97(±0,10)	4	0,978	0,09	90,0	0,907
7	Кетоны	V. fisheri	0,47(±0,04)	0,84(±0,03)	7	0,992	0,11	645,8	0,981
8	Кетоны	T. pyriformis	0,42(±0,13)	0,97(±0,10)	10	0,926	0,39	99,7	0,896
9	Кетоны	P. reticulata	0,49(±0,03)	0,98(±0,02)	7	0,997	0,07	1633,4	0,992
10	Кетоны	P. promelas	0,49(±0,04)	0,98(±0,03)	10	0,992	0,14	1026,4	0,986
11	Сл. эфиры	T. pyriformis	0,39(±0,08)	1,03(±0,08)	9	0,963	0,24	181,9	0,939
12	Сл. эфиры	P. promelas	0,45(±0,09)	0,94(±0,12)	5	0,952	0,20	60,1	0,842
13	Амины	T. pyriformis	0,36(±0,16)	1,05(±0,16)	8	0,882	0,44	44,8	0,833
14	Амины	P. reticulata	0,34(±0,08)	1,04(±0,05)	11	0,980	0,23	431,1	0,968
15	Амины	P. promelas	0,41(±0,08)	1,01(±0,05)	10	0,983	0,20	459,7	0,972
16	Амины	Мыши	0,32(0,01)	1,00(0,01)	4	0,999	0,01	14951,0	0,999
17	Фенолы	T. pyriformis	0,31(±0,04)	1,01(±0,03)	9	0,994	0,07	1172,8	0,989
18	Фенолы	P. reticulata	0,30(±0,14)	1,00(±0,08)	9	0,960	0,19	168,4	0,946
19	Фенолы	P. promelas	0,27(±0,11)	1,02(±0,06)	9	0,975	0,14	270,0	0,962
20	Альдегиды	T. pyriformis	0,24(±0,08)	1,04(±0,10)	9	0,933	0,23	97,7	0,886
21	Альдегиды	P. reticulata	0,22(±0,10)	0,96(±0,11)	9	0,917	0,14	77,1	0,808
22	Кислоты	T. pyriformis	0,19(±0,05)	1,20(±0,13)	8	0,935	0,13	85,9	0,857
23	Кислоты	Мыши	0,13(±0,05)	1,02(±0,07)	16	0,934	0,21	199,6	0,917

ЛИТЕРАТУРА

1. Schultz T.W., Cronin M.T.D., Walker J.D. and Aptula A.O. J. Mol. Struct. (Theochem). 2003. Vol. 622. P. 1–22.
2. Vermue M., Sikkema J., Verheul A., Bakker R. and Tramper J. Biotechnol. Bioeng. 1993. Vol. 42. P. 747–758.

3. *Альберт А.* Избирательная токсичность, Т. 1, Медицина, Москва. 1989, с. 58–60.
4. *Hau K.M., Connell D.W. and Richardson B.J.* Environ. Toxicol. Pharmacol. 1999. Vol. 7(1). P. 59–167.
5. *Devillers J. and Devillers H.* SAR and QSAR in Environ. Res. 2009. Vol. 20(5). P. 467–500.
6. *Yan D., Jiang X., Xu S., Wang L., Bian Y. and Yu G.* Chemosphere. 2008. Vol. 71(10). P. 1809–1815.
7. *Yan D., Jiang X., Yu G., Zhao Z., Bian Y. and Wang F.* Chemosphere. 2006. Vol. 63(5). P. 744–750.
8. *Зенкевич И.Г.* Успехи современного естествознания. 2006. № 7. С. 42–46.
9. *M.T.D. Cronin, and T.W. Schultz.* The Science of the Total Environment. 1997. Vol. 204. P. 75–88.
10. *Zhu H., Tropsha A., Fourches D., Varnek A., Papa E., Gramatica P., Öberg O., Dao P., Cherkasov A. and Tetko I.* J. Chem. Inf. Model. 2008. Vol. 48. P. 766–784.
11. *Ellison C.M., Cronin M.T.D., Madden J.C. and Schultz T.W.* SAR and QSAR in Environ. Res. 2008. Vol. 19(7). P. 751–783.
12. *Katritzky A.R., Tatham D.B. and Maran U.* J. Chem. Inf. Comput. Sci. 2001. Vol. 41(5). P. 1162–1176.
13. *Russom C.L., Bradbury S.P., Broderius S.J., Hammermeister D.E. and Drummond R.A.* Environ. Toxicol. Chem. 1997. Vol. 16(5). P. 948–967.
14. SYMYX Database. <http://www.symyx.com/products/databases/index.jsp>.
15. *Форсайт Дж., Малькольм М., Моулер К.* Машинные методы математических вычислений, Мир, Москва. 1980. С. 86–248.
16. *Nendza M. and Wenzel A.* Environ. Sci. Pollut. Res. 2006. Vol. 13(3). P. 192–203.
17. *Wang X., Sun C., Wang Y. and Wang L.* Chemosphere. 2002. Vol. 46. P. 153–161.
18. *Lessigiarska I., Worth A.P., Sokull-Kluttgen B., Jeram S., Dearden J.C., Netzeva T.I. and Cronin M.T.D.* SAR and QSAR in Environ. Res. 2004. Vol. 15(5–6). P. 413–431.
19. *Vaal M., Van der Wal J.T., Hermens J. and Hoekstra J.* Chemosphere. 1997. Vol. 35(6). P. 1291–1309.
20. *Seward J.R., Hamblen E.L. and Schultz T.W.* Chemosphere. 2002. Vol. 47. P. 93–101.
21. *Dimitrova S.D., Mekenyan O.G., Sinks G.D. and Schultz T.W.* J. Mol. Struct. (Theochem). 2003. Vol. 622. P. 63–70.

*О.А. Раевский¹, С.Л. Солодова¹, О.Е. Раевская¹,
Я.В. Липлавский¹, Р. Манхольд²*

КОМПЬЮТЕРНЫЕ КЛАССИФИКАЦИОННЫЕ МОДЕЛИ ВЗАИМОСВЯЗИ СТРУКТУРЫ ХИМИЧЕСКИХ СОЕДИНЕНИЙ И ЛЕКАРСТВ С ИХ СПОСОБНОСТЬЮ ПРОНИКАТЬ ЧЕРЕЗ ГЕМАТОЭНЦЕФАЛИЧЕСКИЙ БАРЬЕР

*¹Институт физиологически активных веществ РАН,
Черноголовка, Московская область*

²Университет им. Генриха Гейна, Дюссельдорф, Германия

Введение

Часто нарушения деятельности ЦНС являются результатом влияния на организм различных эндогенных и экзогенных факторов. Для того чтобы ЦНС нормально и правильно функционировала, в процессе эволюции организмом был выработан ряд собственных защитных механизмов, которые обеспечивают независимость работы ЦНС даже в агрессивных условиях. Одним из таких важнейших механизмов является гематоэнцефалический барьер (ГЭБ). Под этим термином условно объединяют совокупность анатомических и функциональных систем, которые поддерживают постоянство внутренней среды мозга и регулируют ее состав [1, 2].

ГЭБ регулирует прохождение различных веществ между током крови и мозгом и обладает высокой избирательной проницаемостью. Барьер предотвращает проникновение вредных веществ из крови в мозг и активно выводит их в том случае, если они каким-то образом попали в мозг. ГЭБ также играет важную роль в патологии таких заболеваний, как инсульты, микробные инфекции, черепно-мозговые травмы, нейродегенеративные заболевания, которые, в свою очередь, могут приводить к изменениям проницаемости ГЭБ. С другой стороны, ГЭБ непроницаем для многих лекарственных препаратов, что может служить препятствием при лечении заболеваний [2].

Молекулы веществ могут проходить через такую сложно устроенную систему двумя путями: за счет пассивного и активного транспорта. Пассивная диффузия осуществляется по градиенту концентрации из тока крови в мозг без энергетических затрат. Активный транспорт осуществляется против градиента концентрации, такой перенос требует затраты энергии и осуществляется при помощи белков-транспортёров.

В связи с указанной важностью процесса проникновения лекарств через ГЭБ за последние двадцать лет был проведен целый ряд исследований по установлению взаимосвязи между структурой органических соединений и их способностью проникать через ГЭБ. Здесь можно упомянуть об оценке проницаемости Сасо-2 [3], классификации действия соединений на систему ЦНС [4, 5], а также проницаемости через ГЭБ [6, 7]. Во всех указанных публикациях использовался целый набор различных дескрипторов, и значимость и влияние этих дескрипторов на процесс транспорта веществ в организме продолжает оставаться предметом оживленных дискуссий. Следует отметить, что с годами росло и число соединений, вовлеченных в расчет. Так в публикации [3] были использованы только 17 лекарств, в публикации [5] – 125, а в последующих работах уже сотни химических соединений и лекарств.

В целом сложилось впечатление, что основными факторами, влияющими на проницаемость веществ через ГЭБ, являются стерические взаимодействия, липофильность и способность соединений к образованию водородных связей.

Представленная в настоящем сообщении работа посвящена построению классификационной модели проницаемости веществ через ГЭБ путем пассивного транспорта (проникающие соединения обозначаются как ВВВ+ и непроникающие как ВВВ–) на основе рассмотрения данных для 1513 химических соединений и лекарств. Авторы рассматривают данную работу в качестве первого шага в системном исследовании этого чрезвычайно сложного биологического процесса, и главным результатом этой работы должно было стать выявление физико-химических дескрипторов, оказывающих влияние на изучаемую классификацию.

Использовались два независимых метода классификации соединений: метод структурной аналогии (read-across) и линейный дискриминантный анализ (ЛДА) на основе физико-химических дескрипторов.

Методика

Данные

Во многих работах проникновение химических соединений через ГЭБ отождествляется с непосредственным действием веществ на ЦНС. Нет сомнений в том, что вещества, оказывающие действие на ЦНС, конечно, проходят через барьер, однако существует и ряд веществ, которые проходят

через ГЭБ, но не являются ЦНС-активными. По этой причине для настоящей работы требовался тщательный отбор литературных данных, конкретно связанных с проницаемостью через ГЭБ. Здесь за основу были взяты данные по проницаемости ГЭБ, отобранные и проанализированные в публикации [6]. При этом из указанного массива соединений были исключены данные по проницаемости 91 соединения, которые согласно [6] проявляют свойства гликопротеиновых субстратов. В качестве веществ, проникающих через ГЭБ (BBB+), рассматривались соединения, имеющие величину $\log BB \geq -1$, а вещества непроницающие (BBB-) с $\log BB < -1$. Среди указанных 1513 соединений 1276 являются BBB+ и 237 BBB-.

Дескрипторы

С помощью имеющихся компьютерных программ (SYBYL, HYBOT и DRAGON) в данной работе был рассчитан целый ряд дескрипторов, связанных с различными видами межмолекулярных взаимодействий:

- стерическими (MW, α , AREA);
- электростатическими (ΣQ^-).

Ввиду хорошо известной исключительной роли водородной связи во взаимодействиях органических лигандов с биологическими мишенями в живых организмах в данной работе были привлечены дескрипторы водородной связи, рассчитываемые различными подходами и авторами:

- число водородносвязанных доноров (HD) и акцепторов (HA) в молекулах [4];
- общая водородносвязанная основность (B) и кислотность (A) [7];
- водородносвязанные донорно-акцепторные факторы ($\Sigma Ca(o)$, ΣCd , $\Sigma Ca(o)d$) [8, 9];
- поверхность полярных взаимодействий PSA [10];
- поверхность полярных взаимодействий, рассчитанная с учетом силы водородной связи WEASA, WEDSA, WEASA + WEDSA [11].

Кроме того, в расчетах также привлекался в качестве дескриптора коэффициент распределения соединений в системе вода – останол ($\log P$), который является широко используемым в литературе параметром, связанным с транспортом химических веществ [12].

В таблице 1 представлены все использованные в данной работе физико-химические дескрипторы, их обозначения и ссылки.

Результаты и обсуждение

Использование метода структурной аналогии (read-across) для классификации химических соединений на BBB+ и BBB-

Метод аналогии производит оценку свойств какого-либо соединения на основе этих же свойств одного или нескольких сходных соединений [17, 18]. В англоязычной литературе этот метод именуется read-across.

Подлинными основоположниками данного метода являются Д.И. Менделеев [19] и А.М. Бутлеров [20]. Стоит вспомнить, что свойства еще не открытых элементов Менделеев предсказывал на основании их места в периодической системе, по свойствам их соседей справа и слева, сверху и снизу (read-across!!!) т. е. по соседним членам периода и группы.

Официальное современное определение метода аналогии (read-across) формулируется следующим образом: read-across – это техника для предсказания целевой информации для какого-либо химического соединения, используя информацию о другом химическом соединении, которое рассматривается как сходное каким-либо образом (структурное сходство, сходство свойств или активности) [18].

Краткие основы метода аналогии и примеры его использования для классификации представлены в публикациях [21–24].

В простейшем случае в методе аналогии данные по свойству одного соединения приписываются другому сходному соединению. В данной работе в качестве сходства были выбраны структурные формулы. А степень сходства устанавливалась посредством расчета индексов Танимото (T_c):

$$T_c = C / (A + B - C),$$

где С – общие фрагменты структуры в паре сравниваемых соединений, А и В – число структурных фрагментов в каждом соединении. С целью отбора наиболее родственных структур для каждого из 1513 соединений все остальные 1512 соединений ранжировались в порядке их значений индекса Танимото. Производилась эта процедура с помощью программы MOLDIVS [25, 26]. Для классификации отбиралось нечетное число ближайших структурных аналогов. И рассматриваемому соединению приписывалось свойство, которое было у большинства отобранных аналогов. Результаты этой работы суммированы в таблице 2.

В качестве первого шага классификация была выполнена без какого-либо порогового значения индекса Танимото для структурных аналогов. Это позволило вовлечь в классификацию все 1513 соединений, включая те, которые не имели «хороших» структурных аналогов. Как видно из таблицы, уже использование одного ближайшего структурного аналога привело к хорошей классификации: правильное отнесение было сделано для 222 ВВВ– соединений (93,7%) и для 1262 ВВВ+ соединений (98,9%). Дальнейшее привлечение трех и пяти ближайших структурных аналогов еще улучшили распознавание ВВВ+ соединений (до 99,4% и 99,5% соответственно).

Повышение порога структурного сходства приводит к улучшению правильного распознавания как ВВВ+, так и ВВВ– соединений. Так, при пороге $T_c \geq 0,50$ и использовании пяти ближайших структурных аналогов пра-

вильно распознались 100% ВВВ+ соединений и 98,8% ВВВ– соединений. Однако при этом общее число вовлеченных в классификацию соединений резко сократилось (до 161 ВВВ– и 478 ВВВ+ соединений).

Для рассматриваемого ряда соединений оптимальным вариантом классификации является использование трех ближайших структурных аналогов без ограничений порога структурного сходства. Это позволяет провести классификацию всех соединений и правильно распознать свойства 98,5% соединений (1490 из 1513).

Полученные хорошие результаты классификации ВВВ+ и ВВВ– соединений методом структурной аналогии позволяет предполагать наличие в молекулах структурных фрагментов, которые преимущественно находятся в соединениях, проходящих и не проходящих через ГЭБ. Для определения таких структурных фрагментов была оценена их пропорция в соединениях ВВВ+ и ВВВ– с помощью программы MOLDIVS [25]. При фрагментации с использованием второй координационной сферы данный массив содержал 3870 фрагментов. Поскольку данный ряд (1513 соединений) состоял из 237 ВВВ– соединений (что составляет 15,7% от общего числа) и 1276 ВВВ+ соединений (84,3%), естественно предполагать, что фрагменты, не влияющие на классификацию ВВВ+ и ВВВ–, должны быть распределены с таким же соотношением. В таблице 3 представлены структурные фрагменты, пропорция которых в ВВВ– больше 30%, а в ВВВ+ меньше 70% соединений существенно отличается от указанного соотношения (15,7% и 84,3%).

Как видно из таблицы 3, структурные фрагменты, имеющиеся преимущественно в ВВВ– соединениях, насыщены водородосвязанными донорными и акцепторными группами. Это согласуется с высказанным ранее нами и другими авторами предположением о сильном влиянии водородных связей на процессы транспорта органических соединений в живых организмах. Для дальнейшего обсуждения этого вопроса в настоящей работе были привлечены различные физико-химические дескрипторы.

Взаимосвязи физико-химических дескрипторов с проницаемостью через ГЭБ

На первом этапе этой работы каждый из приведенных в таблице 1 физико-химических дескрипторов ранжировался в порядке их рассчитанных значений. Затем колонка этих значений сопоставлялась с колонкой, содержащей информацию о свойстве (ВВВ+ или ВВВ–). Это дало возможность сравнить на качественном уровне способность различных физико-химических дескрипторов дискриминировать соединения ВВВ+ и ВВВ–.

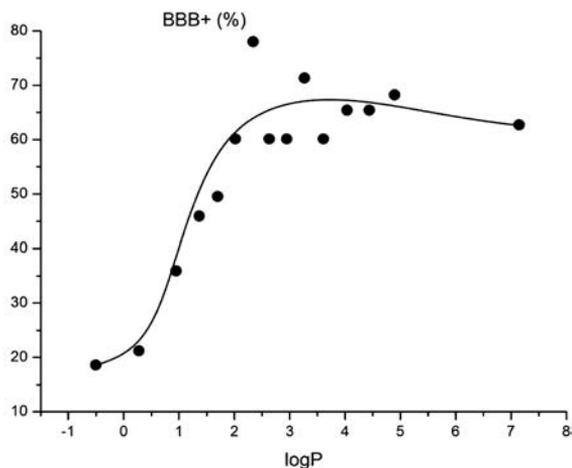


Рис. 1. Графическая зависимость доли ВВВ+ соединений от коэффициента распределения в системе октанол – вода

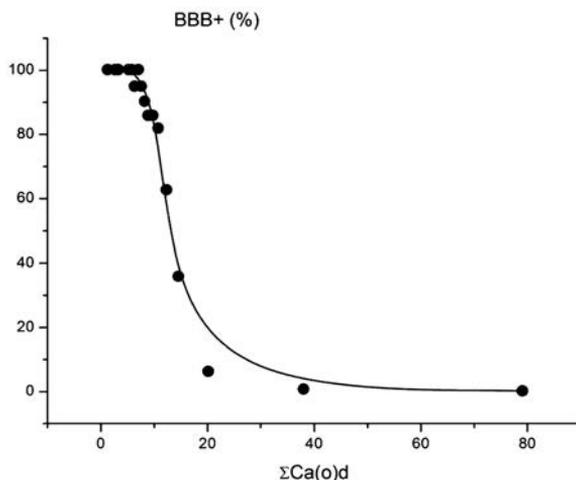


Рис. 2. Графическая зависимость доли ВВВ+ соединений от суммы водородносвязанных факторов

В качестве примера на рисунках 1 и 2 представлены зависимости доли ВВВ+ соединений от липофильности ($\log P$) и суммы водородносвязанных факторов. В случае использования дескриптора $\log P$ зависимость доли ВВВ+ соединений носит S-образный вид и не дает возможности дискриминировать ВВВ+ и ВВВ- соединения. Совершенно иная картина наблюдается при использовании суммы водородносвязанных факторов. В интервале этих факторов от 0,00 до 5,00 все соединения (их 187!) являются ВВВ+. А

при $\sum Ca(o)d \geq 25,00$ все соединения являются ВВВ⁻. В интервале же 5,00 – 25,00 происходит постепенное уменьшение доли ВВВ⁺ соединений по мере увеличения значений этого дескриптора. Тем самым очевидна хорошая способность этого водородносвязанного дескриптора дискриминировать соединения ВВВ⁺ и ВВВ⁻.

На заключительном этапе работы был привлечен метод линейного дискриминантного анализа (ЛДА) для классификации ВВВ⁺/ВВВ⁻ соединений. Использовалась программа SSPS Statistics 17.0 [27]. Результаты представлены в таблице 4.

Как очевидно из таблицы, наилучший результат классификации с помощью одиночной независимой переменной был получен при использовании НУВОТ-дескриптора $\sum Ca(o)d$ (минимальное значение λ Уилкса, максимальное значение критерия Фишера, максимальная чувствительность и точность). Другие водородносвязанные дескрипторы также обеспечивают достаточно хорошую классификацию, показывая несколько худшие статистические критерии дискриминантной функции. Дескриптор, отвечающий за электростатические контакты ($\sum Q^-$) провел дискриминацию с точностью 84,7%. Одиночное использование дескрипторов, связанных со стерическими взаимодействиями (MW, α , AREA, соответственно), показали средний результат (точность до 80%). При использовании параметра гидрофобности logP был получен наихудший результат классификации.

Высокая закоррелированность водородносвязанных дескрипторов между собой не позволяет включать в дискриминантный анализ какое-либо их сочетание. Присоединение же к какому-либо водородносвязанному дескриптору параметров, связанных со стерическими и/или электростатическими взаимодействиями немного улучшает статистику. Однако при этом модели, основанные на этих комбинациях, уступают по статистическим критериям классификации модели с использованием в качестве одиночного дескриптора $\sum Ca(o)d$ (точность 94,6%). Модели же с участием $\sum Ca(o)d$ и α и/или $\sum Q^-$ не увеличили точности классификации, что естественно в виду близости указанных значений к предельной точности самих использованных данных.

Выводы

1. Методы структурной аналогии и дискриминантного анализа на базе водородносвязанных дескрипторов позволили успешно провести классификацию 1513 разнообразных химических соединений по их способности проникать через ГЭБ по пассивному механизму транспорта.

2. Подтверждается предположение Липинского [29], что сумму водородносвязанных донорных и акцепторных факторов ($\sum Ca(o)d$) можно рассматривать в качестве глобальной меры проницаемости веществ в живых ор-

ганизмах и что этот дескриптор должен использоваться при отборе новых эффективных веществ с хорошими параметрами проницаемости (смотри рисунок 3 из [29]).

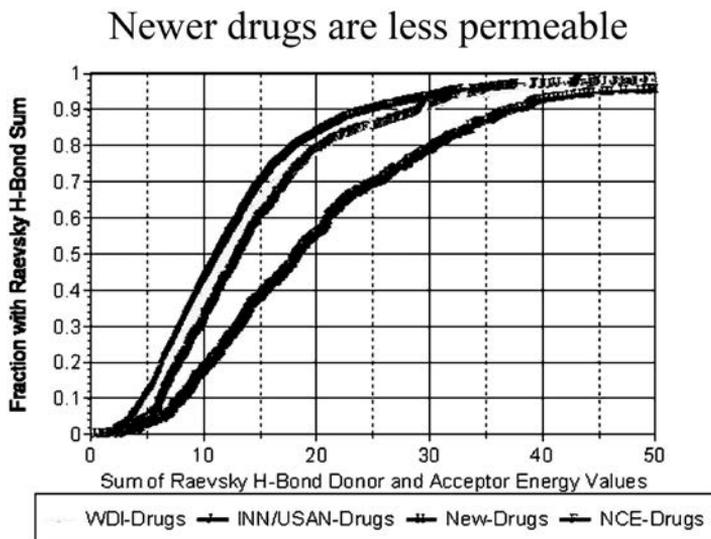


Рис. 3. Графическое сравнение фракции химических соединений в пределах рассмотренных в [29] четырех обширных баз данных, содержащих многие тысячи соединений: Derwent World Drug Index (WDI), International Non-Proprietary Names (INN)/US Adopted Names (USAN), New Chemical Entities (NCE), New Drugs. от $\Sigma Ca(o)d$. Из рисунка очевидно, что только 40% соединений из базы New Drugs имеет значение $\Sigma Ca(o)d \leq 15,00$ (примерный порог проницаемости), в то время как для других трех баз данных при указанном значении $\Sigma Ca(o)d$ фракция таких веществ достигает 70%.

Табл. 1. Физико-химические дескрипторы

N	Название дескриптора	Обозначение	Программа	Ссылка
1	Молекулярный вес	MW	HYBOT	[13]
2	Молекулярная поляризуемость	A	HYBOT	[13]
3	Молекулярная поверхность	AREA	SYBYL	[14]
4	Сумма отрицательных зарядов	ΣQ^-	HYBOT	[13]
5	Число доноров водородной связи	HD	DRAGON	[16]
6	Число акцепторов водородной связи	HA	DRAGON	[16]
7	Общая водородносвязанная кислотность	A	Данные, рассчитанные Algorithm Builder [28] из публикации [7]	[7]

Окончание таблицы 1

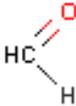
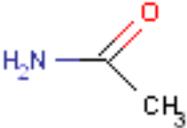
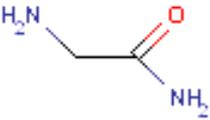
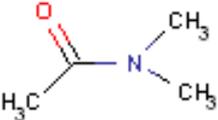
№	Название дескриптора	Обозначение	Программа	Ссылка
8	Общая водородносвязанная основность	B	Данные, посчитанные Algorithm Builder [28] из публикации [7]	[7]
9	Сумма водородносвязанных акцепторных факторов молекулы	$\Sigma Ca(o)$	HYBOT	[13]
10	Сумма водородносвязанных донорных факторов молекулы	ΣCd	HYBOT	[13]
11	Абсолютное значение суммы водородносвязанных донорных и акцепторных факторов	$\Sigma Ca(o)d$	HYBOT	[13]
12	Водородносвязанная энтальпийная акцепторная поверхность на Ван-дер-Ваальсовских сферах	WEASA	3D HYBOT	[15]
13	Водородносвязанная энтальпийная донорная поверхность на Ван-дер-Ваальсовских сферах	WEDSA	3D HYBOT	[15]
14	Поверхность полярных взаимодействий	PSA	SYBYL	[14]
15	Липофильность	logP	DRAGON	[16]

Табл. 2. Структурное сходство

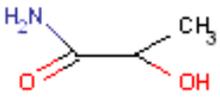
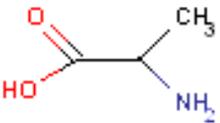
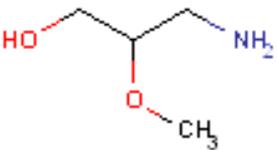
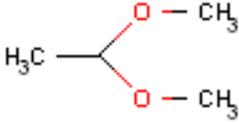
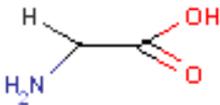
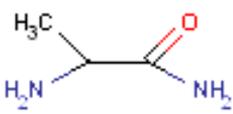
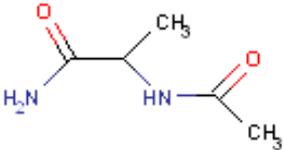
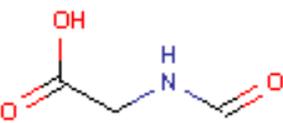
Tс≥	Число ближайших структурных соседей	Число соединений			Правильное распознавание	
		BBB-	BBB+	BBB+/-	BBB-(%)	BBB+(%)
0,0	1	237	1276	1513	222(93,7)	1262(98,9)
0,0	3	237	1276	1513	222(93,7)	1268(99,4)
0,0	5	237	1276	1513	222(93,7)	1270(99,5)
0,1	1	237	1276	1513	222(93,7)	1262(98,9)
0,1	3	237	1276	1513	222(93,7)	1268(99,4)
0,1	5	237	1276	1513	222(93,7)	1270(99,5)
0,2	1	237	1273	1510	222(93,7)	1259(98,9)
0,2	3	237	1266	1503	222(93,7)	1258(99,4)
0,2	5	237	1258	1495	222(93,7)	1252(99,5)
0,3	1	234	1241	1475	219(93,6)	1227(98,9)

Tc ≥	Число ближайших структурных соседей	Число соединений			Правильное распознавание	
		ВВВ–	ВВВ+	ВВВ+/-	ВВВ–(%)	ВВВ+(%)
0,3	3	232	1159	1391	219(94,4)	1152(99,4)
0,3	5	225	1105	1330	210(93,3)	1100(99,5)
0,4	1	224	1114	1338	213(95,1)	1103(99,0)
0,4	3	208	900	1108	201(96,6)	899(99,9)
0,4	5	202	770	972	196(97,0)	769(99,9)
0,5	1	218	941	1159	210(96,3)	932(99,0)
0,5	3	188	618	806	185(98,4)	617(99,8)
0,5	5	161	478	639	159(98,8)	478(100)

Табл. 3. Структурные фрагменты

Функциональная группа	ВВВ+/- (1513)	ВВВ– (237) (15,7%)	ВВВ+ (1276) (84,3%)
	417	171 (41,0%)	246 (59,0%)
	17	7 (41,2%)	10 (58,8%)
	565	175 (30,9%)	390 (69,1%)
	235	132 (56,2%)	130 (43,8%)
	309	130 (42,1%)	179 (57,9%)

Окончание таблицы 3

Функциональная группа	BBB+/- (1513)	BBB- (237) (15,7%)	BBB+ (1276) (84,3%)
	35	22 (62,9%)	13 (37,1%)
	79	62 (78,5%)	17 (21,5%)
	58	36 (62,1%)	22 (37,9%)
	112	62 (55,4%)	50 (44,6%)
	153	131 (85,6%)	22 (14,4%)
	164	131 (79,9%)	33 (20,1%)
	128	120 (93,8%)	8 (6,2%)
	140	129 (92,1%)	11 (7,9%)

**Избранные труды доктора химических наук, профессора О.А. Раевского
в области молекулярного компьютерного дизайна**

Табл. 4. Параметры и статистические критерии дискриминантной функции $y = a_0 + a_1x_1 + \dots + a_nx_n$

x_1	x_2	x_3	x_4	a_0	a_1	a_2	a_3	a_4	λ	F	P	D ²	TP	Чувствительность, %	FN	TN	Специфичность, %	FP	Точность, %
	-	-	-	-2,89	0,008	-	-	-	0,766	461,2	<0,001	2,31	1042	81,7	234	169	71,3	68	80,0
A	-	-	-	-2,96	0,81	-	-	-	0,845	276,2	<0,001	1,38	957	75,0	319	148	62,4	89	73,0
area	-	-	-	-3,79	0,006	-	-	-	0,924	123,5	<0,001	0,62	903	70,8	373	140	59,1	97	68,9
$\Sigma Q-$	-	-	-	2,18	1,25	-	-	-	0,683	701,1	<0,001	3,51	1132	88,7	144	149	62,9	88	84,7
HD	-	-	-	0,817	0,473	-	-	-	0,842	282,8	<0,001	1,42	1097	86,0	179	137	57,8	100	81,6
HA	-	-	-	-1,66	0,319	-	-	-	0,775	437,6	<0,001	2,19	1085	85,0	191	155	65,4	82	82,0
HD+HA	-	-	-	-1,44	0,209	-	-	-	0,772	445,4	<0,001	2,23	1158	90,8	118	148	62,4	89	86,3
HD+HA	MW	-	-	-2,68	0,126	0,005	-	-	0,689	340,7	<0,001	3,41	1132	88,7	144	163	68,8	74	85,6
HD+HA	MW	$\Sigma Q-$	-	2,01	-0,089	0,002	1,169	-	0,643	279,9	<0,001	4,21	1152	90,3	124	167	70,5	70	87,2
A	-	-	-	-0,95	2,27	-	-	-	0,831	307,4	<0,001	1,54	1096	85,9	180	152	64,1	85	82,5
B	-	-	-	-1,92	1,08	-	-	-	0,801	376,3	<0,001	1,88	1098	81,6	178	150	63,3	87	82,5
A+B	-	-	-	-0,228	1,228	-	-	-	0,781	423,5	<0,001	2,12	1145	89,7	131	150	63,3	87	85,6
A+B	A	-	-	-2,67	0,587	0,038	-	-	0,743	260,7	<0,001	2,61	1111	87,1	165	153	64,4	84	83,5
A+B	A	$\Sigma Q-$	-	1,64	-0,30	0,042	1,44	-	0,633	292,2	<0,001	4,39	1153	90,4	123	166	70,0	71	87,2
$\Sigma Ca(o)$	-	-	-	-2,204	0,319	-	-	-	0,573	1124,9	<0,001	5,63	1207	94,6	69	168	70,9	69	90,9
ΣCd	-	-	-	1,09	0,348	-	-	-	0,535	1313,9	<0,001	6,57	1227	96,2	49	165	69,6	72	92,0

Окончание таблицы 4

$\Sigma Ca(o)d$	-	-	-1,91	0,190	-	-	0,448	1583,8	<0,001	7,92	1250	98,0	26	182	76,8	55	94,6
$\Sigma Ca(o)d$	A	-	-1,67	0,200	-0,010	-	0,486	799,7	<0,001	8,01	1247	97,7	29	182	76,8	55	94,4
$\Sigma Ca(o)d$	A	ΣQ	-1,71	0,208	-0,005	0,118	0,485	800,1	<0,001	8,01	1247	97,7	29	182	76,8	55	94,4
$\Sigma Ca(o)$	ΣCd	-	1,80	-0,159	0,223	-	0,486	789,5	<0,001	7,99	1248	97,8	28	180	75,9	57	94,4
$\Sigma Ca(o)$	ΣCd	α	1,69	-0,183	0,213	0,007	0,486	798,4	<0,001	7,99	1246	97,6	30	182	76,8	55	94,4
$\Sigma Ca(o)$	ΣCd	α	-1,71	0,169	-0,215	0,004	0,485	798,5	<0,001	7,99	1247	97,7	29	181	76,4	56	94,4
WEASA	-	-	-1,89	0,136	-	-	0,759	480,9	<0,001	2,41	1136	89,0	140	155	65,4	82	85,3
WEDSA	-	-	-0,89	0,310	-	-	0,799	380,4	<0,001	1,90	1101	86,3	175	160	67,5	77	83,3
WEASA+	-	-	-1,69	0,101	-	-	0,746	514,9	<0,001	2,58	1158	90,8	118	157	66,2	80	86,9
WEDSA	-	-	-2,62	0,079	0,036	-	0,708	311,7	<0,001	3,12	1170	91,7	106	169	71,3	68	88,5
WEASA+	A	ΣQ	1,71	-0,049	0,039	1,34	0,609	323,3	<0,001	4,86	1167	91,5	109	179	75,5	58	88,9
WEDSA	-	-	-1,41	0,018	-	-	0,732	554,5	<0,001	2,77	1145	89,7	131	155	65,4	82	85,9
PSA	A	-	-2,40	0,015	0,035	-	0,692	335,3	<0,001	3,36	1128	88,4	148	163	68,8	74	85,3
PSA	A	ΣQ	1,62	-0,009	0,036	1,28	0,603	331,8	<0,001	4,99	1148	89,9	128	176	74,3	61	87,5
logP	-	-	-1,18	0,491	-	-	0,921	130,6	<0,001	0,65	885	69,4	391	146	61,6	91	68,1

a_p, a_1, \dots, a_n – коэффициенты функции, x_1, x_2, \dots, x_n – дескрипторы, λ – лямбда Уилкса (изменяется от нуля при идеальной дискриминации до единицы при отсутствии дискриминации), F – критерий Фишера, r – значимость, D^2 – расстояние Махаланобиса (мера разделения двух групп), TP – число правильно распознанных соединений среди BBV^+ , TN – число правильно распознанных соединений среди BBV^- , FP – число неправильно распознанных соединений среди BBV^+ , FN – число неправильно распознанных соединений среди BBV^- , Чувствительность = $TP/(TP + FN)$, Специфичность = $TN/(TN + FP)$, Точность = $(TP + TN)/(TP + TN + FN + FP)$

Работа выполнена при поддержке МНТЦ (проект 3777).

ЛИТЕРАТУРА

1. Бредбери М. Концепция гематоэнцефалического барьера (пер. с англ.). М., Медицина, 1983.
2. Покровский В.М., Коротько Г.Ф., Кобрин В.И. и др. Физиология человека: Учебник. М. Медицина, 1997.
3. Waterbeemd H., Camenisch G., Folkers G., Raevsky O.A. Quant. Struct.-Act. Relat. 1996. Vol. 15. P. 480–490.
4. Lipinski C.A., Lombardo F., Dominy B.W., Feeney P.J. Adv. Drug Delivery Rev. 1997. Vol. 23. P. 3–25.
5. van de Waterbeemd H., Camenisch G., Folkers G., Chretien J. R., Raevsky O.A. J. Drug Targeting. 1998. Vol. 6. P. 151–165.
6. Adenot M., Lahana R.J. Chem. Inf. Comput. Sci. 2004. Vol. 44. P. 239–248.
7. Zhao Y.H., Abraham M.H., Ibrahim A., Fish P.V., Cole S., Lewis M.L., de Groot M.J., Reynolds D.P. J. Chem. Inf. Model. 2007. Vol. 47. P. 170–175.
8. Raevsky O.A. Minireview in Medicinal Chemistry. 2004. Vol. 4. P. 1041–1052.
9. Raevsky O.A. In: Molecular Drug Properties. Measurement and Prediction. (R. Mannhold Eds.) 2008. Wiley-VCH Verlag GmbH & Co.KGaA, Weinheim, pp. 127–154.
10. Ertl P. Polar surface area In Molecular Drug Properties. Measurement and Prediction. (R. Mannhold Eds.) 2008. Wiley-VCH Verlag GmbH & Co.KGaA, Weinheim, pp. 127–154.
11. Raevsky O.A., Skvortsov V.S. SAR & QSAR in Environmental Research. 2005. Vol. 15. P. 287–300.
12. Hansch C., Leo A., and Hoekman D. Exploring QSAR: Hydrophobic, Electronic, and Steric Constants. 1995. American Chemical Society, Washington, DC.
13. Раевский О.А., Григорьев В.Ю., Трепалин С.В. (26.02.1999) Программа HYBOT, Роспатентом № 990090.
14. SYBYL: SYBYL 8.0, Tripos International, 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA.
15. Раевский О.А., Скворцов В.С. (27.09.2004) Программа 3D HYBOT, Роспатент № 2004612207.
16. DRAGON (<http://www.vcclab.org/lab/edragon/>)
17. Organization for Economic Cooperation and Development. Guidance on Grouping of Chemicals, OECD, Paris. 2007.
18. European Chemical Agency, Guidance on Information Requirements and Chemical Safety Assessment, Chapter R6, QSAR and grouping of chemicals, ECHA, Helsinki. 2008.
19. Mendelejeff D. Zeitschrift fur Chemie. 1869. Vol. 12. P. 405–406.
20. Butlerov A.M. Z. Chem. Pharm. 1861. Vol. 4. P. 549–555.
21. Dimitrov S., Mekenyan O. An Introduction to Read-Across for the Prediction of the Effects of Chemicals, in: In Silico Toxicology. Principles and Ap-

plications. (M.T.D. Cronin and J.C. Madden Eds.), 2010. RSC, Cambridge, pp. 372–384.

22. *Diderich R.* Tools for Category Formation and Read-Across: Overview of the OECD (Q)SAR Application Toolbox, *ibid.*, pp. 385–407.

23. *Jeliazkova N., Jaworska J. and Worth A.P.* Open Source Tools for Read-Across and Category Formation, *ibid.*, pp. 408–445.

24. *Raevsky O.A., Grigor'ev V.Yu., Modina E.A. and Worth A.* SAR & QSAR in Environmental Research. 2010. Vol. 21. P. 265–275.

25. *Раевский О.А., Герасименко В.А., Трепалин С.В.* (26.02.1999) Программа MOLDIVS, *Роспатент* № 990093.

26. *Gerasimenko V.A., Trepalin S.V., Raevsky O.A.* MOLDIVS – A New Program for Molecular Similarity and Diversity Calculations, in: *Molecular Modeling and Prediction of Bioactivity* (eds. K. Gundertofe and F. Jorgensen), 2000. Kluwer Academic/Plenum Publ., pp. 423–424.

27. SPSS (<http://www.spss.com>)

28. *PharmaAlgorithms. ADME Boxes, Version 3.0*; PharmaAlgorithms, Inc.: 591 Indian Road, Toronto, ON M6P 2C4, Canada.

29. *Lipinski C.A.* Drugs structure and properties, past and present. Can we design drugs with beautiful properties? http://www.iaim.demon.co.uk/spring99/lipins_n.pdf.

В.Ю. Григорьев, О.А. Раевский

ИССЛЕДОВАНИЕ КОЛИЧЕСТВЕННОЙ СВЯЗИ МЕЖДУ ФУНКЦИЕЙ РАДИАЛЬНОГО РАСПРЕДЕЛЕНИЯ И КРИВОЙ ДАВЛЕНИЯ НАСЫЩЕННОГО ПАРА В РЯДУ N-АЛКАНОВ НА ОСНОВЕ МОДИФИЦИРОВАННОЙ ЛИНЕЙНОЙ ДИНАМИЧЕСКОЙ МОДЕЛИ

*Институт физиологически активных веществ РАН,
Черноголовка, Московская область*

Введение

При исследовании количественной связи «структура – активность» (КССА) и «структура – свойство» (КССС) одной из основных целей является конструирование предсказательных моделей биологической активности и физико-химических свойств как функции молекулярной и структурной информации в рядах химических соединений [1, 2]. В настоящее время широкое распространение для изучения таких связей получили такие методы, как множественный регрессионный анализ, проекции на латентные структуры, нейронные сети и метод опорных векторов [3]. Общей характерной чертой рассматриваемых подходов является то, что в качестве независимых переменных выступают различные физико-химические, электронные, топологические дескрипторы, представленные в виде матрицы, а в качестве зависимой переменной – вектор откликов. При геометрической интерпретации каждая молекула представляется точкой в многомерном пространстве признаков. При этом точечные значения для некоторых экспериментальных дескрипторов извлекаются из соответствующих двумерных кривых. Например, нормальная точка кипения вещества присутствует в кривой зависимости давления насыщенного пара от температуры, а величина LD_{50} берется из графика «доза – эффект».

Известно [4], что кривая «доза – эффект» отражает свойства как вещества, так и организма, на который оно действует, т. е. содержит информацию о механизме действия. Очевидно, что количество информации, содержащейся в двумерной кривой, больше, чем в отдельной точке. Поэтому представляет интерес исследование количественной связи между структурой

и свойствами молекул, выраженных в виде двумерных кривых. Одной из форм представления структуры молекул в КССА/КССС-исследованиях является функция радиального распределения, предложенная в работе [5] и нашедшая дальнейшее применение [6–8]. А в качестве одного из подходов для установления количественных закономерностей может быть использована линейная динамическая модель [9].

Целью настоящей работы является исследование количественной связи между структурой молекул, выраженной в виде функции радиального распределения, и кривой зависимости давления насыщенного пара от температуры в ряду *n*-алканов на основе линейной динамической модели.

Экспериментальная часть

Структуру десяти *n*-алканов (пропан, бутан, ... додекан) рассчитывали с использованием программного комплекса HyperChem [10]. В качестве стартовых молекул при проведении полной оптимизации служили генерируемые программой трехмерные структуры. Молекулы оптимизировали на основе полуэмпирического квантовохимического метода AM1 в приближении ограниченного метода Хартри – Фока с использованием алгоритма Полака – Рибера, среднеквадратичным градиентом 0,1 ккал/(Å моль) в вакууме.

Функцию радиального распределения рассчитывали на основе математического аппарата газовой электронографии, модифицированного для использования в КССА/КССС-исследованиях [5]:

$$f(r) = \frac{2}{\pi} \int_0^{s_{\max}} s M(s) \exp(-bs^2) \sin sr ds$$

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

$$s = \frac{4\pi}{\lambda} \sin \frac{\theta}{2}$$

λ – длина волны, θ – угол рассеяния, b – параметр, оцениваемый из уравнения:

$$\exp(-bs_{\max}^2) = 0,1$$

В случае жесткой модели молекулы (отсутствие внутримолекулярных колебаний):

$$M(s) = \sum_{i=1}^N \sum_{j>i}^N C_{ij} \frac{\sin sr_{ij}}{sr_{ij}}$$

здесь N – число атомов в молекуле, C_{ij} – коэффициент, характеризующий рассеивающую способность данной пары атомов i и j .

$$C_{ij} = \frac{Z_i Z_j}{\sum_{k=1}^N (Z_k^2 + Z_k)}$$

здесь Z – порядковый номер атома.

При моделировании в качестве независимой переменной выступало межатомное расстояние r , а в качестве зависимой – функция радиального распределения $f(r)$. При этом минимальное значение r было равно 0 \AA , что было обусловлено физическим смыслом, а максимальное значение 16 \AA соответствовало самому большому межатомному расстоянию в молекуле додекана ($15,57 \text{ \AA}$).

Данные по зависимости давления насыщенного пара (P) n -алканов от температуры (T) (в форме уравнений Антуана) были взяты из работы [11]. P менялось в пределах от 10 мм Hg до 1500 мм Hg , а T варьировало в пределах от 164 К до 520 К . Учитывая, что давление насыщенного пара имело одинаковую величину интервала для всех рассматриваемых углеводородов, оно было выбрано в качестве независимой переменной, а температуры кипения – в качестве моделируемой функции.

Количественную связь между исследуемыми кривыми устанавливали на основе линейной динамической модели с одним входом $x(t)$, одним выходом $y(t)$ и импульсной переходной функцией $h(\tau)$ [9]:

$$x(t) \rightarrow \boxed{h(\tau)} \rightarrow y(t)$$

Связь между процессами $x(t)$ и $y(t)$ может быть выражена следующим образом:

$$y(t) = \int_0^{\infty} h(\tau)x(t - \tau)d\tau$$

При переходе из временной (физической) области в частотную может быть получено следующее уравнение:

$$Y(f) = H(f) X(f) \quad (1)$$

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

$$S_{xy}(f) = H(f) S_{xx}(f)$$

Здесь $S_{xy}(f)$ – взаимная двусторонняя спектральная плотность, $S_{xx}(f)$ – двусторонняя спектральная плотность $x(t)$.

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

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

где $S_{yy}(f)$ – двусторонняя спектральная плотность $y(t)$.

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 [12]. В качестве реализаций 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 реализации различной длины. При использовании $f(r)$ в качестве входной кривой (модель «структура – активность») проводили варьирование длины реализации путем разделения интервала r (0 – 16 Å) на 160 частей и парной комбинации полученных значений $r_{\text{мин}}$ и $r_{\text{макс}}$. При этом длина реализации выходной кривой $T = f(P)$ оставалась неизменной и была максимальной ($P_{\text{мин}} = 10$ мм Hg, $P_{\text{макс}} = 1500$ мм Hg). Аналогичный подход использовали для входной кривой $T = f(P)$ (модель «активность – структура»). В этом случае интервал P (10–1500 мм Hg) делили на 50 частей и комбинировали величины $P_{\text{мин}}$ и $P_{\text{макс}}$. Длина реализации $f(r)$ была постоянной ($r_{\text{мин}} = 0$ Å, $r_{\text{макс}} = 16$ Å). Полученные модели сортировали по величине остаточного стандартного отклонения и отбирали те из них, которые имели минимальное значение.

Регрессионный анализ выполняли с помощью компьютерной программы SVD [13]. Для оценки предсказательной способности исследуемых моделей применяли схему кросс-валидации с выбором по одному. В качестве статистических характеристик уравнений использовали: n – число точек, R^2 – квадрат коэффициента линейной корреляции, s – остаточное стандартное отклонение, q^2 – квадрат коэффициента линейной корреляции в условиях кросс-валидации.

Обсуждение результатов

На рисунке 1 представлен пример $f(r)$ для молекул пропана и додекана.

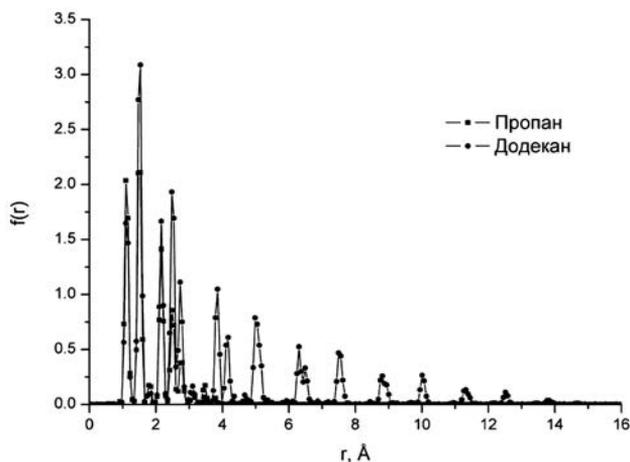


Рис. 1. Функция радиального распределения для молекул пропана и додекана при уровне дискретизации $m = 256$

Максимумы на кривых $f(r)$ соответствуют парным межатомным расстояниям. Так, первый пик, наблюдаемый в районе $r = 1,10 \text{ \AA}$, обусловлен наличием в молекулах валентно связанных атомов С–Н, а второй пик при $r = 1,54 \text{ \AA}$ соответствует валентно связанным атомам С–С. Последующие максимумы отражают расстояния между несвязанными атомами.

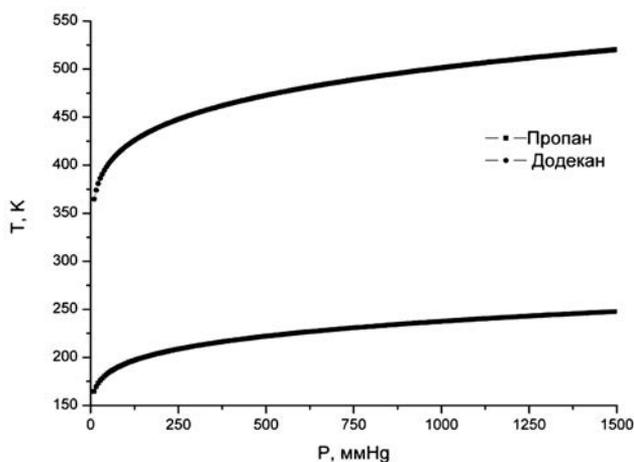


Рис. 2. Обратная функция зависимости давления насыщенного пара от температуры для молекул пропана и додекана при уровне дискретизации $m = 256$

Представленные на рисунке две кривые $T = f(P)$ являются типичными для рассматриваемого ряда *n*-алканов. Они имеют одинаковый интервал изменения по оси абсцисс (10 ... 1500 мм Нг), но отличаются значениями по оси ординат. При этом температура меняется от 164 К до 248 К (пропан) и от 364 К до 520 К (додекан).

В таблице 1 приведены характеристики лучших моделей типа «структура – активность», полученных путем перебора 12880 (161 x 160/2) реализаций различной длины на каждом уровне дискретизации. Из данных таблицы следует, что при увеличении степени дискретизации величина s имеет тенденцию к увеличению. Среднее значение функции когерентности γ^2 ведет себя противоположным образом. Длина реализации $f(r)$ принимает разные значения. При этом $r_{\text{мин}}$ меняется от 0 Å до 1,4 Å, а $r_{\text{макс}}$ меняется от 12,7 Å до 15,6 Å. 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».

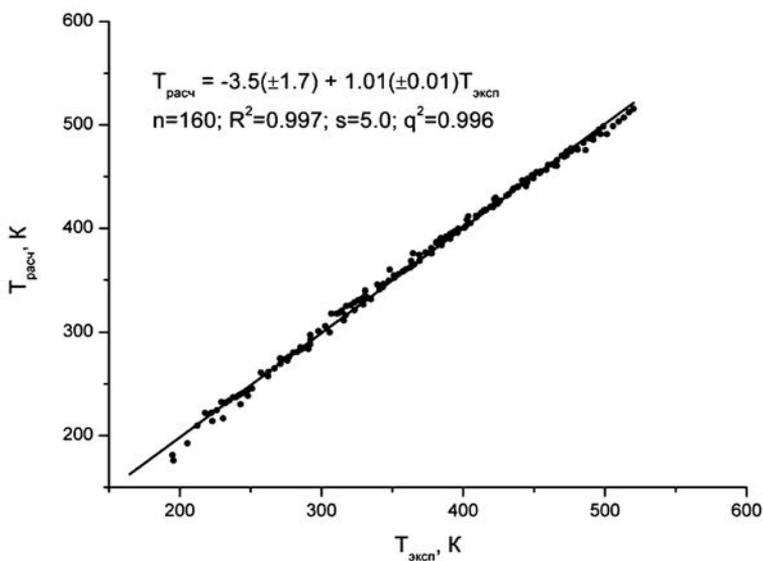


Рис. 3. Зависимость рассчитанных по модели 1 значений температур кипения от экспериментальных величин

Данные, представленные на рисунке 3, свидетельствуют о том, что модель 1 адекватно описывает кривую $T = f(P)$. Об этом говорят как высокие значения характеристик описательной ($R^2 = 0,997$, $s = 5,0$) и предсказательной ($q^2 = 0,996$) способности модели, так и величины коэффициентов.

Следует отметить, что при проведении КССА/КССС-исследований кроме задачи получения адекватных моделей, связывающих структуру молекул с их активностью и свойствами, стоят и другие задачи. В качестве своео-

бразной «сверхзадачи» можно рассматривать предсказание структур с заданными свойствами. Этой теме уделяется большое внимание в литературе [14–20]. Главным направлением в решении этой проблемы является получение качественных КССА/КССС-моделей и компьютерная генерация на их основе молекулярных структур с учетом различных ограничений: физических, химических и др.

В рамках рассматриваемой модифицированной линейной динамической системы эта «сверхзадача» решается естественным образом. Используя уравнение (1), можно при знании $H(f)$ и $Y(f)$ оценивать величину $X(f)$, и наоборот: комбинация известных величин $H(f)$ и $X(f)$ позволяет вычислить $Y(f)$. Путем обратного преобразования Фурье можно вернуться к исходной физической системе. Например, используя модель 1 и задавая кривую $T = f(P)$, можно получить $f(r)$. Однако качество получаемой функции радиального распределения оказывается слишком плохим, чтобы на ее основе делать надежные умозаключения о трехмерной структуре молекул. В самом деле, разрешение по оси ординат (в соответствии с характеристиками модели 1) составляет всего $(13,9-1,1)/(16-1) = 0,85 \text{ \AA}$, а $r_{\text{макс}} = 13,9 \text{ \AA}$, что явно недостаточно для получения надежной структурной информации. Модели 2–5 имеют более высокую разрешающую способность, однако они не охватывают весь диапазон изменения межатомных расстояний. Поэтому мы сконструировали специальную модель «активность – структура», где в качестве входа используется зависимость $T = f(P)$ переменной длины, а в качестве выхода – полноразмерная $f(r)$.

Данные, представленные в таблице 2, свидетельствуют о том, что наиболее «информативным» участком кривой $T = f(P)$ является отрезок 1470,2–1500 мм Нг. При увеличении степени дискретизации s увеличивается с 0,018 до 0,061, а γ^2 уменьшается с 0,992 до 0,940. 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 $(16 - 0)/(256 - 1) = 0,06 \text{ \AA}$, в качестве рабочей модели «активность – структура» была использована зависимость 10. Используя эту модель и схему кросс-валидации с выбором по одной молекуле, были рассчитаны величины $f(r)$ для всех исследованных n -алканов. Ниже представлено соответствующее регрессионное уравнение:

$$f(r)_{\text{расч.}} = -0,001(\pm 0,001) + 1,00(\pm 0,01) f(r)_{\text{эксп.}} \quad (2)$$

$n = 2560; R^2 = 0,971; s = 0,061; q^2 = 0,966$

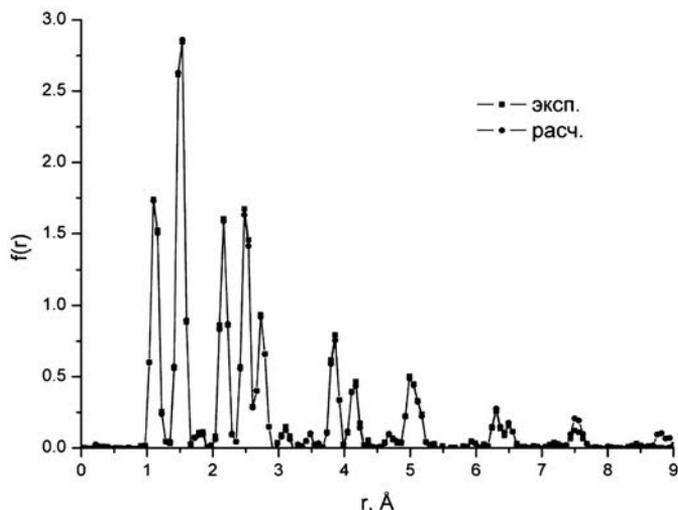


Рис. 4. Экспериментальные и рассчитанные по модели 10 значения функции радиального распределения для молекулы гептана при уровне дискретизации $m = 256$

Параметры и коэффициенты регрессионного уравнения (2) говорят о высоком качестве модели 10, которая может быть применена для вычисления $f(r)$ в исследованном ряду *n*-алканов. Об этом свидетельствует и представленный на рисунке 4 пример расчета функции радиального распределения для молекулы гептана. Из представленных данных следует, что рассчитанные и экспериментальные значения $f(r)$ хорошо совпадают как по абсциссе, так и по ординате.

Таким образом, модифицированная линейная динамическая модель может быть использована в качестве рабочего инструмента при проведении КССА/КССС-исследований не только для получения корреляционных зависимостей между структурой и активностью, но также для извлечения информации о трехмерной структуре молекул с заданными свойствами.

Табл. 1. Порядковый номер модели (N_0), уровни дискретизации (m), минимальная ($r_{\text{мин.}}$) и максимальная ($r_{\text{макс.}}$) величина межатомного расстояния, остаточное стандартное отклонение (s) и среднее значение функции когерентности (γ^2) для моделей «структура – активность»

N_0	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

Табл. 2. Порядковый номер модели (№), уровни дискретизации (m), минимальная ($P_{\text{мин.}}$) и максимальная ($P_{\text{макс.}}$) величина давления насыщенного пара, остаточное 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

ЛИТЕРАТУРА

1. Eriksson L., Jaworska J., Worth A.P. et al. Environmental Health Perspectives. 2003. Vol. 111. N 10. P. 1361.
2. Nantasenamat C., Isarankura-Na-Ayudhya C., Naenna T., Prachayasittikul V. EXCLI Journal. 2009. Vol. 8. P. 74.
3. Liu P., Long W. Int. J. Mol. Sci. 2009. Vol. 10. P. 1978.
4. Куценко С.А. Медлайн.РУ. 2003. Т. 4. Ст. 119. С. 188.
5. Новиков В.П., Раевский О.А. Хим.-фармац. журн. 1982. Т. 16. № 5. С. 574.
6. Раздольский А.Н., Трепалин С.В., Раевский О.А. Хим.-фармац. журн. 2000. Т. 34. № 12. С. 26.
7. Vračko M., Gasteiger J. Internet Electronic Journal of Molecular Design. 2002. Vol. 1. N 10. P. 527.
8. Gonzalez M.P., Teran C., Teijeira M., Helguera A.M. Bulletin of Mathematical Biology. 2007. Vol. 69. P. 347.
9. Бендат Дж., Пирсол А. Прикладной анализ случайных данных. М.: Мир, 1989. 540 с.
10. <http://www.hyper.com/>
11. Рид Р., Праусниц Дж., Шервуд Т. Свойства газов и жидкостей. Л.: Химия, 1982. 592 с.
12. Preuss R. IEEE Transactions on acoustics, speech, and signal processing. 1982. Vol. 30. N 4. P. 595.
13. Форсайт Дж., Малькольм М., Моулер К. Машинные методы математических вычислений. М.: Мир, 1980. 280 с.
14. Miyao T., Arakawa M., Funatsu K. Mol. Inf. 2010. Vol. 29. P. 111.
15. Мельников А.А., Палюлин В.А., Радченко Е.В., Зефирова Н.С. Доклады Академии Наук. 2007. Том 415. № 6. С. 763.

16. *Fink T., Reymond J.-L.* J. Chem. Inf. Model. 2007. Vol. 47. N 2. P. 342.
17. *Varnek A., Fourches D., Solov'ev V.P. et al.* J. Chem. Inf. Comput. Sci. 2004. Vol. 44. N 4. P. 1365.
18. *Brown N., McKay B., Gasteiger J.* J. Comput. Aided Mol. Des. 2004. Vol. 18. N 12. P. 761.
19. *Wieland T., Kerber A., Laue R.* J. Chem. Inf. Comput. Sci. 1996. Vol. 36. N 3. P. 413.
20. *Weis D.C., Faulon J.-L., LeBorne R.C., Visco D.P.* Ind. Eng. Chem. Res. 2005. Vol. 44. N 23. P. 8883.

*О.А. Раевский, А.Н. Раздольский, Я.В. Липлавский,
О.Е. Раевская, А.В. Ярков*

ОЦЕНКА ОСТРОЙ ТОКСИЧНОСТИ ОРГАНИЧЕСКИХ СОЕДИНЕНИЙ ПРИ ВНУТРИВЕННОМ ВВЕДЕНИИ ПО ОТНОШЕНИЮ К МЫШАМ НА ОСНОВЕ МЕЖВИДОВЫХ КОРРЕЛЯЦИЙ, ПАРАМЕТРОВ ЛИПОФИЛЬНОСТИ И ФИЗИКО-ХИМИЧЕСКИХ ДЕСКРИПТОРОВ

*Институт физиологически активных веществ РАН,
Черноголовка, Московская область*

Экспериментальная оценка биологической активности и токсичности химических соединений является дорогостоящей процедурой и сопровождается безвозвратными потерями большого числа животных. Вследствие этого задача резкого уменьшения количества тестируемых животных в Российской Федерации за счет использования альтернативных методов, например, методов *in vitro*, качественных и количественных соотношений «структура – активность» (токсичность) или информации о структурно родственных соединениях является чрезвычайно актуальной. Необходимость использования альтернатив, прямо не связанных с испытаниями на животных, подчеркивается и в официальных документах Евросоюза [1, 2].

Одним из ключевых параметров, используемых в оценке риска при применении химических соединений в промышленности, сельском хозяйстве, медицине, является острая токсичность по отношению к теплокровным, оцениваемая преимущественно на грызунах. Острая токсичность описывает неблагоприятные для жизни эффекты при однократном или многократном введении химического соединения в организм. Острая летальная доза, при которой 50% животных погибает (LD_{50}) является наиболее распространенной количественной мерой токсичности. При установлении количе-

ственных соотношений «структура – активность» (токсичность) для LD_{50} используется размерность, выраженная в молях химического соединения на килограмм живого веса испытываемого объекта (моль/кг).

Хотя значения LD_{50} широко используются в медицинской практике, этот параметр не является идеальным при установлении количественной связи «структура – активность» (КССА). Прежде всего необходимо отметить, что конвертирование сложных явлений, происходящих в целом организме и вызывающих острую токсичность, в какую-либо количественную величину сопровождается потерей части информации. Кроме того, данные по LD_{50} , полученные при различных условиях эксперимента в различных лабораториях на животных разного пола и возраста, являются весьма переменными и затрудняют процесс КССА-моделирования. Вероятно, по этим причинам опубликовано относительно мало КССА-моделей и экспертных систем для предсказания острой токсичности химических соединений по отношению к млекопитающим [3–5]. А проведенный детальный анализ 150 КССА-моделей острой токсичности по отношению к мышам и крысам привел к заключению, что большинство этих моделей имеет ограниченную пользу вследствие скромного статистического качества и малых размеров выборок, вовлеченных в моделирование [6].

В вышеуказанном обзоре [6] в качестве возможных КССА-подходов для острой токсичности по отношению к грызунам описывались межвидовые корреляции, использование параметров липофильности и различных дескрипторов. Однако отбор соединений для этих корреляций не носил системный характер. Поэтому мы решили провести подобное КССА-моделирование для отдельных рядов соединений, преимущественно связанных с определенным типом действия (mode of action, MOA). В широко используемой в КССА водной токсичности классификации Верхаара [7] выделено четыре типа действия: инертные соединения, менее инертные, реактивные и специфические.

В настоящем сообщении представлены результаты КССА-моделирования для инертных соединений.

Экспериментальная часть

Использованы данные по острой токсичности органических соединений при внутривенном введении по отношению к мышам (mice intravenously, mice iv) из *SYMUX Toxicity Database* [8], по отношению к рыбкам гуппи (*Guppy*), тупорылому пескарю (*Fathead Minnow, FHM*) и радужной форели (*Rainbow Trout, RT*) [9], к водному рачку Дафния Магна (*Daphnia Magna, DM*) [10] и к инфузории Тетрахимена (*Tetrahymena Pyriformis, TP*) [11,12].

В качестве количественной меры токсичности использовались величины $\log 1/LD_{50}$ (где LD_{50} – концентрации веществ, вызывающие летальное воздействие на 50 % испытываемых мышей), $\log 1/LC_{50}$ – концентрации веществ,

вызывающие летальное воздействие на 50 % испытуемых рыб и *Daphnia Magna*, и $pIGC_{50}$ – концентрация соединений, подавляющая на 50 % размножение *Tetrahyumena Pyriformis* (growth inhibitory concentration values).

Расчет физико-химических дескрипторов осуществлялся с использованием программы НУВОТ [13].

Результаты и их обсуждение

Анализ данных по токсичности для всех инертных органических соединений по отношению к рассмотренным в данной работе биологическим объектам позволил отобрать 25 общих соединений. В таблице приведены данные по токсичности и физико-химическим параметрам этих соединений.

Межвидовые корреляции. Прежде всего мы провели линейный регрессионный анализ межвидовых корреляций. Для токсичности при внутривенном введении по отношению к мышам были получены следующие уравнения:

$$\log(1/LD_{50}^{mice-iv}) = 0,18(0,07) + 0,78(0,05)\log(1/pIGC_{50}^{TP}) \quad (1)$$
$$N = 19 \quad R^2 = 0,93 \quad Q^2 = 0,90 \quad SD = 0,20 \quad SD_{CV} = 0,23 \quad F = 226,0,$$

где N – число соединений, включенных в корреляцию; R – коэффициент корреляции, Q – коэффициент корреляции при процедуре cross validation, SD – стандартное отклонение, SDCV – стандартное отклонение при процедуре cross validation, F – критерий Фишера.

$$\log(1/LD_{50}^{mice-iv}) = 0,20(0,20) + 0,55(0,10)\log(1/LC_{50}^{DM}) \quad (2)$$
$$N = 12 \quad R^2 = 0,74 \quad Q^2 = 0,64 \quad SD = 0,39 \quad SD_{CV} = 0,48 \quad F = 29,7$$

$$\log(1/LD_{50}^{mice-iv}) = 1,75(0,24) + 0,68(0,02)\log(1/LC_{50}^{Guppy}) \quad (3)$$
$$N = 12 \quad R^2 = 0,92 \quad Q^2 = 0,89 \quad SD = 0,21 \quad SDCV = 0,26 \quad F = 121,8$$

$$\log(1/LD_{50}^{mice-iv}) = -0,25(0,09) + 0,50(0,05)\log(1/LC_{50}^{FHM}) \quad (4)$$
$$N = 14 \quad R^2 = 0,87 \quad Q^2 = 0,77 \quad SD = 0,27 \quad SDCV = 0,36 \quad F = 84,9$$

$$\log(1/LD_{50}^{mice-iv}) = -0,10(0,14) + 0,66(0,08)\log(1/LC_{50}^{RT}) \quad (5)$$
$$N = 5 \quad R^2 = 0,96 \quad Q^2 = 0,85 \quad SD = 0,23 \quad SDCV = 0,53 \quad F = 68,1$$

Статистические критерии всех вышеуказанных уравнений являются удовлетворительными, а стандартные отклонения – на уровне определения экспериментальной токсичности. По величинам коэффициентов уравнений очевидна близость токсичности по отношению к мышам и *Tetrahyumena Pyriformis*. И в случае других, изученных в данной работе объектов, наклон прямых различается незначительно (коэффициент B от 0,50 до 0,78), что также свидетельствует о схожести токсических процессов в данных организмах.

Токсичность и физико-химические параметры изученных соединений

N	Соединение	CAS	Токсичность							Физико-химические параметры (logP – экспериментальные, остальные рассчитаны программой HUBOT)	
			Mice lg(1/LD ₅₀)	Tetrahyemena Pyriformis lg(1/pIGC ₅₀)	Daphnia Magna lg(1/LC ₅₀)	Guppy lg(1/LC ₅₀)	Fathead Minnow lg(1/LC)	Rainbow Trout lg(1/LC ₅₀)	A	SumCa(O)	
1	Метанол	67-56-1	-2,17	-2,67	-2,82	-5,19	-2,96	-2,80	3,246	1,63	
2	Гептанол	53535-33-4	0,10	0,11	0,37		0,53		14,256	1,65	
3	Этанол	9003-99-0	-1,63	-1,99	-2,88	-4,90	-2,50	-2,35	5,081	1,66	
4	Пропан-2-ол	8013-70-5	-1,40	-1,88	-2,69	-4,83	-2,16		6,916	1,69	
5	2-Метилпропан-2-ол	75-65-0	-1,31	-1,79		-4,71	-1,94		8,751	1,6	
6	Бутан-2-ол	78-92-2	-1,01	-1,54			-1,69		8,751	1,62	
7	2-Метилбутан-2-ол	75-85-4	-0,84	-1,17					10,586	1,63	
8	2-Метилпропан-2-ол	78-83-1	-0,75	-1,37	-2,10	-3,50	-1,29	-1,25	8,751	1,57	
9	2-Метилпентан-2-ол	565-67-3	-0,50	-0,77					12,421	1,62	
10	Циклогексанол	108-93-0	-0,43	-0,77	-1,66	-3,51	-0,85		11,647	1,67	
11	3,3-Диметилбутан-1-ол	624-95-3	-0,38						12,421	1,63	
12	Пентан-1-ол	71-41-0	-0,32	-1,03	-1,27		-0,73	-0,66	10,586	1,61	
13	Октан-1-ол	72-69-5	0,28	0,58	0,70	-2,04	0,98	0,87	16,091	1,61	
14	Октан-2-ол	6169-06-8	0,30	0,0					16,091	1,6	

Избранные труды доктора химических наук, профессора О.А. Раевского
в области молекулярного компьютерного дизайна

Окончание таблицы

N	Соединение	CAS	Токсичность						Физико-химические параметры (logP – экспериментальные, остальные рассчитаны программой HUVOT)		
			Mice lg(1/LD ₅₀)	Tetrahyemena Pyriformis lg(1/pIC ₅₀)	Daphnia Magna lg(1/LC ₅₀)	Guppy lg(1/LC ₅₀)	Fathead Minnow lg(1/LC)	Rainbow Trout lg(1/LC ₅₀)	A	SumCa(O)	
15	Пропан-1-ол	71-23-8	-1,06	-1,75	-2,51			-1,88		6,916	1,61
16	Бутан-1-ол	82115-62-6	-0,71	-1,43	-2,04	-3,78		-1,37		8,751	1,6
17	3-Метилбутан-1-ол	6423-06-9	-0,42	-1,04	-1,62					10,586	1,62
18	Изоборнеол	507-70-0	0,44	0,15						18,213	1,6
19	Гексан-1-ол	29917-35-5	0,00	-0,38	-1,36	-2,80		0,02		12,421	1,59
20	1,2,4,5-Тетраметилбензол	95-93-2	-0,13							17,774	0,74
21	Пентан-3-он	96-22-0	-0,77	-1,46		-3,97				10,103	1,88
22	Ундекан-6-он	927-49-1	0,16							21,113	1,96
23	Нонан-5-он	502-56-7	0,01	0,07		-2,29				17,443	1,96
24	Диэтиловый эфир	75678-39-4	-1,13			-3,79		-1,54		8,751	1,54
25	Пропиловый эфир	111-43-3	-0,30							12,421	1,5

Следует отметить, что возможность теоретического обоснования модели является определяющей в выборе типа уравнений для оценки исследуемой активности. В литературе по КССА часто наряду с линейными рассматриваются и параболические модели [14]. Ниже приведены межвидовые параболические корреляции токсичности (уравнения 6–10). По сравнению с аналогичными линейными моделями (уравнения 1–5) статистические параметры существенно улучшаются. Это свидетельствует о нелинейной зависимости токсичностей химических соединений в разных организмах. Так, например, в случаях *Daphnia Magna* и Fathead Minnow имеется явная точка перегиба на графике «токсичность для мышей – токсичность для дафнии или пескаря».

$$\log(1/LD_{50}^{\text{mice-iv}}) = 0,15(0,05) + 0,45(0,09)\log(1/pIGC_{50}^{\text{TP}}) - 0,17(0,04)\log(1/pIGC_{50}^{\text{TP}})^2 \quad (6)$$

$$N = 19 \quad R^2 = 0,96 \quad Q^2 = 0,94 \quad SD = 0,14 \quad SD_{CV} = 0,18 \quad F = 221,8$$

$$\log(1/LD_{50}^{\text{mice-iv}}) = 0,28(0,11) - 0,005(0,129)\log(1/LC_{50}^{\text{DM}}) - 0,250(0,052)\log(1/LC_{50}^{\text{DM}})^2 \quad (7)$$

$$N = 12 \quad R^2 = 0,93 \quad Q^2 = 0,85 \quad SD = 0,22 \quad SD_{CV} = 0,31 \quad F = 58,6$$

$$\log(1/LD_{50}^{\text{mice-iv}}) = 0,70(0,81) + 0,06(0,46)\log(1/LC_{50}^{\text{Guppy}}) - 0,09(0,06)\log(1/LC_{50}^{\text{Guppy}})^2 \quad (8)$$

$$N = 12 \quad R^2 = 0,94 \quad Q^2 = 0,88 \quad SD = 0,20 \quad SD_{CV} = 0,27 \quad F = 66,9$$

$$\log(1/LD_{50}^{\text{mice-iv}}) = -0,06(0,04) + 0,33(0,03)\log(1/LC_{50}^{\text{FHM}}) - 0,135(0,016)\log(1/LC_{50}^{\text{FHM}})^2 \quad (9)$$

$$N = 14 \quad R^2 = 0,98 \quad Q^2 = 0,96 \quad SD = 0,11 \quad SD_{CV} = 0,16 \quad F = 294,9$$

$$\log(1/LD_{50}^{\text{mice-iv}}) = -0,002(0,027) + 0,43(0,03)\log(1/LC_{50}^{\text{RT}}) - 0,122(0,012)\log(1/LC_{50}^{\text{RT}})^2 \quad (10)$$

$$N = 5 \quad R^2 = 0,999 \quad Q^2 = 0,996 \quad SD = 0,04 \quad SD_{CV} = 0,09 \quad F = 1166,2$$

Таким образом, проведенный корреляционный анализ токсичности инертных химических соединений по отношению к разным живым организмам дает возможность рассчитывать токсический эффект на одном из изученных объектов по токсичности другого. Это открывает возможность достоверной оценки токсичности инертных соединений по отношению к теплокровным на основании данных о токсичности к более простым водным организмам, таким как инфузории, рачки и рыбки.

Корреляции с параметрами липофильности

Согласно [7] действие инертных токсикантов связано с их распределением внутри и вне мембранного пространства клеток. Традиционно этот процесс количественно моделируется коэффициентом распределения химических соединений в системе «октанол – вода» ($\log P$) [14]. В уравнениях 11–28 показано соотношение между изученными в данной работе видами токсичности соединений и их гидрофобностью. В данной работе во всех случаях использовались только экспериментальные значения $\log P$.

Линейные модели:

$$\log(1/pIGC_{50}^{TP}) = -1,97(0,07) + 0,76(0,04)\log P_{exp} \quad (11)$$

$N = 19$ $R^2 = 0,96$ $Q^2 = 0,95$ $SD = 0,18$ $SD_{CV} = 0,20$ $F = 442,8$

$$\log(1/LC_{50}^{DM}) = -2,68(0,14) + 0,95(0,09)\log P_{exp} \quad (12)$$

$N = 12$ $R^2 = 0,91$ $Q^2 = 0,85$ $SD = 0,36$ $SD_{CV} = 0,47$ $F = 104,2$

$$\log(1/LC_{50}^{Guppy}) = -4,63(0,08) + 0,86(0,06)\log P_{exp} \quad (13)$$

$N = 12$ $R^2 = 0,96$ $Q^2 = 0,95$ $SD = 0,22$ $SD_{CV} = 0,25$ $F = 228,9$

$$\log(1/LC_{50}^{FHM}) = -2,23(0,04) + 1,03(0,03)\log P_{exp} \quad (14)$$

$N = 14$ $R^2 = 0,99$ $Q^2 = 0,99$ $SD = 0,11$ $SD_{CV} = 0,13$ $F = 1274,4$

$$\log(1/LC_{50}^{RT}) = -2,07(0,04) + 0,95(0,02)\log P_{exp} \quad (15)$$

$N = 5$ $R^2 = 0,99$ $Q^2 = 0,99$ $SD = 0,07$ $SD_{CV} = 0,10$ $F = 1596,9$

$$\log(1/LD_{50}^{mice-iv}) = -1,25(0,10) + 0,46(0,05)\log P_{exp} \quad (16)$$

$N = 25$ $R^2 = 0,80$ $Q^2 = 0,73$ $SD = 0,30$ $SD_{CV} = 0,35$ $F = 89,4$

Сравнение статистических критериев уравнений 11–16 свидетельствует о том, что в случае изученных водных организмов существует устойчивая линейная зависимость токсичности с гидрофобностью, а в случае токсичности при внутривенном введении по отношению к мышам статистические критерии такой зависимости хуже. Это побудило нас протестировать и нелинейные зависимости токсичности и гидрофобности для изученных биологических объектов. Ниже приведены результаты корреляции параметров токсичности и гидрофобности в рамках билинейной и параболической моделей.

Параболические модели:

$$\log(1/pIGC_{50}^{TP}) = -1,97(0,06) + 0,76(0,10)\log PeXp - 0,0003(0,03)\log PeXp^2 \quad (17)$$

$N = 19$ $R^2 = 0,96$ $Q^2 = 0,94$ $SD = 0,18$ $SD_{CV} = 0,22$ $F = 208,4$

$$\log(1/LC_{50}^{DM}) = -2,65(0,08) + 0,45(0,12)\log P_{exp} + 0,21(0,05)\log P_{exp}^2 \quad (18)$$

N = 12 R² = 0,98 Q² = 0,96 SD = 0,20 SD_{CV} = 0,25 F = 176,7

$$\log(1/LC_{50}^{Guppy}) = -4,64(0,09) + 0,94(0,14)\log P_{exp} - 0,034(0,049)\log P_{exp}^2 \quad (19)$$

N = 12 R² = 0,96 Q² = 0,94 SD = 0,23 SD_{CV} = 0,29 F = 108,8

$$\log(1/LC_{50}^{FHM}) = -2,22(0,04) + 0,997(0,068)\log P_{exp} + 0,014(0,025)\log P_{exp}^2 \quad (20)$$

N = 14 R² = 0,99 Q² = 0,99 SD = 0,12 SD_{CV} = 0,14 F = 601,9

$$\log(1/LC_{50}^{RT}) = -2,06(0,05) + 0,96(0,06)\log P_{exp} - 0,004(0,025)\log P_{exp}^2 \quad (21)$$

N = 5 R² = 0,998 Q² = 0,961 SD = 0,09 SD_{CV} = 0,54 F = 539,2

$$\log(1/LD_{50}^{mice-iv}) = -1,46(0,06) + 0,95(0,07)\log P_{exp} - 0,14(0,02)\log P_{exp}^2 \quad (22)$$

N = 25 R² = 0,94 Q² = 0,92 SD = 0,17 SD_{CV} = 0,20 F = 165,7

Билинейные модели:

$$\log(1/pIGC_{50}^{TP}) = -1,96(0,06) + 0,74(0,05)\log P_{exp} - 0,122(0,190)\log(-0,0008P_{exp} + 1) \quad (23)$$

N = 19 R² = 0,96 Q² = 0,90 SD = 0,18 SD_{CV} = 0,30 F = 213,9

$$\log(1/LC_{50}^{DM}) = -3,29(0,19) - 0,50(0,41)\log P_{exp} + 1,759(0,485)\log(0,9992P_{exp} + 1) \quad (24)$$

N = 12 R² = 0,96 Q² = 0,94 SD = 0,24 SD_{CV} = 0,31 F = 122,0

$$\log(1/LC_{50}^{Guppy}) = -4,59(0,17) + 0,96(0,37)\log P_{exp} - 0,121(0,432)\log(0,9992P_{exp} + 1) \quad (25)$$

N = 12 R² = 0,96 Q² = 0,92 SD = 0,23 SD_{CV} = 0,32 F = 104,0

$$\log(1/LC_{50}^{FHM}) = -2,23(0,04) + 0,96(0,11)\log P_{exp} + 0,104(0,152)\log(0,27P_{exp} + 1) \quad (26)$$

N = 14 R² = 0,991 Q² = 0,988 SD = 0,12 SD_{CV} = 0,14 F = 609,3

$$\log(1/LC_{50}^{RT}) = -2,01(0,12) + 1,05(0,20)\log P_{exp} - 0,116(0,237)\log(1,17P_{exp} + 1) \quad (27)$$

N = 5 R² = 0,998 Q² = 0,981 SD = 0,08 SD_{CV} = 0,29 F = 596,5

$$\log(1/LD_{50}^{mice-iv}) = -1,47(0,06) + 0,78(0,05)\log P_{exp} - 1,051(0,132)\log(0,133P_{exp} + 1) \quad (28)$$

N = 25 R² = 0,95 Q² = 0,93 SD = 0,16 SD_{CV} = 0,19 F = 196,1

Сравнение уравнений линейной, параболической и билинейной зависимости токсичности и гидрофобности показывает, что для рыбок (*Guppy*, *Fathead Minnow* и *Rainbow Trout*) использование более сложных параболических и билинейных моделей не приводит к улучшению значений коэффициентов корреляций и стандартных отклонений. А критерий Фишера значительно понизился, очевидно, за счет увеличения числа использованных переменных. В случае же *Daphnia Magna* и мышей использование нелинейных моделей приводит к существенному улучшению статистических критериев соответствующих уравнений. Таким образом, для изученных водных

организмов линейные зависимости токсичности от гидрофобности являются стабильными и дают возможность оценивать токсичность инертных соединений по величине $\log P$. В случае же *Daphnia Magna* целесообразно использовать параболическую модель, а для оценки токсичности для мыши билинейную.

В заключении этого раздела необходимо отметить, что использование экспериментальных значений $\log P$ сопряжено с трудностями. Во-первых, не для всех веществ имеются в литературе такие данные. Во-вторых, экспериментальное определение значений $\log P$ требует синтеза соединений и привлечения квалифицированных специалистов для оценки величин $\log P$. Такая работа сопряжена с существенными финансовыми затратами. По этой причине целесообразно для расчета токсичности химических соединений привлечение не экспериментально определяемых параметров (токсичность по другому организму, липофильность), а дескрипторов, рассчитываемых какой-либо программой.

Корреляции токсичности с физико-химическими параметрами

В работе [15] показано, что липофильность органических соединений количественно описывается с помощью двух дескрипторов программы НУВОТ [13]: поляризуемость соединений (α) и суммой водородно связанных акцепторных факторов молекул SumCa(O). Включение указанных дескрипторов в корреляции их с токсичностью химических соединений на изученных объектах приводит к следующим соотношениям:

$$\log(1/pIGC_{50}^{TP}) = -1,76(0,93) + 0,21(0,01)\alpha - 0,93(0,57)\text{SumCa(O)} \quad (29)$$

$N = 19$ $R^2 = 0,94$ $Q^2 = 0,91$ $SD = 0,24$ $SD_{CV} = 0,28$ $F = 119,2$

$$\log(1/LC_{50}^{DM}) = -4,43(0,37) + 0,289(0,036)\alpha \quad (30)$$

$N = 12$ $R^2 = 0,87$ $Q^2 = 0,77$ $SD = 0,44$ $SD_{CV} = 0,59$ $F = 65,2$

$$\log(1/LC_{50}^{Guppy}) = -3,45(1,35) + 0,26(0,03)\alpha - 1,70(0,86)\text{SumCa(O)} \quad (31)$$

$N = 12$ $R^2 = 0,92$ $Q^2 = 0,88$ $SD = 0,33$ $SD_{CV} = 0,38$ $F = 48,9$

$$\log(1/LC_{50}^{FHM}) = -4,24(0,20) + 0,321(0,020)\alpha \quad (32)$$

$N = 14$ $R^2 = 0,96$ $Q^2 = 0,94$ $SD = 0,24$ $SD_{CV} = 0,28$ $F = 266,3$

$$\log(1/LC_{50}^{RT}) = -3,77(0,04) + 0,289(0,004)\alpha \quad (33)$$

$N = 5$ $R^2 = 0,999$ $Q^2 = 0,998$ $SD = 0,04$ $SD_{CV} = 0,07$ $F = 4177,4$

$$\log(1/LD_{50}^{\text{mice-iv}}) = -2,14(0,16) + 0,137(0,013)\alpha \quad (34)$$

$N = 25$ $R^2 = 0,82$ $Q^2 = 0,77$ $SD = 0,28$ $SD_{CV} = 0,52$ $F = 107,5$

Сопоставление уравнений 29–34 с соответствующими уравнениями с параметром липофильности показывает, что хотя статистика первых несколько хуже, их использование вполне приемлемо для оценки токсичности инертных химических соединений по отношению ко всем изученным биологическим объектам. Существенным же преимуществом использования указанных теоретических рассчитываемых дескрипторов следует считать возможность интерпретации процесса формирования токсичности. Из уравнений 29 и 34 очевиден положительный вклад в токсичность поляризуемости и отрицательный вклад водородносвязанных акцепторных факторов молекулы.

ЛИТЕРАТУРА

1. EC (2006). Regulation (EC) No. 1907 / 2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999 / 45 / EC and repealing Council Regulation (EEC) No. 793 / 93 and Commission Regulation (EC) No. 1488 / 94 as well as Council Directive 76 / 769 / EEC and Commission Directives 91/155/EEC, 93 / 67 / EEC, 93 / 105 / EC and 2000 / 21 / EC. Official Journal of the European Union L 396 / 1 of 30.12.2006. Office for Official Publications of the European Communities (ОПОСЕ), Luxembourg.
2. ECHA (2008). Guidance on Information Requirements and Chemical Safety Assessment. Chapter R6. European Chemicals Agency, Helsinki, Finland. Available from: http://guidance.echa.europa.eu/docs/guidance_document/information_requirements_en.htm?time=1252064523#r6.
3. *Tsakovska I., Lessigiarska I., Netzeva T. and Worth A.* QSAR & Combi. Sci. 2008. Vol. 27. P. 41–48.
4. *Jackel H. and Klein W.* Quantitative Structure-Activity Relat. 1991. Vol. 10. P. 198–204.
5. *Juranic I.O., Drakulic B.J. et al.* Chemosphere. 2006. Vol. 62(4). P. 641–649.
6. *Devillers J. and Devillers H.* SAR & QSAR in Environmental. Res. 2009. Vol. 20. P. 467–500.
7. *Verhaar H.J.M., Van Leeuwen C.J., Hermens J.L.M.* Chemosphere. 1992. Vol. 25. P. 471–491.
8. SYMYX Toxicity Database. <http://www.symyx.com/products/databases/index.jsp>.
9. *Raevsky O.A., Grigor'ev V.Yu., Weber E.E. and Dearden J.C.* QSAR & Combinatorial Sci. 2008. Vol. 27(11–12). P. 1274–1281.
10. *Раевский О.А., Раздольский А.Н., Тонкопий В.Д. и др.* Хим.-фарм. журн. 2008. Vol. 42(6). С. 22–27.

11. *Zhu H., Tropsha A., Fourches D.et. al.* // J. Chem. Inf. Model. 2008. No. 48. P. 766–784.

12. *Ellison C.M., Cronin M.T.D., Madden J. C., Schultz T. W.* SAR and QSAR in Envir. Res. 2008. Vol. 19(7). P. 751–783.

13. HYBOT (Hydrogen Bond Thermodynamics) program package, O.A. Raevsky, V.Ju. Grigor'ev, S.V. Trepalin, Registration by Russian State Patent Agency № 990090 of 26.02.99.

14. *Стьюпер Э., Брюггер У., Джурс П.* Машинный анализ химической структуры и биологической активности, Мир, Москва. 1982.

15. *Raevsky O.A., Schaper K.-J., H.van de Waterbeemd, McFarland J.W.* in: Molecular Modelling and Prediction of Bioactivity, K. Gundertofe and F. Jorgensen (eds.), Kluwer Academic / Plenum Publ. 2000, pp. 221–227.

О.А. Раевский¹, С.Л. Солодова¹, А.А. Лагунин², В.В. Поройков²

КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ ПРОНИЦАЕМОСТИ ФИЗИОЛОГИЧЕСКИ АКТИВНЫХ ВЕЩЕСТВ ЧЕРЕЗ ГЕМАТОЭНЦЕФАЛИЧЕСКИЙ БАРЬЕР

*¹Институт физиологически активных веществ РАН,
Черноголовка, Московская область*

*²Научно-исследовательский институт биомедицинской химии
имени В.Н. Ореховича, Москва*

1. Природа гематоэнцефалического барьера (ГЭБ) (строение, функции)

Нарушения деятельности центральной нервной системы (ЦНС) во многих случаях являются результатом влияния на организм различных эндо- и экзогенных факторов. Для обеспечения нормального функционирования ЦНС организмом в процессе эволюции был выработан ряд защитных механизмов, которые обеспечивают надежность функционирования ЦНС даже в агрессивных условиях. Одним из важнейших защитных механизмов является гематоэнцефалический барьер (ГЭБ). Под этим термином условно объединяют совокупность анатомических и функциональных систем, которые поддерживают постоянство и регулируют состав внутренней среды головного и спинного мозга [1, 2].

ГЭБ регулирует распределение различных веществ между током крови и мозгом и обладает высокой избирательностью. Барьерная функция состоит в предотвращении проникновения вредных веществ из крови в мозг и их активном выведении в тех случаях, если они каким-то образом попали в ЦНС. ГЭБ играет ведущую роль в возникновении таких патологий, как инсульты, микробные инфекции, черепно-мозговые травмы, нейродегенеративные заболевания, которые, в свою очередь, могут приводить к изменениям проницаемости ГЭБ. С другой стороны, ГЭБ ограничивает проницаемость в ЦНС многих лекарственных препаратов, что может служить препятствием при терапии ряда заболеваний ЦНС [2].

Основной функцией ГЭБ является поддержание гомеостаза мозга [3]. ГЭБ препятствует попаданию в центральную нервную систему циркулирующих в крови микроорганизмов, токсинов, клеточных и гуморальных факторов иммунной системы, которые воспринимают ткань мозга как чужеродную и могут вызвать ее необратимые повреждения [4]. ГЭБ выполняет функцию высокоселективного фильтра, через который из кровеносного русла в мозг поступают питательные вещества, а в обратном направлении выводятся продукты жизнедеятельности нейронов [3]. С помощью ГЭБ осуществляется избирательная проницаемость в головной мозг нейромедиаторов и нейропептидов центрального действия, таким образом, не происходит избыточного воздействия на нейроны головного мозга нейромедиаторов периферического действия. Кроме того, одной из важнейших функций ГЭБ является сохранение баланса концентрации ионов и необходимого объема жидкости в головном мозге для создания оптимальных условий функционирования нейронов [3].

Механизмы проникновения через ГЭБ как эндогенных, так и экзогенных веществ обусловлены его морфологическими и биохимическими особенностями, поэтому при исследовании механизмов проницаемости через ГЭБ в первую очередь должно быть рассмотрено его строение.

Принято условно выделять два компонента, принимающих участие в образовании морфологической структуры ГЭБ. Первый, физический, или структурный компонент, образован клетками эндотелия (эндотелиоцитами), выстилающими стенки микрокапилляров мозга. Эндотелиоциты образуют плотные контакты, которые препятствуют проникновению инородных веществ в ЦНС. В капиллярах мозга плотность межклеточных контактов значительно превышает плотность таких контактов в других тканях [5, 6]. Плотные контакты между эндотелиоцитами значительно ограничивают проницаемость через ГЭБ ионов Na^+ , Cl^- , в результате чего создается высокое трансэндотелиальное электрическое сопротивление ($> 1000 \text{ Ом/см}^2$).

Среди белков, которые играют роль в формировании плотных контактов, наиболее важными являются окклюдин и клаудины. Окклюдин — белок с молекулярной массой 60–65 кДа, который взаимодействует с белком 1 «зоны замыкания» (ZO-1) на мембране эндотелиоцита. Основная функция окклюдина — регуляция образования плотных контактов между эндотелиоцитами [7, 8]. Экспрессия клаудинов-3, -5 и -12, по-видимому, связана с поддержанием высокого трансэндотелиального электрического сопротивления [9]. Белки адгезии JAM-A, JAM-B и JAM-C экспрессируются в эндотелии головного мозга (ГМ) и осуществляют функцию регуляции образования и сохранения плотных контактов. Трансмембранные белки взаимодействуют в цитоплазме с белками, расположенными на периферии мембраны эндотелиоцита с образованием больших белковых комплексов. Эти комплексы в дальнейшем участвуют в передаче сигнала и регуляции образования и/или восстановления плотных контактов. Среди трансмембранных белков, спо-

собных к образованию таких комплексов, наиболее изучены белки «зоны замыкания» – 1, 2, 3 (ZO-1, 2, 3), кальций-зависимая сериновая протеинкиназа (CASK), мембраносвязанные гуанилаткиназы – 1, 2 и 3 (MAGI-1, 2, 3). Известна роль ZO-1 связанного белка, взаимодействующего с нуклеиновыми кислотами (ZONAB) в регуляции транскрипции белков зоны замыкания. К настоящему времени не выявлены механизмы, способствующие увеличению проницаемости низкомолекулярных веществ через ГЭБ, обусловленные взаимодействием с белками, образующими плотные контакты.

2. Основные механизмы транспорта через ГЭБ

Как уже было сказано выше, ГЭБ не только ограничивает проницаемость различных веществ из крови в мозг, но и выполняет противоположную функцию – транспортирует необходимые вещества для метаболизма тканью мозга.

В настоящее время известно, что существует множество механизмов проникновения веществ из кровеносного русла в мозг. Сюда входят различные транспортные системы, каналы клеточной мембраны, везикулярный транспорт. Однако большинство веществ проходит через ГЭБ посредством диффузии [10, 11]. Диффузия может осуществляться через клеточные мембраны эндотелиоцитов и, в меньшей степени, через плотные межклеточные контакты. Пассивная диффузия осуществляется по градиенту концентрации из тока крови в мозг без энергетических затрат.

Второй, биохимический, компонент образован транспортными системами, которые расположены на люминальной (обращенной в просвет сосуда) и аблюминальной (внутренней или базальной) мембране эндотелиоцита. Транспортные системы могут осуществлять как перенос веществ из кровотока к мозгу (influx), так и/или обратный перенос из ткани мозга в кровоток (efflux) [5]. Работа белков-транспортёров требует затрат энергии, так как осуществляется против градиента концентрации (активный транспорт). Энергия для переноса высвобождается либо при гидролизе аденозинтрифосфата (т. н. АТФ-связывающие транспортёры, АТФ cassette binding carriers, ABC), либо при работе Ca^{2+} - или Na^+/K^+ -зависимых АТФаз, использующих градиент концентрации ионов для получения энергии (т. н. вторичные транспортёры, solute-carrier transporters, SLC). Na^+/K^+ -АТФ-азы и белки-переносчики, использующие их энергию, экспрессируются на аблюминальной мембране эндотелиоцита, в то время как Ca^{2+} -зависимые АТФазы экспрессируются на люминальной и аблюминальной мембранах [12, 13]. Следует отметить, что для некоторых SLC-транспортёров показана возможность осуществлять транспортировку независимо от градиента концентрации ионов, по механизму облегченной диффузии [14, 15].

SLC-транспортёры включают в себя ряд переносчиков, которые осуществляют доставку в ткань мозга питательных веществ – глюкозы, аминокислот, нуклеиновых кислот. SLC-транспортёры могут участвовать в переносе низкомолекулярных веществ, подобных по структуре эндогенным молекулам, из кровотока в ткань мозга [10, 14].

Нейтральные аминокислоты, такие как фенилаланин, тирозин и лейцин, транспортируются через ГЭБ посредством так называемой системы L, работающей независимо от градиента концентрации ионов и включающей в себя белки, кодируемые генами LAT-1 [16, 17] и LAT-2 [18–20]. Предполагается, что посредством транспортёров системы L через ГЭБ транспортируются лекарственные препараты габапентин и баклофен.

Белки-транспортёры глюкозы могут играть важную роль в транспортировке лекарственных препаратов через ГЭБ, поскольку транспортная система, осуществляющая перенос глюкозы, обеспечивает более высокую скорость по сравнению с другими транспортными системами [14]. Два основных семейства белков, осуществляющие перенос глюкозы через ГЭБ, – это: (1) семейство SGLT, использующее энергию Na^+/K^+ -зависимой АТ-Разы, (2) семейство GLUT, не использующее градиент концентрации ионов. Предполагается, что основным транспортёром глюкозы, экспрессируемым в эндотелии капилляров головного мозга (ГМ), является GLUT1. Возможно, посредством GLUT1 через ГЭБ транспортируются низкомолекулярные сахара. [21]. Однако возможности переноса низкомолекулярных веществ через ГЭБ с помощью GLUT1-транспортёра требуют дополнительного изучения, поскольку вследствие высокой потребности ГМ в глюкозе использование транспортёров глюкозы для переноса экзогенных веществ может быть связано с серьёзными побочными эффектами.

Транспортные системы, осуществляющие перенос через ГЭБ монокарбоксилатов (МСТ), включают в себя, главным образом, транспортёры, осуществляющие перенос метаболитов, таких как молочная кислота, и других из ткани ГМ в кровоток и для переноса пирувата и монокарбоксилатов в головной мозг [22–24].

Было показано, что многие низкомолекулярные вещества, в частности, пропионат, бутират, бензойная кислота, салициловая кислота, никотиновая кислота, некоторые β -лактамные антибиотики, также транспортируются через ГЭБ посредством МСТ-системы [25]. Один из восьми белков семейства МСТ экспрессируется на люминальной и аблюминальной мембранах эндотелиоцитов ГМ, мышцах, тонком кишечнике. Существуют предположения о том, что транспортёры МСТ работают по механизму «антипорта» и осуществляют обмен отрицательно заряженных молекул [14].

Органические заряженные вещества транспортируются через ГЭБ посредством транспортёров семейства OAT (organic anion transporters) и OCNT (organic carnitine/cation transporters). Транспортёры семейства OAT, OAT-1 и OAT-3 экспрессируются на люминальной и аблюминальной мем-

бране капилляров эндотелиоцитов и контролируют перенос через ГЭБ таких молекул, как холаты, 17- β эстрадиол-глюкуронид, сердечные гликозиды (дигоксин) [26]. Существуют данные о том, что токсичность дигоксина и его побочные эффекты на ЦНС возникают при накоплении этого препарата в ЦНС, ключевую роль в этом процессе играют транспортеры OAT [14, 26].

Известно, что транспортеры семейства OCNT принимают участие в транспортировке карнитина [27] и тетраэтиламмония [28]. OCNT-транспортеры осуществляют рН-зависимый, ион-независимый перенос через ГЭБ органических катионов. Предполагается, что эти транспортеры играют роль в переносе через ГЭБ антагонистов H_1 -рецептора гистамина, таких как метирамин, фексофенадин [28].

В ГМ также существуют специализированные переносчики небольших пептидов, такие как PerT1, PerT2, NPT, PTS 1-5, которые принимают участие в переносе через ГЭБ энкефалинов, тиреотропин-рилизинг гормона (ТТГ), α -меланоцит-стимулирующего гормона и т. п. [29, 30].

Транспортная система ГЭБ представлена также нуклеозидными транспортерами двух типов: (1) равновесные транспортеры нуклеозидов (equilibrative nucleoside transporters, SLC29), которые в зависимости от градиента концентрации вводят или выводят нуклеозиды из клетки и натрий-зависимые нуклеозидные транспортеры, работающие против градиента концентрации (concentrative nucleoside transporters, SLC28) [31]. Известно несколько лекарственных препаратов, субстратов нуклеозидных транспортеров. Например, противоопухолевый препарат гемцитабин и противовирусный препарат зидовудин (AZT) являются субстратами натрий-зависимых нуклеозидных транспортеров [32, 33].

Следует отметить, что для некоторых лекарственных препаратов известно, что они поступают в ткань ГМ путем активного транспорта при участии белков-переносчиков, однако не исследованы системы, которые осуществляют перенос этих веществ через ГЭБ. Примерами таких препаратов являются фентанил, пентазоцин [14]. Кроме того, экспериментальные исследования транспортных систем ГЭБ продолжаются, что может привести в дальнейшем к открытию новых белков-переносчиков и новых механизмов доставки лекарственных препаратов в ЦНС.

При разработке новых лекарственных препаратов, взаимодействующих с белками-переносчиками эндогенных субстанций, в течение некоторого времени использовали стратегию «тройного коня», при этом вещество с низкой проницаемостью через ГЭБ связывали с эндогенными веществами для того, чтобы обеспечить взаимодействие полученного комплекса с белком-транспортером и его перенос из кровотока в ЦНС. Однако у такой стратегии есть недостатки. Во-первых, очень часто комплекс эндогенного и экзогенного вещества не распознается белками-транспортерами, либо комплекс после связывания с белком-транспортером распознается как чужеродный и разрушается в лизосомах эндотелиоцитов. Во-вторых, комплекс

должен распадаться после переноса в ЦНС, но не до этого – выполнения данного условия достаточно трудно достигнуть в живом организме из-за особенностей фармакокинетики эндогенных веществ [11].

Сложность исследования *in silico* взаимодействия с белками-транспортерами, ответственными за транспорт веществ из крови, обусловлена тем, что в литературе нет достаточного количества данных о взаимодействии низкомолекулярных веществ с конкретным белком-транспортером, что значительно затрудняет построение QSAR/QSPR (QSAR/QSPR (Quantitative Structure-Activity/Property Relationships) модели – модели количественных взаимосвязей «структура – активность» / «структура – свойство» – моделей в отношении взаимодействия с белками-транспортерами.

Семейство ABC-транспортеров включает в себя Р-гликопротеин (Pgp), обеспечивающий выведение (efflux) лекарств, а также белки, обеспечивающие устойчивость к лекарственным препаратам (Multidrug Resistance Protein Family, MRP).

Взаимодействие низкомолекулярных веществ с белками системы ABC, ответственными за выброс веществ, обуславливает возникновение множественной устойчивости к лекарственным препаратам – MDR (Multidrug Resistance). Эта устойчивость обусловлена чрезвычайно развитыми системами активного выброса веществ из ЦНС в кровоток, в частности Р-гликопротеином (Pgp, ABCB1). Кроме Р-гликопротеина, этот процесс осуществляет семейство белков, обуславливающих множественную лекарственную устойчивость (MRP, ABCC). Все белки семейства MRP содержат трансмембранные домены (TMD), состоящие из шести α -спиралей (так называемый Р-гликопротеин-подобный домен) и цитоплазматический домен L_0 , который выполняет функцию транспорта (Рисунок 1) [34, 35].

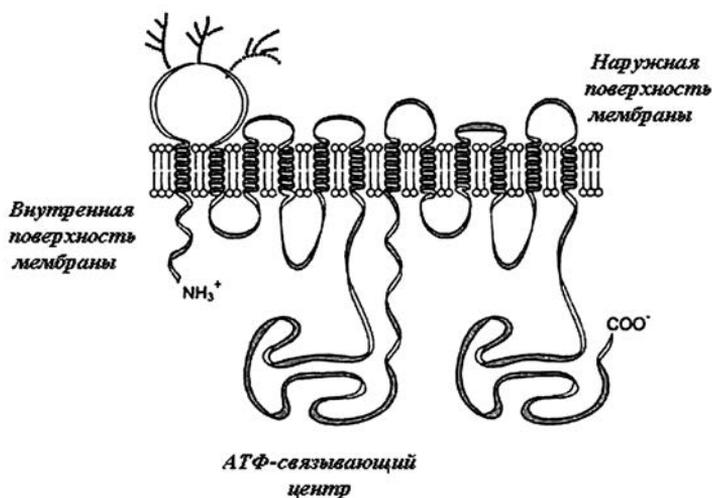


Рис. 1. Топологическая модель строения MRP [36]

Идентифицировано около девяти белков этого семейства. MRP-1 является белком, экспрессируемым во многих тканях, в том числе и в головном мозге [37]. Наиболее важна роль MRP-1 в транспорте глутатиона, глюкуроноидов, конъюгированных лейкотриенов. Известно, что глутатион увеличивает транспорт субстратов MRP-1 [38]. Предполагают, что сайт связывания MRP содержит два «кармана», с одним карманом могут взаимодействовать комплексы глутатиона и низкомолекулярных соединений, а с другим – отдельные молекулы. Глутатион транспортируется вне комплекса с низкомолекулярным соединением в очень низких концентрациях, следовательно, его основная роль состоит в поддержании транспорта катионов и нейтральных молекул. Анионы переносятся MPR вне комплекса с глутатионом. Участие семейства белков MRP в эффлюксе отдельных лекарственных препаратов находится на стадии изучения.

P-гликопротеин является наиболее распространенным эффлюкс-транспортером, входящим в состав люминальных мембран эндотелиоцитов.

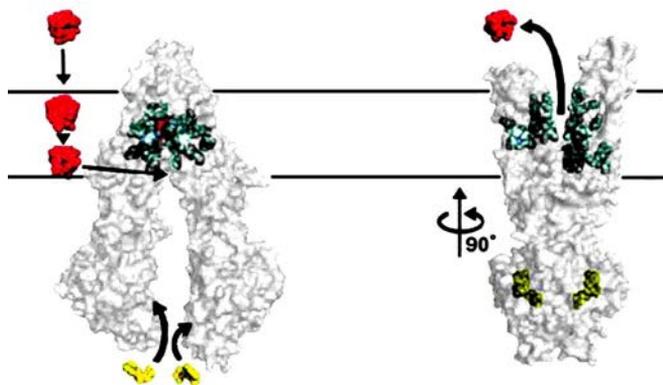


Рис. 2. Модель транспорта субстрата с помощью P-гликопротеина

Он осуществляет выброс из клеток молекул с молекулярной массой от 300 до 4000 Да в кровоток [39, 40] (рисунок 2). Таким образом, именно P-гликопротеин является наиболее важным транспортером, препятствующим достижению терапевтических концентраций ряда физиологически активных веществ в паренхиме головного мозга [41]. В результате недавних исследований также было показано, что P-гликопротеин экспрессируется не только на люминальной мембране, но и на внутриклеточных мембранах (эндоплазматический ретикулум, мембрана ядра) [42]. Взаимодействие низкомолекулярных соединений с P-гликопротеином происходит по весьма сложному механизму, который обусловлен тем, что этот белок содержит несколько сайтов связывания, а также тем, что в результате связывания P-гликопротеина с АТФазой происходит изменение его конформации, что также влияет на связывание белка с субстратом [43].

Таким образом, блокирование транспорта субстрата Р-гликопротеина другим субстратом не всегда предполагает наличие обратного эффекта. Например, циклоспорин ингибирует транспорт дигоксина в эксперименте на клеточных линиях, но дигоксин не влияет на транспорт циклоспорина [44]. Стимуляция функции Р-гликопротеина, приводящая к увеличению активности АТФазы и усиленному транспорту молекул, является одним из аспектов проявления множественной устойчивости к лекарственным препаратам. Например, известно, что прогестерон, пропранолол, дилтиазем и amitриптилин, являясь субстратами Р-гликопротеина, стимулируют транспорт верапамила [44]. Низкомолекулярные соединения могут быть как конкурентными, так и неконкурентными ингибиторами Р-гликопротеина; также возможно одновременное взаимодействие нескольких соединений, приводящее к усилению эффекта [44]. Изучение взаимосвязи между структурой химических соединений и их взаимодействием с Р-гликопротеином осложняется тем, что до сих пор не установлена с высоким разрешением его пространственная структура. В работе [45] была сопоставлена

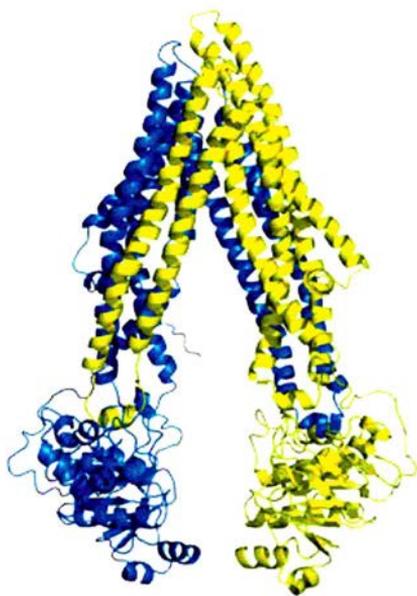


Рис. 3. Структура Р-гликопротеина

трехмерная структура Р-гликопротеина (рисунок 3), определенная методом электронной кристаллографии с разрешением ~ 20 Å, для свободного белка и его комплекса с негидролизуемым аналогом АТФ (AMP-PNP). Показано, что при связывании нуклеотида возникают значительные конформационные перестройки трансмембранных доменов, распространяющиеся на всю глубину мембраны. Недавно также была расшифрована с разрешением ~ 8 Å пространственная структура Р-гликопротеина в комплексе с AMP-PNP [46], а также структура Р-гликопротеина мыши – с разрешением около 4 Å (PDB IDs: 3G61, 3G60) [47].

Изучение зависимости между структурой химического соединения и его способностью к взаимодействию с Р-гликопротеином необходимо для оценки проницаемости соединения через ГЭБ. Одним из критериев для оценки взаимодействия вещества с Р-гликопротеином, который применялся при построении QSPR-моделей *in silico* [48], является способность вещества к образованию водородных связей. Липофильность является дополнительным фактором, обуславливающим активность субстрата при взаимодействии с Pgp [48, 49].

Еще одним механизмом транспорта через ГЭБ является рецептор-опосредованный транцитоз [50]. Несмотря на то, что в литературе есть данные о том, что некоторые лекарственные препараты могут проникать через ГЭБ посредством рецептор-опосредованного транцитоза [10], тем не менее, наиболее распространенными механизмами проницаемости низкомолекулярных веществ через ГЭБ считаются трансмембранная диффузия и взаимодействие с белками-транспортёрами [10, 11, 14].

3. Исследование по установлению взаимосвязей между структурой лигандов и их способностью преодолевать ГЭБ по механизму пассивного транспорта

Каждая биологическая мембрана обладает своими собственными биохимическими и физико-химическими характеристиками. Поэтому следует ожидать влияния физико-химических характеристик физиологически активных веществ на прохождение через биологический барьер. Целый ряд экспериментальных и теоретических подходов, связанных с измерением и/или с расчетом проницаемости через ГЭБ, был реализован в девяностые годы XX века [51–63]. В частности было продемонстрировано, что на основе параметра $\Delta \log P$, определяемого как разность между значениями липофильности $\log P_{\text{октанол-вода}}$ и $\log P_{\text{октан-вода}}$, можно предсказать проницаемость веществ в мозг [56, 58, 64, 65]. Были сформулированы правила, согласно которым можно ожидать оптимальной проницаемости через ГЭБ для соединений с характеристиками $0 < \log D_{\text{октанол}} < 3$ и $\Delta \log P < 2$ [58], где $\log D$ – коэффициент распределения октанол/вода при фиксированном значении pH. Продемонстрировано, что эти значения $\Delta \log P$ определяют в неявной форме главным образом способность к образованию водородной связи физиологически активных веществ с компонентами мембраны [56, 58, 59]. Тем самым предполагается, что образование межмолекулярных водородных связей играет важную роль в транспорте веществ через ГЭБ.

Было показано [66–69], что $\log P$ или $\log D$ являются составными параметрами, зависящими от двух характеристик, а именно – размера (V) и полярности (Λ), что можно выразить уравнением (1):

$$\log P = a \cdot V - \Lambda, \quad (1)$$

где V молярный объем, Λ – полярность и a – коэффициент регрессии.

Мембранная проницаемость может быть записана в виде уравнения [70] как

$$P_{erm} = P_m \cdot \bar{D}/h, \quad (2)$$

где P_{erm} – коэффициент проницаемости ($\text{см} \cdot \text{с}^{-1}$), \bar{D} – коэффициент диффузии в мембране ($\text{см}^2 \cdot \text{с}^{-1}$), P_m – коэффициент распределения между мембранной

тканью и водным окружением, h – толщина мембраны (см). Относительная липофильность растворенного вещества часто выражается через величину его коэффициента распределения в системе «октанол – вода», которая, как предполагают, линейно связана с коэффициентом распределения между мембранной и водной фазами в виде уравнения Колландера [71]:

$$\log P_m = a \cdot \log P_{\text{oct}} + b. \quad (3)$$

Однако в некоторых исследованиях было продемонстрировано, что это соотношение справедливо только тогда, когда свойства мембраны и октанола не слишком различаются между собой. Поэтому в работах [72, 73] предложено модифицировать уравнение Колландера следующим образом:

$$\log P_m = a \cdot \log P_{\text{oct}} + b \cdot \text{НВ} + c,$$

где НВ – это разница в способности образования водородной связи между двумя средами. (4)

Диффузия через мембрану связана с размером растворенного вещества. Для описания этой зависимости предложено несколько уравнений, например, [74]:

$$\bar{D} \sim e^{-0,005V} \cdot V^{0,8}, \quad (5)$$

где V – молярный объем растворенного вещества, или в логарифмической форме уравнение:

$$\log \bar{D} \sim 0,8 \log V \sim a \log MW, \quad (6)$$

где MW – молекулярный вес растворенного вещества. Таким образом, уравнение (2) может быть представлено как:

$$\log P_{\text{cm}} = a \cdot \log P_{\text{oct}} + b \cdot \text{НВ} + a \log MW + c. \quad (7)$$

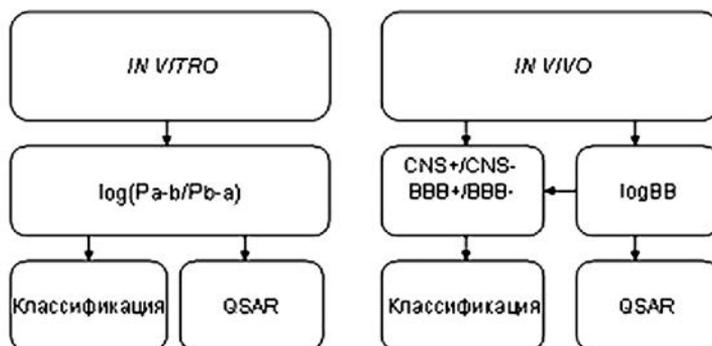
Для большинства лекарств $MW = 200 - 600$ D, и, следовательно, $\log MW$ является почти константой. Использование уравнения (1) позволяет получить уравнение:

$$\log P_{\text{cm}} = p \cdot MW + q \cdot \text{НВ} + r, \quad (8)$$

где p , q и r – коэффициенты множественной регрессии. Это уравнение связывает мембранную проницаемость, включая проницаемость через ГЭБ, с основными факторами, которые вносят вклад в этот процесс, а именно – с размером молекулы и способностью растворенного вещества к образованию водородной связи.

Указанные выше параметры используются практически во всех работах, посвященных установлению связи «структура – свойство» и опубликованных в девяностые годы XX века [63, 64, 75–77]. При этом рассматриваемые выборки содержали до нескольких десятков соединений, достаточно близких по структуре. В результате построенные в разных работах QSAR-модели включали в качестве зависимых переменных разные дескрипторы: например, в одной модели предпочтение отдавалось дескрипторам, связанным с размером или объемом молекулы, а в другой – дескрипторам, отражающим способность вещества к образованию водородной связи [78]. Обзор этих работ приведен в статье [78].

В 2000-х годах появились публикации, основанные на анализе обучающих выборок существенно большего размера, с использованием различных параметров для описания свойств молекулы и разных методов построения моделей QSAR [79–83]. Схематически эти подходы можно представить следующим образом:



Используются как регрессионные (QSAR), так и классификационные модели; обучающие выборки содержат данные о проницаемости веществ через ГЭБ, полученные *in vitro* ($\log(P_{a-b}/P_{b-a})$) или *in vivo* ($\log BB$ для QSAR-моделей и о наличии/отсутствии проницаемости через ГЭБ (BBB+/BBB–) для качественных классификационных моделей), либо о наличии/отсутствии действия веществ на ЦНС (ЦНС+/ЦНС–).

Обзор работ по компьютерному моделированию проницаемости веществ через гематоэнцефалический барьер приведен в публикациях [78,84–88]; в таблице 1 мы представили работы, опубликованные за последние пять лет.

4. Классификационные модели проницаемости веществ через ГЭБ

В литературе используется два различных параметра для классификации проницаемости веществ через ГЭБ: наличие/отсутствие действия вещества на ЦНС (ЦНС+/ЦНС–) либо превышение веществом некото-

рого порогового значения $\log BB$ ($BBB+/BBB-$) [81, 87, 88]. Нужно подчеркнуть, что эти параметры не обязательно коррелируют друг с другом, поскольку не все вещества, проникающие через ГЭБ, воздействуют на биологические мишени в ЦНС.

При классификации $BBB+/BBB-$ существенным является выбор порогового значения $\log BB$, который осуществляется различными авторами произвольно (в опубликованных работах эта величина находится в интервале от -1 до $0,63$). Понятно, что сравнение результатов классификации, полученной в различных работах [84], возможно только при одинаковом пороговом значении.

С нашей точки зрения, наиболее интересные результаты по классификации ЦНС+/- получены в работах [70, 87, 122, 123]. Рассматриваемые ряды соединений содержали от 125 соединений [70] до многих тысяч из коммерчески доступных баз и собственной базы компании Pfizer [123]. Во всех четырех публикациях со всей определенностью делается вывод о решающей роли в проницаемости веществ через ГЭБ способности веществ к образованию водородных связей. Вместе с тем имеются и некоторые различия. Так, в публикациях [70] и [87] наряду с дескрипторами, связанными с образованием водородной связи, используются также дескрипторы, характеризующие объем молекулы, и структурные фрагменты [87]. В то же время в нашей работе [122] показано, что использование только одного дескриптора, а именно полного водородносвязанного потенциала (ΣC_{ad}) позволяет произвести классификацию ЦНС+/- с точностью 97%. А в работе [123] полный водородносвязанный потенциал был использован как «глобальная мера проницаемости» и было установлено, что новые лекарства обладают меньшей проницаемостью через ГЭБ по сравнению с препаратами, введенными в медицинскую практику до 2000 года [123].

5. Количественные (QSAR) модели проницаемости веществ через ГЭБ

Как уже указывалось выше, большое число работ посвящено количественному описанию QSAR-моделей на уровне *in vivo*. В этих работах в качестве зависимой величины используется логарифм отношения концентрации вещества в мозге и крови ($\log BB = \log(C_{\text{brain}}/C_{\text{blood}})$). Несмотря на обилие публикаций, до сегодняшнего дня не установлено четко, какова зависимость между величиной $\log BB$ и молекулярными дескрипторами. В подавляющем большинстве работ для построения QSAR-моделей применяется линейная регрессия [89, 90, 92, 109, 116, 117, 121], а построенные модели основаны на использовании в качестве зависимых переменных различных физико-химических дескрипторов. Однако есть свидетельства и нелинейного характера такой зависимости [94, 124, 125].

Среди работ, основанных на множественной линейной регрессии (MLR), можно выделить публикации, в которых использовался набор сольватохромных дескрипторов [78]. В работе [78] были построены линейные регрессионные модели на основе анализа обучающей выборки, содержащей 302 соединения. В результате проведенного анализа были получены уравнения с очень хорошими статистическими характеристиками со стандартным отклонением в расчете $\log BB$ на уровне 0,30. Однако по целому ряду причин эти результаты необходимо рассматривать достаточно критически. Прежде всего это связано с тем, что между дескрипторами, использованными в работе, имеется значительная корреляция (R^2 порядка 0,70). Кроме того, для некоторых соединений в обучающей выборке имелось несколько различных значений $\log BB$; таким образом, при построении моделей число различных соединений не соответствовало числу значений $\log BB$. И, наконец, в работе [78] использовались данные, полученные в разных условиях, взятые из публикаций различных авторов. Обычно расхождение таких экспериментальных данных в логарифмической шкале составляет 0,5; следовательно, рассчитанное стандартное отклонение для модели существенно меньше ошибки эксперимента, что противоречит здравому смыслу. Таким образом, нет оснований считать, что полученная в работе [78] модель соответствует реальной действительности.

Самой сложной проблемой в работах с данными, полученными *in vivo*, является формирование обучающих выборок для моделирования проницаемости через ГЭБ. Все указанные классификационные и количественные модели на уровне *in vivo* построены на предположении реализации транспорта через ГЭБ путем диффузии, и авторы вынуждены на основе различной косвенной информации исключать из рассмотрения все соединения, которые могут проникать через ГЭБ по другим механизмам. Процедура такого изъятия сложная и достаточно произвольная. В частности, к таким соединениям относят вещества, для которых расчетные оценки имеют существенные отклонения от построенной модели («выбросы»). Если для анализируемого вещества экспериментальное значение существенно ниже расчетного, считают, что такое соединение выводится обратно в кровь за счет связывания с Р-гликопротеином, либо что его концентрация снижается за счет метаболизма. Чтобы не ухудшать статистические характеристики модели, такие соединения из расчета изымаются, что не позволяет считать удовлетворительным качество таких QSAR-моделей.

Существенно более надежную информацию о проницаемости веществ через ГЭБ можно получить в экспериментах *in vitro*. В такого рода исследованиях можно отдельно оценить поток вещества из крови в мозг и обратно; более того, можно использовать ингибиторы Р-гликопротеина и тем самым предотвратить обратный поток. В настоящее время существует ряд интересных публикаций по классификации и установлению количественных соотношений «структура – активность», основанных на таких данных.

Из публикаций в этой области заслуживает особого внимания статья [126]. В этой работе используются количественные данные по проницаемости через ячейку Caco-2 (культура клеток Caco-2 используется для моделирования пассивного транспорта; эти клетки получают из колоректальной карциномы) 53 лекарств в направлении как прямого ($a \rightarrow b$), так и обратного транспорта ($b \rightarrow a$), а также данные по ингибированию Pgp. Для построения моделей использовались молекулярные дескрипторы Volsurf [127], а также GRIND – дескрипторы, основанные на фрагментах структуры [128]. В качестве наиболее информативных дескрипторов Volsurf были использованы переменные, полученные из параметров полей взаимодействия лигандов с водой, полей карбоксильной группы, наличие доноров водородной связи и др. В результате полученная по методу частичных наименьших квадратов (PLS) дискриминантная модель правильно распознала 19 из 22 субстратов Р-гликопротеина (86,4%) и 28 из 31 веществ, не являющихся субстратами (90,3%) – средняя точность распознавания составила 88,7%. Затем полученная модель была применена к распознаванию соединений независимой тестовой выборки. Тестовая выборка содержала 115 субстратов Р-гликопротеина и 157 веществ, не являющихся субстратами. Для тестовой выборки качество распознавания было хуже: модель правильно предсказывала свойства для 72% соединений (61% для субстратов Р-гликопротеина и 81% для веществ, не являющихся субстратами). Авторы полагают, что эта модель может использоваться для компьютерного скрининга при отборе соединений, взаимодействующих (или не взаимодействующих) с Р-гликопротеином. Помимо описанной работы с указанными дескрипторами Volsurf, на основе программы GRIND авторы работы [126] построили модель для прогноза, является ли вещество субстратом или ингибитором Р-гликопротеина. В качестве обучающей выборки были использованы данные по 9 субстратам и 14 ингибиторам. Построенная «великолепная» (по определению авторов [126]) модель дискриминации между субстратами и ингибиторами была использована для прогноза свойств веществ из независимой тестовой выборки, содержащей 125 соединений (69 субстратов и 56 ингибиторов Р-гликопротеина). Правильное предсказание было получено для 61 субстрата (88,4%), и 42 ингибиторов (75%) – средняя точность предсказания составила 82,4%. В работе [126] отмечается, что наиболее важными дескрипторами являются дескрипторы, связанные с формированием водородной связи. Более того, авторами были установлены три области взаимодействия субстратов с Р-гликопротеином: O–O (области вокруг доноров водородной связи), N1–N1 (области вокруг акцепторов водородной связи) и O–N1 (области вокруг доноров и акцепторов водородной связи), размещенных на расстояниях 6,5 Å, 15,0 Å, 12,5 Å соответственно. Этот результат не противоречит выполненной ранее аналогичной работе [129], где также использовалась комбинация GRIND и Volsurf дескрипторов, в которой были выявлены две водородносвязанные акцепторные группы с расстояниями 11,5 Å и 15,0 Å в качестве ключевого элемента для распознавания субстратов. Однако при всех

достоинствах этой работы необходимо отметить существенное влияние на ее результаты точности определения пространственной (3D) структуры изучаемых соединений. В работе [126] для конвертации структуры из двумерной в трехмерную была использована компьютерная программа CORINA [130], что считается достаточно грубым приближением для обсуждения тонких деталей лиганд-рецепторных взаимодействий.

В работе [131] рассматриваются другие расстояния между акцепторными атомами азота и кислорода, как оптимальные для взаимодействия с Р-гликопротеином, а именно 2,5 Å и 4,6 Å. Не исключено, что в указанных статьях речь идет о разных сайтах связывания с Р-гликопротеином. Очевидно, что вопрос о точной геометрии взаимодействия донорных и акцепторных центров лигандов с Р-гликопротеином требует более детального изучения и обсуждения.

В работе [132] с использованием метода опорных векторов построена модель для распознавания субстратов Р-гликопротеина на основе анализа обучающей выборки, содержащей 197 веществ (99 – субстраты Р-гликопротеина, 98 – вещества, не являющиеся субстратами Р-гликопротеина). Для независимой тестовой выборки, содержащей 32 соединения, показано, что средняя точность распознавания составила около 80%. На основе гомологии построена модель Р-гликопротеина человека по рентгеноструктурным данным, полученным для Р-гликопротеина мыши. Показано [132], что с использованием молекулярного докинга удается успешно предсказывать геометрию комплексов лиганд-Р-гликопротеин. Авторами данной работы реализован Веб-сервер (<http://pgp.althotas.com>), позволяющий прогнозировать, является ли вещество субстратом Рgp, а также оценивать геометрию соответствующего комплекса лиганд-Р-гликопротеин.

Таким образом, несмотря на то, что за последнее десятилетие опубликовано множество работ, посвященных установлению взаимосвязи между структурой химических соединений и лекарств и их проницаемостью через ГЭБ с привлечением самых разнообразных дескрипторов (от топологических до физико-химических и квантово-химических) и применением самых современных математических моделей для описания зависимостей, в настоящее время создание общей картины количественного описания транспорта через ГЭБ нельзя считать завершенным. Нет даже единого мнения о виде функции взаимосвязи (линейная или нелинейная), описывающей транспорт веществ через биологические мембраны; не установлены детальные механизмы лиганд-рецепторных взаимодействий. Следует, однако, отметить, что большинство авторов опубликованных работ пришли к заключению об исключительной роли формирования водородных связей как при пассивном транспорте веществ через ГЭБ, так в процессе взаимодействия веществ с Р-гликопротеином.

Дальнейшие исследования должны быть направлены на расшифровку пространственной структуры полноразмерной молекулы Р-гликопротеина с высоким разрешением и создание QSAR-моделей, описывающих количественную взаимосвязь между структурой и активным транспортом вещества через ГЭБ.

Табл. 1. Обзор работ по компьютерному моделированию проницаемости веществ через ГЭБ

Дескрипторы	Модель	Свойство	Метод	Ссылка
Расчитанные молекулярные дескрипторы	Корреляция	logBB	Множественная линейная регрессия (MLR)	[45] [89]
Топологические дескрипторы (TD(EMCS))	Корреляция	logBB	Множественная линейная регрессия MLR	[46] [90]
Дескрипторы Абрахама	Классификация	BBB+/BBB-	Рекурсивный анализ разделения	[47] [87]
Расчитанные молекулярные дескрипторы	Корреляция	logBB	COSMO-RS	[48] [91]
Дескрипторы LFER	Корреляция	logBB	Метод к ближайших соседей (kNN), линейные соотношения свободных энергий-множественная линейная регрессия (LFER-MLR)	[49] [92]
Квантовохимические дескрипторы	Корреляция	logBB	Гауссовский процесс (GP)	[50] [93]
Log P, структурные дескрипторы	Корреляция	logBB	Линейная регрессия (LR), множественная линейная регрессия (MLR), метод частичных наименьших квадратов (PLS)	[51] [94]
Молекулярные дескрипторы, рассчитанные с помощью домашней программы	Корреляция	logBB	Гауссовский процесс (GP)	[52] [95]
Расчитанные молекулярные дескрипторы	Корреляция	logBB	Метод опорных векторов (SVM)	[53] [96]
Молекулярные дескрипторы, рассчитанные с помощью CODESA	Классификация	logBB	Искусственная нейронная сеть (ANN)	[54] [97]
Молекулярный вес, полярный атом	Корреляция	logBB	Множественный регрессионный анализ (MRA)	[55] [98]
Квантовохимические дескрипторы	Корреляция	logBB	Gaussian 03	[56] [99]
Расчитанные молекулярные дескрипторы	Корреляция	logBB	Комбинаторный QSAR	[44] [86]
Дескрипторы DRAGON	Корреляция	logBB	Генетический алгоритм (GA)	[57] [100]
Дескрипторы DRAGON	Корреляция	logBB	Метод Монте-Карло	[58] [101]

Продолжение таблицы 1

Расчитанные молекулярные дескрипторы	Корреляция + классификация	logBB	Искусственная нейронная сеть (ANN)	[59] [102]
LogP, logD, logPS	Классификация	logBB	Специфические ионизационные модели QSAR	[60] [103]
LogD, Vx (собственная программа)	Корреляция	$f_{\text{p}} \cdot f_{\text{ND}}^{-1}$	Метод Монте-Карло	[61] [104]
Дескрипторы Абрахама, P-gr(H), logP, поверхность полярных взаимодействий (PSA), объем	Корреляция + классификация	logBB	Искусственная нейронная сеть (ANN)	[62] [105]
Молекулярные дескрипторы, рассчитанные с помощью программы ADRIANA	Классификация	BBB+/BBB-	Самоорганизующаяся нейронная сеть Кохонена (KohNN), метод опорных векторов (SVM)	[63] [106]
ISIDA-дескрипторы	Корреляция	H/tissue, R/tissue ²	Многозадачное обучение (MTL), метод частичных наименьших квадратов (PLS)	[64] [107]
Расчитанные и измеренные молекулярные дескрипторы	Корреляция	Кр.щ, brain ³	Метод частичных наименьших квадратов (PLS)	[65] [108]
Молекулярные дескрипторы, рассчитанные с помощью программ SAS (version 9), PrologD (version 2.01)	Корреляция	РАМРА ⁴	Линейная регрессия (LR)	[66] [109]
Молекулярные дескрипторы, рассчитанные с помощью программ ACD Labs PhysChem Prediction Software	Корреляция	logBB	Поверхностная активность профилирования (SAP)	[67] [110]
Молекулярные дескрипторы, рассчитанные с помощью программ HuretChem® 7.51 и MarvinBeans 4.1.8 software	Корреляция	logBB	2D-количественная связь «структура — свойство» (QSPR)	[68] [111]
2D, 3D рассчитанные дескрипторы	Классификация	BBB+/BBB-	Метод опорных векторов (SVM)	[69] [112]
Расчитанные молекулярные дескрипторы	Классификация	BBB+/BBB-	Гауссовский процесс GP	[70] [113]
Молекулярный вес (MW), поверхность полярных взаимодействий (PSA), logP, топологические дескрипторы	Корреляция + классификация	logBB	Генетическая функциональная аппроксимация (GFA)	[71] [114]

Окончание таблицы 1

Дескрипторы	Модель	Свойство	Метод	Ссылка
Дескрипторы Абрахама	Корреляция	logBB	Линейные соотношения свободных энергий (LFER), стат. информационный критерий Акаике	[72] [115]
Дескрипторы Абрахама	Корреляция	РАМРА ⁴ , logBB	Множественная линейная регрессия (MLR), метод частичных наименьших квадратов (PLS)	[73] [116]
Молекулярные дескрипторы, рассчитанные с помощью программы Prism 3.0	Корреляция	Сасо-25	Линейная регрессия (LR)	[74] [117]
Дескрипторы Абрахама	Корреляция	РАМРА4 logBB	Линейные соотношения свободных энергий (LFER)	[75] [118]
2D-, 3D-дескрипторы (домашняя программа)	Классификация	Кр.иц, brain3	Метод опорных векторов (SVM), Random forest (RF), метод частичных наименьших квадратов (PLS)	[76] [119]
Дескрипторы программы DRAGON 5.4	Корреляция + классификация	Сасо-2	Кластерный анализ методом k-средних (k-MCA)	[77] [120]
Рассчитанные молекулярные дескрипторы (MW, logP, PSA)	Корреляция	Сасо-2	Линейная регрессия (LR)	[78] [121]
молекулярные дескрипторы, рассчитанные с помощью программ SYBYL, HYBOT и DRAGON	Классификация	BBB+/BBB-	Read-across, линейный дискриминантный анализ (ЛДА)	[79] [122]

fP, fND¹ – концентрация свободного вещества в плазме и ткани, соответственно; H/tissue, R/tissue² – коэффициент распределения вещества в системе «ткань – воздух»; Кр, иц, brain³ – коэффициент распределения свободного вещества в мозге; РАМРА⁴ – проницаемость искусственных параллельных мембран; Сасо-25 – клетки Сасо-2

Работа выполнена при финансовой поддержке Министерства образования и науки Российской Федерации
(госконтракт № 07.514.11.4118).

ЛИТЕРАТУРА

1. Бредбери М. Концепция гематоэнцефалического барьера (пер. с англ.), Медицина, М. 1983.
2. Покровский В.М., Коротько Г.Ф., Кобрин В.И. и др. Физиология человека: Учебник, Медицина, М. 1997.
3. Wolf S., Seehaus B., Minol K. *Naturwissenschaften*. 1996. **83**, pp. 302–311.
4. Risau W., Engelhardt B., Wekerle H. *The Journal of Cell Biology*. 1990. **110**, 1757–1766.
5. Abbott N.J., Rönnbäck L., Hansson E. *Nat. Rev. Neurosci*. 2006. **7**, pp. 41–53.
6. Lee J. *Progress in neuropathology*. 1971. **1**. pp. 84–145.
7. Hawkins B.T., Davis T.P. *Pharmacol. 2005. Rev.*, **57**, pp. 173–185.
8. Yu A.S., McCarthy K.M., Francis S.A., McCormack J.M., Lai J., Roger R.A., Lynch R.D., Schneeberger E.E. *Am. J. Physiol. Cell Physiol*. 2005. **288**, pp. 1231–1241.
9. Wolburg H., Lippoldt A. *Vasc. Pharmacol*. 2002. **38**, pp. 323–337.
10. Lee G., Dallas S., Hong M., Bendayan R. *Pharmacol. Rev*. 2001. **53**, pp. 569–596.
11. Banks W.A. *BMC Neurol*. 2009. **9** (Suppl 1), S1–S3.
12. Bernacki J. *Pharmacol. Reports*. 2008. **60**, pp. 600–622.
13. Roberts L.M., Black D.S., Raman C., et al. *Neuroscience*. 2008. **155**, pp. 423–438.
14. Tsuji A. *NeuroRx*. 2005. **2**, pp. 54–62.
15. Dahlin A., Royall J., Hohmann J.G., Wang J. *J. Pharmacol. Exp. Ther*. 2009. **329**, pp. 558–570.
16. Kanai Y., Segawa H., Miyamoto K., Uchino H., Takeda E., Endou H. *J. Biol. Chem*. 1998. **273**. pp. 23629–23632.
17. Mastroberardino L., Spindler B., Pfeiffer R., Skelly P.J., Loffing J. *Nature*. 1998. **395**, pp. 288–291.
18. Pineda M., Fernandez E., Torrents D., Estevez R. et al. *J. Biol. Chem*. 1999. **274**, pp. 19738–19744.
19. Segawa H., Fukasawa Y., Miyamoto K., Takeda E., Endou H. *J. Biol. Chem*. 1999. **274**, pp. 19745–19751.
20. Kido Y., Tamai I., Uchino H., Sai Y., Suzuki F., Tsuji A. *J. Pharm. Pharmacol*. 2001. **53**, pp. 497–503.
21. Kumagai A.K., Dwyer K.J., Pardridge W.M. *Biochem. Biophys. Acta*. 1994. **91**, pp. 24–30.
22. Oldendorf W.H. *Eur. Neurol*. 1971. **6**, pp. 49–55.
23. Oldendorf W.H. *Am. J. Physiol. (Lond)*. 1973. **224**, pp. 1450–1453.
24. Nemoto E.M., Severinghaus J.W. *Stroke*. 1997. **5**, pp. 81–84.
25. Kang Y.S., Terasaki T., Tsuji A. *J. Pharmacobiodyn*. 1990. **13**, pp. 158–163.
26. Kusahara H., Sekine T., Utsunomiya-Tate N., Tsuda M., et al. *J. Biol. Chem*. 1999. **274**, pp. 13675–13680.

27. Tamai I., Ohashi R., Nezu J., Yabuuchi H., et al. J. Biol. Chem. 1998. **273**, pp. 20378–20382.
28. Wu X., Kekuda R., Huang W., Fei Y.-J., et al. J. Biol. Chem. 1998. **273**, pp. 32776–32786.
29. Banks W.A., Audus K., Davis T.P. Peptides. 1992. **13**, 1289–1294.
30. Begley D.J. J. Pharm. Pharmacol. 1996. **48**, pp. 136–146.
31. Molina-Arcas M., Casado F.J., Pastor-Anglada M. Curr. Vasc. Pharmacol. 2009. **7** (4), pp. 426–434.
32. Bhutia Y.D., Hung S.W., Patel B., Lovin D., Govindarajan R. Cancer Res. 2011. **71** (5), pp. 1825–1835.
33. Nagai K., Nagasawa K., Fujimoto S. Cancer Chemother. Pharmacol. 2005. **55**(3), pp. 222–230.
34. Roninson I.B. Structure and evolution of P-glycoprotein. From , edited by I. Roninson, pp. 189–209. Plenum Press, New York. 1991.
35. C.F. Higgins Res. Microbiol. 2001. **152**, pp. 205–210.
36. Bakos E. Eur. J. Physiol. 2007. **453**, pp. 621–641.
37. Dean M., Hamon Y., Chimini G. J. Lipid. Res. 2001. **42**, pp. 1007–1017.
38. Bakos E., Homolya L. Pflugers. Arch. 2007. **453**(5), pp. 621–641.
39. Ramachandra M., Ambudkar S. V., Chen D. et al. Biochemistry. 1998. **37** (14), pp. 5010–5019.
40. Lam F. C., Liu R., Lu P., et al. J. Neurochem. 2001. **76**, pp. 1121–1128.
41. Miller D.S., Bauer B., Hartz A.M.S. Pharmacol. Rev. 2008. **60**, pp. 196–209.
42. Bendayan R., Ronaldson P.T., Gingras D., Bendayan M. J. Histochem. Cytochem. 2006. **54**, pp. 1159–1167.
43. Higgins C.F., Linton K.J. Nat. Struct. Mol. Biol. 2004. **11**(10), 918–926.
44. Aszalos A. Drug. Discov. 2007. Today, **12**, pp. 833–837.
45. Rosenberg M.F., Kamis A.B., Callaghan R., Higgins C.F., Ford R.C. J. Biol. Chem. 2003. **278**, pp. 8294–8299.
46. Rosenberg M.F., Callaghan R., Modok S., Higgins C.F., Ford R.C. J. Biol. Chem. 2005. **280**, pp. 2857–2862.
47. Aller S.G., Yu J., Ward A., Weng Y., et al. Science. 2009. **323**, pp. 1718–1722.
48. Seelig A., Landwojtowicz E. Eur. J. Pharm. Sci. 2000. **12**, pp. 31–40.
49. Ecker G., Huber M., Schmid D., Chiba P. Mol Pharmacol. 1999. **56**, pp. 791–796.
50. Торшин Н. А., Власова В. И. Основы физиологии человека - Учебник для студентов вузов, обучающихся по медицинским и биологическим специальностям. 2-е издание, исправленное. РУДН, М. 2001.
51. Desai M.C., Thadeio P.F., Lipinski Chr.A., Liston D.R., Spencer R.W and Williams I.H. Bioorg. Med. Chem. Lett. 1991. **8**, pp. 411–414.
52. Abraham M.H., Lieb W.R. and Franks N.P. J. Pharm. Sci. 1991. **80**, pp. 719–724.
53. Abraham M.H., Chadha H.S. and Mitchell R.C. J. Pharm. Sci. 1995. **84**, pp. 1257–1268.

54. *Abraham M.H., Chadha H.S. and Mitchell R.C.* Drug Des. Discov. 1995. **13**, pp. 123–131.
55. *Abraham M.H., Takacs-Novak K. and Mitchell R.C.* J. Pharm. Sci. 1997. **86**, pp. 310–315.
56. *Van de Waterbeemd H. and Kansy M.* Chimia. 1992. **46**, pp. 299–303.
57. *Seelig A., Gottschlich R. and Devant R.M.* Proc. Natl. Acad. Sci. USA. 1994. **91**, pp. 68–72.
58. *Ter Laak A.M., Tsai R.S., Donne-Op den Kelder G.M., Carrupt P.A. and Testa B.* Eur. J. Pharm. Sci. 1994. **2**, pp. 373–384.
59. *Chikhale E.G., Ng K.-Y., Burton Ph.S. and Borchardt R.T.* Pharm. Res. 1994. **3**, pp. 412–419.
60. *Calder J.A. and Ganellin C.R.* Drug Des. Discov. 1994. **11**, pp. 259–268.
61. *Basak S.C., Gute B.D. and Drewes L.R.* Pharm. Res. 1996. **13**, 775–778.
62. *Kai J., Nakamura K., Masuda T., Ueda I. and Fujiwara H.* J. Med. Chem. 1996. **39**, pp. 2621–2624.
63. *Lombardo F., Blake J.F. and Curatolo W.J.* J. Med. Chem. 1996. **39**, pp. 4750–4755.
64. *Young R.C., Mitchell R.C., Brown Th.H., Ganellin C.R., et al.* J. Med. Chem. 1988. **31**, pp. 656671.
65. *NGoka V., Schlewler G., Linget J.M., et al.* J. Med. Chem. 1991. **34**, pp. 2547–2557.
66. *Testa B. and Seiler P.* Drug Res. 1981. **31**, pp. 1053–1058.
67. *Van de Waterbeemd H. and Testa B.* Adv. Drug Res. 1987. **16**, pp. 85–225.
68. *Tayar El., Tsai N., Testa R.-S., B. et al.* J. Pharm. Sci. 1991. **80**, pp. 590–598.
69. *Tayar El., Testa N., Carrupt P.-A.* J. Phys. Chem. 1992. **96**, pp. 1455–1459.
70. *Van de Waterbeemd H., Camenisch G., Folkers G., Chretien J. R., Raevsky O.A.* J. Drug Targeting. 1998. **6**, pp. 151–165.
71. *Leo A., Hansch C. and Elkins D.* Chem. Rev. 1971. **71**, pp. 525–616.
72. *Camenisch G., Folkers G. and Van de Waterbeemd H.* Pharm. Acta Helv. 1996. **71**, pp. 309–327.
73. *Van de Waterbeemd H., Kansy M., Wagner B. and Fischer H.* In Lipophilicity in Drug Action and Toxicology (Pliska, V., Testa, B. and Van de Waterbeemd, H. eds.), Weinheim: VCH. 1996.
74. *Xiang T.X. and Anderson B.D.* J. Membr. Biol. 1994. **140**, pp. 111–122.
75. *Abraham M.H. et al.* J. Pharm. Sci. 1994. **83**, pp. 1257–1268.
76. *Luco J.M.* J. Chem. Inf. Comput. Sci. 1999. **39**, pp. 396–404.
77. *Clark D.E.* J. Pharm. Sci. 1999. **88**, pp. 815–821.
78. *Abraham M.H., Ibrahim A., Zhao Y., Acree W. E.* J. Pharm. Sci. 2006. **95**, pp. 2091–2100.
79. *Subramanian G., Kitchen D.B.* J. Comp.-Aided Mol. Design. 2003. **17**, pp. 643–664.
80. *Sun H.* J. Chem. Inf. Comput. Sci. 2004. **44**, pp. 748–757.
81. *Adenot M., Lahana R.* J. Chem. Inf. Comput. Sci. 2004. **44**, pp. 239–248.
82. *Liu X., Ma P.X.* Drug Metab. Dispos. 2004. **32**, 132–139.

83. *Cabrera M.A., Bermejo M., Perez M., Ramos R.J.* Pharm. Sci. 2004. **93**, pp. 1701–1717.
84. *Mensch J., Oyarzabal J., Mackie C., Augustijns P. J.* Pharm. Sci. 2009. **98**, pp. 4429–4468.
85. *Mehdipour A.R., Hamidi M.* Drug Discovery Today. 2009. **14**, pp. 1030–1036.
86. *Zhang L., Zhu H., Oprea T.I., Golbraikh A., Tropsha A.* Pharm. Res. 2008. **25**, pp. 1902–1914.
87. *Zhao Y.H., Abraham M.H., et al.* J. Chem. Inf. Model. 2007. **47**, pp. 170–175.
88. *Li H., Yap C. W., Ung C. Y., Xue Y., et al.* J. Chem. Inf. Model. 2005. **45**, p. 1376–1384.
89. *Al-Fahemi J.H.A., Cooper D. L., Allan N.L.* J. Molecular Graphics and Modelling. 2007. **26**, pp. 607–612.
90. *Cuadrado M.U., Ruiz I.L., A'ngel M.* J Comput Chem. 2007. **28**, pp. 1252–1260.
91. *Wichmann K. et al.* J. Chem. Inf. Model. 2007. **47**, pp. 228–233.
92. *Konovalov D. A. et al.* J. Chem. Inf. Model. 2007. **47**, pp. 1648–1656.
93. *Obrezanova O, Csa'nyi G., Gola J.M.R., Segall M.D.* J. Chem. Inf. Model. 2007. **47**, pp. 1847–1857.
94. *Wan H. et al.* J. Med Chem. 2007. **50**, pp. 4606–4615.
95. *Obrezanova O. et al.* J. Comput. Aided Mol. 2008. **22**, pp. 431–440.
96. *Kortagere S., Chekmarev D., Welsh W.J., Ekins S* Pharmaceutical Research. 2008. **25**, pp. 1836–1845.
97. *Guerra A., Pa'ez J.A., Campillo N. E.* QSAR Comb. Sci. 2008. **27**, pp. 586–594.
98. *Fu X.-C., Wang G.-P., Shan H.-L. et al.* European Journal of Pharmaceutics and Biopharmaceutics. 2008. **70**, pp. 462–466.
99. *Van Damme S., Langenaeker W., Bultinck P.* Journal of Molecular Graphics and Modelling. 2008. **26**, pp. 1223–1236.
100. *Shen J., Dub Y., Zhaoa Y., Liua G., Tanga Y.* QSAR Comb. Sci. 2008. **27**, No.6, pp. 704–717.
101. *Konovalov D.A., Sim N., Deconinck E. et al.* J. Chem. Inf. Model. 2008. **48**, pp. 370–383.
102. *Karelson M. et al.* ARKIVOC. 2008. pp. 38–60.
103. *Lanevskij K., Japertasa P., Didziapetrisa R., Petrauskasa A.* Chemistry & Biodiversity. 2009. **6**, pp. 2050–2054.
104. *Guo Q., Brady M., Gunn R. N.* J Nucl Med. 2009. **50**, pp. 1715–1723.
105. *Chen Y., Zhub Q.-J., Pana J., Yanga Y., Wuc X.-P.* Computer methods and programs in biomedicine. 2009. **95**, pp. 280–287.
106. *Wang Z., Yan A., Yuan Q.* QSAR Comb. Sci. 2009. **28**, pp. 989–994.
107. *Varnek A., Gaudin C., Marcou G. et al.* J. Chem. Inf. Model. 2009. **49**, pp. 133–144.
108. *Friden M., Winiwarter S., Jerndal G., Bengtsson O., Wan H.* J. Med. Chem. 2009. **52**, pp. 6233–6243.

109. Di L., Kerns E.H., Bezar I.F., Petusky S.L., Huangdi Y. J. Pharm Sciences. 2009. **98**, pp. 1980–1991.
110. Petereit A.C., Swinney K., Mensch J., Mackie C., Stokbroekx S., Brewster M., Dressman J. B. Eur. J. Pharm. Biopharm. 2010. **75**, pp. 405–410.
111. Sá M.M., Pasqualoto K.F. M., Rangel-Yagui C.O. Brazilian J. Pharm. Sci. 2010. **46**, pp. 741–751.
112. Shen J., Cheng F., Xu Y., Li W. and Tang Y. J. Chem. Inf. Model. 2010. **50** (6), pp. 1034–1041.
113. Obrezanova O. and Segall M.D. J. Chem. Inf. Model. 2010. **50**, pp. 1053–1061.
114. Fan Y. J. Chem. Inf. Model. 2010. **50**, pp. 1123–1133.
115. Abraham M. et al. J. Pharm. Sci. 2010. **99**, pp. 2492–2501.
116. Mensch J. et al. International Journal of Pharmaceutics. 2010. **395**, pp. 182–197.
117. Mabondzo A., Bottlaender M., Guyot A.-C., Tsaouin K., Robert Deverre J., Balimane P. V. Molecular Pharmaceutics. 2010. **7**, pp. 1805–1815.
118. Tsinman O., Tsinman K., Sun N., Avdeef A. Pharm Res. 2011. **28**, pp. 337–363.
119. Chen H. et al. J. Mol. Graph. Model. 2011. **29**, pp. 985–995.
120. The et al. H. P. Mol. Inf. 2011. **30**, pp. 376–385.
121. Lacombe O., Videau O., Chevillon D., Guyot A.-C. et al. Mol. 2011. **8**, pp. 651–663.
122. Raevsky O. A., Solodova S. L., Raevskaya O. E., Liplavskiy Y. V., and Mannhold R. Biochemistry (Moscow) Supplement Series B: Biomedical Chemistry. 2012. **6**, pp. 31–38.
123. Lipinski C.A. Drugs structure and properties, past and present. Can we design drugs with beautiful properties? http://www.iainm.demon.co.uk/spring99/lipins_n.pdf.
124. Hughes J.D. et al. Bioorg. Med. Chem. Lett. 2008. **18**, pp. 4872–4875.
125. Zhang, H. J. Pharm. Sci. 2004. **93**, pp. 1595–1604.
126. Crivori P., Reinach B., Pezzetta D., Poggese I. Mol. Pharm. 2006. **3**, pp. 33–44.
127. Volsurf, version 4.0; available from Molecular Discovery Ltd., London, U.K. (www.moldiscovery.com).
128. GRID, version 22; available from Molecular Discovery Ltd., London, U.K. (www.moldiscovery.com).
129. Cianchetta G., Singleton R. W., Zhang M., Wildgoose M., Giesing D., Fravolini A., Cruciani G., Vaz R. J. J. Med. Chem. 2005. **48**, pp. 2927–2935.
130. CORINA, version 3.2; available from Molecular Networks, GmbH, Computerchemie, Erlangen, Germany (www.mol-net.de).
131. Seelig A. Eur. J. Biochem. 1998. **251**, pp. 252–261.
132. Bikadi Z, Hazai I, Malik D, Jemnitz K, Veres Z, et al. PLoS ONE. 2011. **6** (10), e25815.

*В.Ю. Григорьев¹, А.Н. Раздольский¹, А.О. Загребин²,
В.Д. Тонкопий², О.А. Раевский¹*

КЛАССИФИКАЦИОННЫЕ QSAR-МОДЕЛИ ОСТРОЙ ТОКСИЧНОСТИ ОРГАНИЧЕСКИХ СОЕДИНЕНИЙ ПО ОТНОШЕНИЮ К *DAPHNIA MAGNA*

*¹Институт физиологически активных веществ РАН,
Черноголовка, Московская область*

²Институт озераведения РАН, Санкт-Петербург

Среди различных характеристик вещества особое место занимает острая токсичность. Знание количественных величин этого свойства совершенно необходимо для определения степени опасности химического соединения по отношению к здоровью человека и состоянию окружающей среды. Для оценки острой токсичности широкое распространение в качестве объектов изучения приобрели водные организмы: инфузории [1], рыбы [2], бактерии [3], членистоногие [4], водоросли [5]. Важным элементом при проведении комплексного токсикологического исследования вещества является его отнесение к той или иной группе на основе классификационных моделей. В настоящее время широкое применение получили методы «количественная связь “структура – активность”» (QSAR), использующие классификацию химических соединений на основе типов токсического действия (MOA – mode of action) [6]. В литературе описаны различные схемы MOA, включающие, например, 4-MOA [7], 8-MOA [8]. Около половины используемых в промышленности соединений вызывают у водных организмов острые токсические эффекты, соответствующие концепции «неполярной неспецифической токсичности» [9]. Поэтому в случае бинарной классификации целесообразным представляется выделение двух классов соединений, обладающих неспецифической и специфической острой токсичностью. В качестве критериев отнесения соединений к тому или иному классу могут выступать тип химической связи и биологическая мишень в организме [10]. С этой точки зрения к первому классу можно отнести соединения, образующие обратимые нековалентные комплексы с липидами биологических мембран. К этой группе относятся соединения с MOA неполярный и полярный наркоз [11].

Ко второму классу, в рамках рассматриваемой задачи, можно отнести соединения с другими МОА. Решающее значение для получения хороших классификационных моделей имеет качество обучающих выборок. В случае водных организмов можно выделить наборы данных для *Pimephales promelas* (617 соединений, 8 МОА [8]), *Poecilia reticulata* (293 соединения, 4 МОА [12]), *Tetrahymena pyriformis* (983 соединения [13]), (514 соединений, 1 МОА [14]). Для *Daphnia magna* известны большие ряды соединений (например, 300 соединений, различные МОА [15]), однако они сформированы на основе различных литературных источников. Поэтому разработка нового токсикологического ряда экспериментальных данных для *Daphnia magna* является актуальной задачей. За основу был взят ряд из 119 органических соединений с экспериментально определенными величинами токсичности [16].

Целью работы является создание бинарных классификационных моделей неспецифической и специфической острой токсичности по отношению к *Daphnia magna* на основе нового расширенного экспериментального ряда органических соединений.

Экспериментальная часть

Величины LC_{50} определяли по методике, описанной в работе [16]. Для исследования острой токсичности использовали культуру генетически однородных организмов *Daphnia magna* двух-трехдневного возраста. При определении токсичности в стаканчики с 50 мл раствора различных концентраций исследуемого вещества помещали по 10 рачков. Растворы готовили на воде, используемой для культивирования дафний (водопроводной, дехлорированной) при температуре 18–19 °С. Растворы веществ использовали в 5–10 возрастающих по логарифмической шкале концентрациях. Каждая серия опытов сопровождалась контрольными испытаниями. Токсичность исследуемых соединений определяли по величине LC_{50} – концентрации (ммоль/л) соединений, приводящей к гибели 50% организмов при 24-часовой инкубации. Величины LC_{50} рассчитывали по методу пробит-анализа Литчфилда и Уилкоксона в модификации В. Прозоровского [17].

Физико-химические, топологические и электронные 2D-дескрипторы 443 химических соединений рассчитывали с использованием компьютерных программ HUBOT [18] (26 дескрипторов) и DNESTR [19] (51 дескриптор). Фрагментацию молекул проводили на основе программы MOLDIVS [20]. Отбор дескрипторов осуществляли с помощью модифицированного подхода, основанного на отбрасывании дескрипторов, имеющих постоянное значение в более чем 95% случаев, и удалении дескрипторов с учетом пороговых величин коэффициентов корреляции [21] и коэффициентов вариации. Классификацию молекул проводили с использованием линейного дискриминантного анализа (ЛДА) и k-ого ближайшего соседа (кБС) из библиоте-

ки компьютерных программ IMSL [22], метода опорных векторов (МОВ), реализованного в программе LibSVM [23], искусственных нейронных сетей (ИНС, многослойный перцептрон) в составе статистического пакета программ SPSS 17.0 [24]. Результаты бинарной классификации оценивали с помощью «чувствительности» (Sens), «специфичности» (Spec) и «точности» (Acc), которые рассчитывали по формулам:

$$\text{Sens} = \text{TP}/(\text{TP} + \text{FN}), \text{Spec} = \text{TN}/(\text{TN} + \text{FP}), \text{Acc} = (\text{TP} + \text{TN})/(\text{TP} + \text{FN} + \text{TN} + \text{FP}),$$

где TP и TN – число правильно классифицированных молекул в первой и второй группе, FN и FP – число неправильно классифицированных молекул в первой и второй группе соответственно. Исходный ряд из 443 соединений разбивали на обучающую (355 соединений) и тестовую (88 соединений) выборки путем упорядочивания соединений по убыванию токсичности и отбора каждого пятого соединения в тестовый ряд. При этом в группу с неспецифической токсичностью попадали соединения с МОА неполярный и полярный наркоз, а в группу со специфической токсичностью – остальные соединения.

Результаты и их обсуждение

Используемый в настоящей работе структурный ряд (приложение 1) состоит из 443 органических соединений, большинство из которых являются полифункциональными и содержат молекулярные фрагменты, характерные для спиртов, кислот, кетонов, простых и сложных эфиров, аминов, нитросоединений, фосфорильных и тиофосфорильных соединений, галогенпроизводных, нитрилов, ароматических и других соединений. Величины токсичности $\log(1/\text{LC}_{50})$ (LC_{50} , ммоль/л) меняются в широком интервале от $-2,90$ до $5,66$ при среднем значении $0,86$ и стандартном отклонении $1,36$. Все соединения были отнесены к одному из трех типов МОА: неполярный наркоз (279 молекул, $63,0\%$), полярный наркоз (37 молекул, $8,4\%$), другие (127 молекул, $28,6\%$).

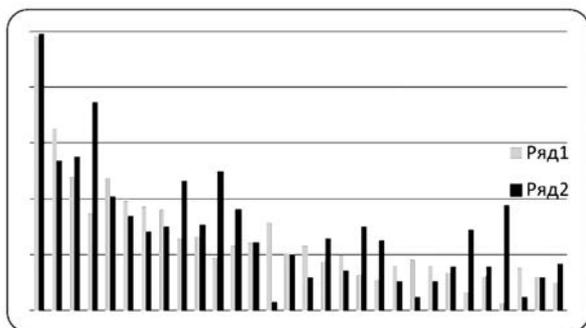
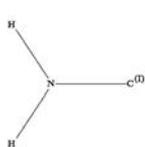


Рис. 1. Относительная частота встречаемости 30 молекулярных фрагментов в ряду соединений с неспецифической (ряд 1) и специфической (ряд 2) токсичностью

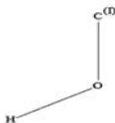
кул, $63,0\%$), полярный наркоз (37 молекул, $8,4\%$), другие (127 молекул, $28,6\%$).

Для выявления структурных различий между соединениями с неспецифической и специфической токсичностью был проведен анализ частоты встречаемости 30 наиболее распространенных молекулярных фрагментов (рисунок 1).

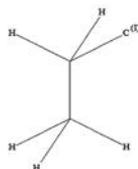
Наиболее существенные отличия в частоте встречаемости (в три и более раза) были выявлены для фрагментов 14, 22, 25 и 27:



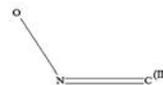
14



22



25



27

Фрагменты 14 и 22 преимущественно встречаются в соединениях с неспецифической токсичностью, в то время как фрагменты 25 и 27 характерны для соединений со специфической токсичностью.

В таблице 1 представлены результаты бинарной классификации молекул с использованием методов ЛДА, кБС, МОВ и ИНС (№ 1–16). Для обучающей выборки степень общего правильного распознавания (Acc) в большинстве случаев составляет 0,9 и выше. Максимальная величина Acc = 0,955 наблюдается при использовании МОВ и 77 дескрипторов. В тестовой выборке достигается величина Acc = 0,932 при применении кБС в пространстве из 77 дескрипторов. Максимальная степень правильной классификации в первой группе Sens = 0,984 (кБС, 77 дескрипторов). Во второй группе наилучшие результаты (Spec = 0,875) наблюдаются в случае ИНС (42 и 77 дескрипторов). При увеличении числа дескрипторов от 19 до 77 при проведении классификации наблюдается общая закономерность, заключающаяся в увеличении общей степени правильного распознавания. В случае кБС величина Acc меняется от 0,909 до 0,932, МОВ 0,886 и 0,920, ИНС 0,875 и 0,920 соответственно. При использовании ЛДА также наблюдается отмеченная закономерность для случая с 19 и 31 дескриптором.

Табл. 1. Результаты классификации органических соединений (Sens – «чувствительность», Spec – «специфичность», Acc – «точность») на две группы: неспецифическая/специфическая острая токсичность по отношению к *Daphnia magna* с использованием методов линейного дискриминантного анализа (ЛДА), к-ближайшего соседа (кБС), метода опорных векторов (МОВ) и искусственных нейронных сетей (ИНС) при различных величинах дескрипторного пространства.

№	Метод	Число дескрипторов	Обучающая выборка			Тестовая выборка		
			Sens	Spec	Acc	Sens	Spec	Acc
1	ЛДА	19	0,940	0,689	0,868	0,938	0,667	0,864
2	кБС	19	0,937	0,786	0,893	0,969	0,750	0,909
3	МОВ	19	0,960	0,854	0,930	0,953	0,708	0,886
4	ИНС	19	0,980	0,795	0,926	0,969	0,625	0,875

№	Метод	Число дескрипторов	Обучающая выборка			Тестовая выборка		
			Sens	Spec	Acc	Sens	Spec	Acc
5	ЛДА	31	0,933	0,806	0,896	0,906	0,833	0,886
6	кБС	31	0,944	0,796	0,901	0,953	0,792	0,909
7	МОВ	31	0,948	0,757	0,893	0,969	0,708	0,898
8	ИНС	31	0,975	0,783	0,919	0,969	0,792	0,920
9	ЛДА*	42	–	–	–	–	–	–
10	кБС	42	0,933	0,806	0,896	0,938	0,792	0,898
11	МОВ	42	0,952	0,777	0,901	0,969	0,750	0,909
12	ИНС	42	0,970	0,892	0,947	0,938	0,875	0,920
13	ЛДА*	77	–	–	–	–	–	–
14	кБС	77	0,940	0,767	0,890	0,984	0,792	0,932
15	МОВ	77	0,976	0,903	0,955	0,953	0,833	0,920
16	ИНС	77	0,930	0,843	0,905	0,875	0,875	0,875
17	КМ**	77	0,952	0,825	0,915	0,969	0,833	0,932

Примечание. *Сингулярная матрица; **КМ – консенсусная модель (кБС + МОВ + ИНС).

Учитывая, что лучшие результаты классификации в двух группах достигаются на основе разных методов, выглядит целесообразным разработка консенсусной модели (КМ) с участием кБС, МОВ и ИНС в пространстве из 77 дескрипторов. В этом случае однозначное отнесение молекул к той или иной группе соединений проводилось на основе простого голосования с использованием трех оценок. Полученный результат (таблица 1, № 17) свидетельствует о том, что КМ по своим классификационным характеристикам в обучающей выборке (Acc = 0,915) превосходит методы кБС и ИНС и уступает только МОВ. А в тестовой выборке КМ показывает такой же результат (Acc = 0,932), что и метод кБС.

В литературе нам удалось обнаружить данные по классификации органических соединений на две группы: наркоз и остальные МОА [15, 25, 26]. Для корректного сравнения с этими данными дополнительно была проведена классификация 443 исследуемых соединений на две группы: неполярный наркоз / остальные МОА с использованием модели МОВ в пространстве из 77 дескрипторов. При этом было установлено, что величина Acc в обучающей и тестовой выборке составила 0,966 и 0,886 соответственно. Полученный результат вполне сопоставим с литературными данными по классификации химических соединений с МОА наркоз. Так, авторами публикации [15] раз-

работана схема дискриминации, основанная на использовании базовой токсичности и структурных фрагментов, которая позволяет отделять соединения с МОА наркоз от остальных соединений с точностью 0,92. В работе [25] при классификации 194 соединений с МОА наркоз при использовании ЛДА и четырех физико-химических дескрипторов общая степень правильного распознавания составила 0,95. Подобный результат ($Acc = 0,90$) был достигнут авторами работы [26] при бинарной классификации 88 химических соединений, относящихся к неполярным наркотикам и реактивным соединениям, на основе методов линейного дискриминантного анализа и логистической регрессии.

Таким образом, разработанные классификационные QSAR-модели по своим статистическим параметрам находятся на одном уровне с приводимыми в литературе данными по дискриминации соединений с МОА наркоз. Это свидетельствует о том, что предложенный в настоящей работе ряд из 443 органических соединений с охарактеризованными типами токсического действия и экспериментальными величинами LC_{50} может быть использован в качестве источника данных для проведения дальнейших токсикологических исследований.

ЛИТЕРАТУРА

1. Schwöbel J.A.H, Madden J.C., Cronin M.T.D. Chemosphere. 2011. 85 (6), pp. 1066–1074.
2. Mager E.M., Esbaugh A.J., Brix K.V. et al. Comp. Biochem. Physiol. C. 2011. 153 (1), pp. 82–90.
3. Shi J.-Q., Cheng J., Wang F.-Y. et al. Ecotoxicol. Environ. Saf. 2012. 78 (1), pp. 134–141.
4. S. Kar, K. Roy, J. Hazard. Mater. 2010. 177, pp. 344–351.
5. Mitchell R.J., Myers A.L., Mabury S.A. et al. Ecotoxicol. Environ. Saf. 2011. 74 (8), pp. 2260–2267.
6. Netzeva T.I., Pavan M. and Worth A.P. QSAR Comb. Sci. 2008. 27 (1), pp. 77–90.
7. Verhaar H.J.M., C.J. van Leeuwen, and Hermens J.L.M. Chemosphere. 1992. 25 (4), pp. 471–491.
8. Russom C.L., Bradbury S.P., Broderius S.J. et al. Environ. Toxicol. Chem. 1997. 16 (5), pp. 948–967.
9. Nendza M., Wenzel A. Environ. Sci. Pollut. Res. 2006. 13 (3), pp. 192–203.
10. Escher B.I., Hermens J.L.M. Environ. Sci. Technol. 2002. 36 (20), pp. 4201–4217.
11. Schultz T.W., Cronin M.T.D., Walker J.D. et al. J. Mol. Struct. (Theochem). 2003. 622 (1–2), pp. 1–22.
12. Katritzky A.R., Tatham D.B. J. Chem. Inf. Comput. Sci. 2001. 41 (5), pp. 1162–1176.

13. Zhu H., Tropsha A., Fourches D. et al. J. Chem. Inf. Model. 2008. 48 (4), pp. 766–784.
14. Ellison C.M., Cronin M.T.D., Madden J.C. et al. SAR QSAR Environ. Res. 2008. 19 (7–8), pp. 751–783.
15. P.C. von der Ohe, Kuhne R., Ebert R.-U. et al. Chem. Res. Toxicol. 2005. 8(3), pp. 536–555.
16. Раевский О.А., Раздольский А.Н., Тонкопий В.Д. и др. Хим.-фармац. ж. 2008. 42 (5), pp. 49–55.
17. Прозоровский В.Б., Прозоровская М.П., Демченко В.М. Фармакол. и токсикол. 1978. 41 (4), pp. 497–502.
18. Раевский О.А., Григорьев В.Ю., Тrepали С.В. НУВОТ. Свидетельство РОСПАТЕНТА об официальной регистрации программы для ЭВМ № 990090. 26.02.1999.
19. Raevsky O., Sapegin A., Zefirov N., in: QSAR: Rational Approaches to the Design of Bioactive Compounds, C. Silipo and A. Vittoria, (eds.), Elsevier, Amsterdam.1991. pp. 189–192.
20. Gerasimenko V.A., Trepalin S.V., Raevsky O.A., in: Molecular Modeling and Prediction of Bioactivity, K. Gundertofte and F.S. Jorgensen, (eds.), Kluwer Academic/Plenum Publishers, New York.2000. pp. 423–424.
21. Lessigiarska I., Cronin M.T.D., Worth A.P. et al. SAR QSAR Environ. Res. 2004. 15 (3), pp. 169–190.
22. URL: <http://software.intel.com>.
23. URL: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
24. URL: <http://www.spss.ru>.
25. S. Ren, Toxicol. Lett. 2002. 133, pp. 127–139.
26. S. Ren, T.W. Schultz, Toxicol. Lett. 2002. pp.129, 151–160.

*О.А. Раевский, С.В. Трепалин, В.Ю. Григорьев, С.Л. Солодова,
А.В. Ярков, О.Е. Раевская*

КОМПЬЮТЕРНЫЙ РАСЧЕТ ПРОНИЦАЕМОСТИ ФАРМАКОЛОГИЧЕСКИХ ВЕЩЕСТВ ЧЕРЕЗ ГЕМАТОЭНЦЕФАЛИЧЕСКИЙ БАРЬЕР

*Институт физиологически активных веществ РАН,
Черноголовка, Московская область*

Нарушения деятельности центральной нервной системы (ЦНС) во многих случаях являются результатом влияния на организм различных эндо- и экзогенных факторов. Для обеспечения нормального функционирования ЦНС организмом в процессе эволюции был выработан ряд защитных механизмов, которые обеспечивают надежность функционирования ЦНС даже в агрессивных условиях. Одним из важнейших защитных механизмов является гематоэнцефалический барьер (ГЭБ). Под этим термином условно объединяют совокупность анатомических и функциональных систем, которые поддерживают постоянство и регулируют состав внутренней среды головного и спинного мозга [1, 2].

ГЭБ регулирует распределение различных веществ между током крови и мозгом и обладает высокой избирательностью. Барьерная функция состоит в предотвращении проникновения «вредных» веществ из крови в мозг и их активном выведении в тех случаях, если они каким-то образом попали в ЦНС. ГЭБ играет ведущую роль в возникновении таких патологий, как инсульты, микробные инфекции, черепно-мозговые травмы, нейродегенеративные заболевания, которые, в свою очередь, могут приводить к изменениям проницаемости ГЭБ. С другой стороны, ГЭБ ограничивает проницаемость в ЦНС многих лекарственных препаратов, что может служить препятствием при терапии ряда заболеваний центральной нервной системы. Детально этот вопрос обсуждается в недавнем нашем обзоре [3].

Транспорт химических соединений из крови в мозг преимущественно осуществляется посредством пассивной диффузии. Пассивная диффузия

реализуется по градиенту концентрации из тока крови в мозг без энергетических затрат. Происходит также транспорт из мозга в кровь за счет связывания веществ с полигликопротеином (P-gp). Структура этого белка в настоящее время установлена не полностью, что затрудняет процесс моделирования его взаимодействия с веществами.

В связи с важностью выше указанных процессов в литературе в настоящее время накопилась масса публикаций, посвященных установлению взаимосвязи структуры органических соединений и лекарств и их проницаемостью через ГЭБ методом пассивной диффузии. В этих работах используются различные методы установления количественной связи «структура – активность» (КССА), включая такие мощные современные методы, как искусственные нейронные сети и метод опорных векторов. А в качестве дескрипторов привлекаются самые разнообразные топологические, квантовохимические, физико-химические, двух- и трехмерные дескрипторы. Тем не менее, к данному моменту не установлена точно даже форма (линейная или нелинейная) зависимости проницаемости органических соединений и лекарств от структурных и физико-химических параметров. Недавно нами были опубликованы две статьи, связанные с рассматриваемой тематикой, в которых демонстрируется решающая роль водородных связей в транспорте химических соединений и лекарств в организме человека [4, 5].

В настоящей работе описывается метод расчета проницаемости органических соединений и лекарств через ГЭБ.

Экспериментальная часть

Обучающая выборка была сформирована на основе публикации Абрахама [6] и Келдера [7], содержащей сведения о проницаемости через ГЭБ путем пассивного транспорта 298 соединений. С целью проверки устойчивости модели исходные 298 соединений разбивались на внутреннюю обучающую (239) и тестовую (59) выборки. Использовались также в качестве внешней тестовой выборки данные по logBB для пяти дополнительных соединений, не включенных в общую обучающую выборку или внутреннюю обучающую и тестовую выборки.

Описание программы

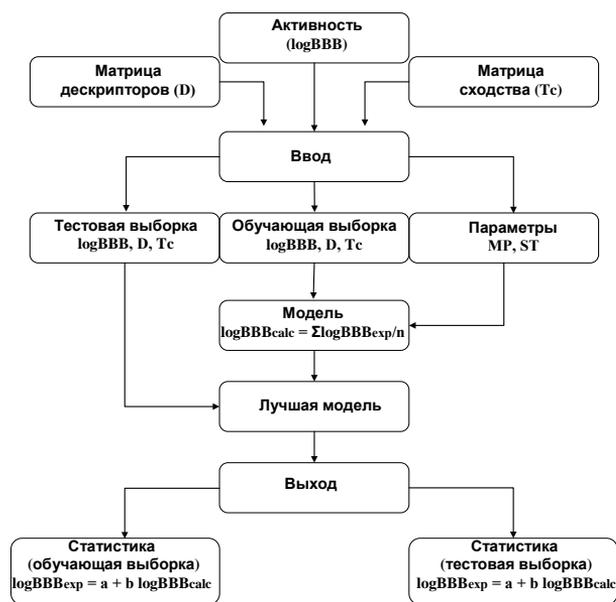
В основу алгоритма расчета положена модель АМТ, развитая недавно нами [8, 9].

В рамках этой модели значение рассматриваемого свойства рассчитывается как среднее арифметическое значение свойства в одной или нескольких парах ближайших структурно родственных соединений. При этом в каждой паре один из соседей должен иметь большее, а второй сосед меньшее (или наоборот) значение используемого дескриптора по отношению к дескриптору рассматриваемого соединения. В этом подходе каж-

дый дескриптор используется отдельно, то есть один дескриптор – одна модель, и из совокупности таких моделей выбирается одна или несколько моделей с наилучшими статистическими критериями. В этом методе расчета использовались 33 физико-химических дескриптора, рассчитанные программой НУВОТ [10]. Проведенная работа по использованию этих дескрипторов показала, что три модели, основанные на различных одиночных дескрипторах: $\sum(Ead)$ – сумма энтальпийных акцепторных и донорных факторов, $\sum(Cad)$ – сумма свободно энергетических водородно-связанных акцепторных и донорных факторов, а также композитный дескриптор, $\sum(Cad)/\alpha$, где α – молекулярная поляризуемость, – приводят к стабильным моделям.

На общей обучающей выборке из 298 соединений модели, основанные на отдельном использовании каждого из указанных дескрипторов, привели к следующим значениям квадрата коэффициента корреляции и стандартного отклонения: 0,594 и 0,406; 0,588 и 0,409; 0,577 и 0,415, соответственно.

Как указывалось выше, в указанных моделях используются одиночные дескрипторы. Однако на финальной стадии можно получить согласованную модель с использованием результатов расчета рассматриваемого свойства. В этом случае в качестве конечного результата можно использовать среднее арифметическое рассчитанных значений из моделей по отдельным дескрипторам. Применение такого подхода для трех указанных выше лучших моделей ($\sum(Ead)$,



D – дескрипторы; Tc – индекс Танимото; МР – число молекулярных пар; ST – порог сходства

Рис. 1. Принципиальная схема расчета

$\sum(Cad)$, $\sum(Cad)/\alpha$) приводит к следующим результатам: $R^2 = 0,641$ и $SD = 0,382$. Последняя цифра близка к значению SD , равному 0,370, полученному при корреляции экспериментальных значений $\log BB$ одних и тех же соединений, взятых из разных источников. Это свидетельствует о том, что полученная нами финальная модель имеет стандартное отклонение на уровне ошибки экспериментального определения этого параметра. Схема расчета представлена ниже.

Согласно указанной схеме на первом этапе рассчитываются матрицы используемых дескрипторов и структурного сходства для рассматриваемых соединений, а затем осуществляется расчет указанной модели. Результаты расчета logBB по указанной фиксированной модели регистрируются на выходном файле.

Описанный метод расчета использован нами в созданной программе, зарегистрированной в Федеральной службе по интеллектуальной собственности [11].

На заключительном этапе для тестирования описанной схемы расчета были привлечены данные по проницаемости пяти дополнительных соединений [12, 13] (пять дополнительных соединений с точным указанием характера действия), это три вещества с пассивным механизмом его транспорта (loperamide, chlorpromazine, Y-G16) и два вещества с активным механизмом транспорта (terfenadine, Y-G9). Результаты расчета приведены в таблице 1. Из рассмотрения этой таблицы очевидно, что не только соединения с пассивным транспортом, но и два соединения, имеющие существенный вклад активного транспорта, хорошо описываются предложенной моделью (таблица 1). Это позволяет предположить, что разработанный метод расчета может быть применен к соединениям, проникающим через ГЭБ различными механизмами. Привлечение в дальнейшей нашей работе в этой области новых дополнительных данных для соединений с различным механизмом транспорта позволит проверить это предположение.

Очевидно, что указанный подход представляет особую ценность для крупных фармацевтических компаний, которые занимаются постоянным поиском новых оригинальных лекарственных веществ и имеют собственные базы данных, содержащие множество структурно разнообразных химических соединений.

Табл. 1. Результаты расчета величин logBB тремя различными моделями, основанными на одном дескрипторе

Название	Smile	LogBB AMP (ΣCad)	LogBB AMP (ΣCad/α)	LogBB AMP (TPSA)	LogBB AMP D/3	LogBB _{exp}	
Внешний тестовый ряд							
1	Loregamide	<chem>Clc1ccc(cc1)C4(O)CCN(CCC(c2ccccc2)(c3ccccc3)C(=O)N(C)C)CC4</chem>	0,45	0,94	0,45	0,61	0,77
2	Chlorpromazine	<chem>CN(C)CCCN1c2ccccc2Sc3c1cc(cc3)Cl</chem>	1,02	1,02	0,70	0,91	0,84
3	Y-G16	<chem>NCCc1nccs1</chem>	0,25	0,25	-0,14	0,12	0,19
4	Terfenadine	<chem>OC(c1ccccc1)(c2ccccc2)C4CCN(CCCC(O)c3ccccc3)C(C)C)CC4</chem>	0,45	0,70	0,45	0,53	0,64
5	Y-G9	<chem>[O-][N+](=O)C=C(\NC)NCCSc1oc(cc1)CN(C)C</chem>	-1,03	-1,03	-1,27	-1,11	-1,23

Программа создана в рамках государственного контракта № 07.514.11.4118

ЛИТЕРАТУРА

1. Бредбери М. Концепция гематоэнцефалического барьера (пер. с англ.), Медицина, Москва. 1983.
2. Покровский В.М., Коротько Г.Ф., Кобрин В.И. и др. Физиология человека: Учебник. Медицина, Москва. 1997.
3. Раевский О.А., Солодова С.Л., Лагунин А.А., Поройков В.В. Журнал биомедицинской химии (в печати).
4. Raevsky O.A., Solodova S.L., Raevskaya O.E., Liplavskiy O.E., Mannhold R., Biochem. (Moscow) Suppl. Ser. B: Biomed. Chem. 2012. 6, pp. 31–38.
5. Раевский О.А., Солодова С.Л., Раевская О.Е., Манхольд Р. Химико-фармацевтический журнал. 2012. 46 (3), с. 3–8.
6. Abraham M.H., Ibrahim A., Zhao Y., Acree W. E. J. Pharm. Sci. 2006. 95, pp. 2091–2100.
7. Kelder J., Grootenhuis P.D.J., Bayada D. M., Delbressine L.P.C., Ploemen J.-P. Pharm. Res. 1999. 16, pp. 1514–1519.
8. Raevsky O.A., Grigor'ev V.Yu., Modina E.A., Worth A. SAR QSAR Environ. Res. 2010. 21, pp. 265–275.
9. Raevsky O.A., Grigor'ev V.Yu., Modina E.A., Worth A. Mol. Inf. 2011. 30, pp. 267–275.
10. Raevsky O.A., Skvortsov V.S. 3D HYBOT program, Russian State Patent Department, № 004612207.
11. Программа зарегистрирована Федеральной службой по интеллектуальной собственности (№ 2012614339); правообладатель – ЗАО НБК «ВИСТ», авторы программы являются авторами этой статьи.
12. Crivori P., Reinach B., Pezzetta D., Poggesi I. J. Mol. Pharm. 2006. 3, pp. 33–44.
13. .Mente S.R., Lombardo F. J. Comp. Aid. Mol. Design. 2005. 19, pp. 465–481.

*В.Ю. Григорьев¹, С.Л. Солодова¹, Д.Е. Полианчик¹,
Дж. С. Дьерден², О.А. Раевский¹*

КЛАССИФИКАЦИОННЫЕ МОДЕЛИ ВЗАИМОСВЯЗИ СТРУКТУРЫ ЛЕКАРСТВЕННЫХ СОЕДИНЕНИЙ И ИХ Р-ГЛИКОПРОТЕИНОВОЙ АКТИВНОСТИ

*¹Институт физиологически активных веществ РАН,
Черноголовка, Московская область*

*²School of Pharmacy and Biomolecular Sciences,
Liverpool John Moores University, Liverpool, U.K.*

Введение

Р-гликопротеин (Pgp) представляет собой крупный трансмембранный белок с молекулярной массой 170 кДа, состоящий из 1280 аминокислотных остатков [1]. Pgp является наиболее важным транспортером, препятствующим достижению терапевтических концентраций ряда физиологически активных веществ в головном мозгу [2].

Взаимодействие низкомолекулярных соединений с Pgp происходит по весьма сложному механизму, который обусловлен тем, что этот белок содержит несколько сайтов связывания, а также тем, что в результате связывания Р-гликопротеина с АТФазой происходит изменение его конформации, что также влияет на связывание белка с субстратом [3].

Пространственная структура этого белка в настоящее время мало изучена. Так, в работе [4] была представлена трехмерная структура Pgp с разрешением $\sim 8 \text{ \AA}$ в комплексе с АМР–PNP и в публикации [5] структура Pgp мышцы – с разрешением около 4 \AA . Хотя эти результаты и дают представление об общей трехмерной структуре и функционировании Pgp, достигнутая точность еще далека от необходимой для молекулярного моделирования взаимодействия Pgp с субстратами и ингибиторами. Ввиду этого понятно наличие в литературе многих публикаций, представленных в обзоре [6], связанных с моделированием *in silico* взаимосвязи между структурой соединений и их транспортом в центральную нервную систему (ЦНС).

Анализ этих публикаций показывает, что к исследованиям в этой области привлечены практически все современные QSAR-технологии (методы, дескрипторы, модели). Тем не менее здесь можно отметить два важных обстоятельства: использование при построении моделей достаточно большого числа дескрипторов, что затрудняет интерпретацию полученных результатов, и недостаточно глубокое описание водородного связывания. Первое обстоятельство препятствует созданию устойчивых предсказательных механистических моделей «структура – активность» для модификации структуры химических соединений с целью получения веществ с заданными свойствами. Что же касается водородного связывания, в настоящее время это взаимодействие считается одним из важных (если не самым важным) в межмолекулярных взаимодействиях вообще и в субстрат-рецепторных комплексах в частности. Однако количественное описание этого эффекта в рамках QSAR или молекулярного моделирования подчас проводится не достаточно корректно.

В настоящем исследовании поставлена задача создания устойчивых предсказательных классификационных QSAR-моделей взаимодействия химических соединений и лекарств с Pgp и их транспорта через гематоэнцефалический барьер (ГЭБ) на основе небольшого числа физико-химических дескрипторов с включением в эти модели дескрипторов водородной связи.

Методика

Данные

В настоящей работе были использованы тестовые литературные данные по определению субстратной специфичности соединений к Р-гликопротеину, основанные на изучении клеточной проницаемости. При этом в качестве основной характеристики были использованы величины ER (Efflux Ratio), которые представляют собой отношение коэффициентов проницаемости соединений в двух противоположных направлениях. Соединения, имеющие величины $ER > 2$, рассматривались как субстраты, остальные – как несубстраты. В обучающую выборку были включены данные по субстратной специфичности соединений, полученные на основе линии клеток MDCK-MDR1 [7], а в тестовую выборку вошли данные из работы [8], основанные на использовании клеток Caco-2, исключая общие соединения из этих двух публикаций. Таким образом, данные по классификации «субстрат/несубстрат» включали 221 соединение. 177 соединений обучающей выборки содержали 71 субстрат и 106 несубстратов, а 44 соединения тестового ряда содержали 20 субстратов и 24 несубстрата. На рисунке 1 представлено распределение соединений по наличию в них 17 наиболее распространенных в органической химии функциональных групп. Из рисунка очевидно пропорциональное присутствие одних и тех же функциональных групп в обучающей и тестовой выборках.

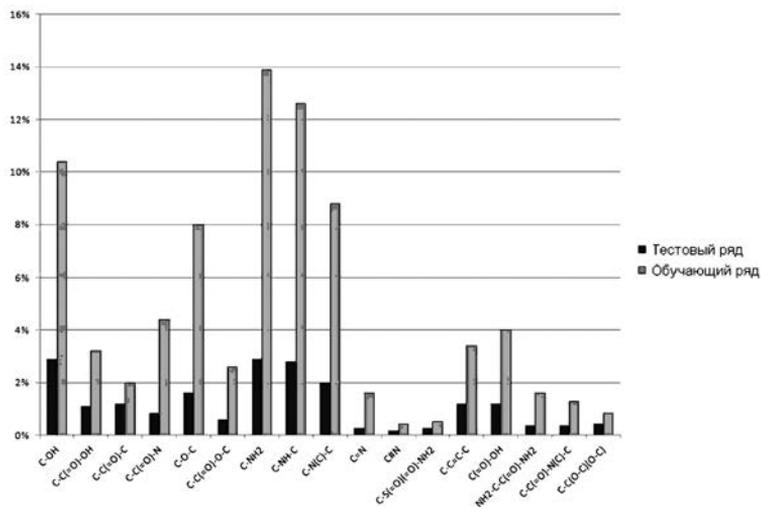


Рис. 1. Частота встречаемости (%) 17 наиболее распространенных функциональных групп в соединениях обучающего и тестового ряда

Дескрипторы

С использованием компьютерных программ HVBOT [9], MOLTRA [10], DRAGON [11] для описания молекулярной структуры соединений были рассчитаны 15 дескрипторов, заметная роль которых в транспорте веществ в живых организмах была уже установлена. Это молекулярные дескрипторы, связанные, в основном, с описанием межмолекулярного взаимодействия, размера молекул, липофильности. Дальнейший отбор дескрипторов был связан с необходимостью обеспечения линейной независимости переменных. Для этого был проведен анализ корреляционной матрицы дескрипторов с использованием итерационной процедуры, которая состояла из ряда шагов: 1) выбор наиболее информативного дескриптора с максимальной величиной коэффициента вариации; 2) формирование вокруг этого выбранного дескриптора кластера родственных дескрипторов на основе граничной величины коэффициента корреляции (0,95); 3) удаление из матрицы родственных дескрипторов; 4) повторение шагов 1–3 до остановки процедуры. В результате было отобрано семь дескрипторов (таблица 1).

Табл. 1. Физико-химические молекулярные дескрипторы

Дескриптор	Интерпретация	Интервал	Программа
ΣQ^-	Сумма отрицательных атомных зарядов	$-7,78 \div -0,66$	HVBOT
ΣCad	Сумма свободноэнергетических Н-акцепторных и Н-донорных дескрипторов	$2,70 \div 38,50$	HVBOT

Дескриптор	Интерпретация	Интервал	Программа
MW	Молекулярная масса, а.е.м.	141,17 ÷ 1202,63	HYBOT
PSA _{ca}	Парциальный поверхностный дескриптор суммарной энтальпийной Н-акцепторной способности, Å ²	1,80 ÷ 45,96	MOLTRA
PSA _{ca}	Парциальный поверхностный дескриптор суммарной свободноэнергетической Н-акцепторной способности, Å ²	1,64 ÷ 54,53	MOLTRA
PSA _{cd}	Парциальный поверхностный дескриптор суммарной энтальпийной Н-донорной способности, Å ²	0,00 ÷ 15,95	MOLTRA
TPSA _{tot}	Общая поверхность полярных взаимодействий, Å ²	3,24 ÷ 278,80	DRAGON

Методы классификации

Бинарную классификацию соединений проводили с использованием трех методов: линейного дискриминантного анализа (ЛДА), случайного леса (СЛ) и опорных векторов (ОВ). Учитывая небольшое число отобранных дескрипторов (7) для расчета классификационных моделей, был использован метод полного перебора всех возможных комбинаций дескрипторов. В качестве критерия для отбора наиболее значимых моделей использовали коэффициент корреляции Мэтьюза (МСС). При этом применяли два подхода. В первом случае отбирали модели с максимальным значением кросс-валидированных значений МСС. Во втором случае выбор осуществляли на основе максимальных величин МСС тестовой выборки.

Линейный дискриминантный анализ проводили с использованием компьютерной программы LDA [12]. В качестве метрики в исследуемом пространстве дескрипторов использовали расстояние Махаланобиса.

Для классификации на основе метода случайного леса использовали программу rf5new [13]. При этом параметры метода имели следующие значения: число деревьев ($jbt = 500$), число случайно выбираемых переменных ($mtry0 = (M + 0,5)^{0,5}$), где M – общее число дескрипторов в модели.

Один из вариантов опорных векторов [14] был применен для создания классификационных моделей. Для решения задачи использовали радиальную базисную функцию с заданными параметрами (без их оптимизации). Дескрипторы автошкалировали на основе формулы: $\hat{A}_i = (A_i - \bar{A})/S$, где A_i и \hat{A}_i – исходные и автошкалированные величины дескриптора для i -й молекулы, \bar{A} – средняя величина и S – стандартное отклонение дескриптора.

Бинарные классификационные модели оценивали на основе TP (число правильно распознанных соединений, принадлежащих к первому классу (субстраты)), FN (число ошибочно распознанных соединений среди перво-

го класса), TN (число правильно распознанных соединений, принадлежащих ко второму классу (несубстраты)) и FP (число ошибочно распознанных соединений среди второго класса). С использованием величин TP, FN, TN и FP рассчитывали общую точность ($ACC = (TP + TN) / (TP + FN + TN + FP)$) и корреляционный коэффициент Мэтьюза ($MCC = (TP * TN - FP * FN) / ((TP + FP) * (TP + FN) * (TN + FP) * (TN + FN))$), 0,5.

Валидацию полученных классификационных моделей проводили на основе двух подходов. В первом случае использовали внутреннюю кросс-валидацию обучающей выборки с числом блоков $k = 5$. (В методе СЛ аналогичную роль выполняла процедура ООВ (out-of-bag)). Во втором случае применяли внешнюю валидацию на основе тестовой выборки.

Результаты и обсуждение

Классификационные модели, полученные оптимизацией кросс-валидированных величин MCC

Среди специалистов по QSAR-моделированию уже достаточно длительное время ведется дискуссия о предпочтительности внутренней или внешней валидации для оценки предсказательной способности моделей [15]. В связи с этим при выборе классификационных моделей мы использовали два подхода. В первом случае после расчета на основе обучающей выборки моделей с одинаковым числом дескрипторов выбиралась та модель, которая имела максимальную кросс-валидированную величину MCC_{cv} (таблица 2).

Табл. 2. Статистические характеристики классификационных моделей, отобранных на основе кросс-валидированных величин MCC_{cv} обучающего ряда

Модель	Метод	Дескрипторы	Обучающий ряд		Тестовый ряд	
			ACC_{cv}	MCC_{cv}	ACC	MCC
1	ЛДА	ΣC_{ad}	0,785	0,546	0,818	0,639
2	ЛДА	ΣC_{ad} , MW	0,802	0,584	0,727	0,456
3	ЛДА	ΣQ^- , ΣC_{ad} , MW	0,819	0,620	0,636	0,262
4	ЛДА	ΣQ^- , ΣC_{ad} , MW, PSA_{ca}	0,808	0,597	0,659	0,310
5	ЛДА	ΣQ^- , ΣC_{ad} , MW, PSA_{ca} , $TPSA_{tot}$	0,797	0,572	0,591	0,232
6	ЛДА	ΣQ^- , ΣC_{ad} , MW, PSA_{ca} , PSA_{ca} , PSA_{ed} , $TPSA_{tot}$	0,751	0,473	0,591	0,232
7	СЛ	$TPSA_{tot}$	0,678	0,335	0,727	0,467
8	СЛ	MW, $TPSA_{tot}$	0,797	0,575	0,591	0,203
9	СЛ	ΣC_{ad} , MW, $TPSA_{tot}$	0,780	0,537	0,636	0,283

Модель	Метод	Дескрипторы	Обучающий ряд		Тестовый ряд	
			ACC _{cv}	MCC _{cv}	ACC	MCC
10	СЛ	ΣC_{ad} , MW, PSA _{ca} , PSA _{ca}	0,780	0,539	0,591	0,192
11	СЛ	ΣQ^- , ΣC_{ad} , MW, PSA _{ca} , TPSA _{tot}	0,780	0,537	0,682	0,365
12	СЛ	ΣQ^- , ΣC_{ad} , MW, PSA _{ca} , PSA _{ca} , PSA _{cd} , TPSA _{tot}	0,763	0,500	0,659	0,324
13	ОВ	MW	0,729	0,437	0,659	0,309
14	ОВ	MW, TPSA _{tot}	0,734	0,442	0,682	0,356
15	ОВ	ΣC_{ad} , MW, TPSA _{tot}	0,740	0,467	0,727	0,466
16	ОВ	ΣQ^- , ΣC_{ad} , MW, TPSA _{tot}	0,734	0,454	0,750	0,509
17	ОВ	ΣQ^- , ΣC_{ad} , MW, PSA _{ca} , TPSA _{tot}	0,723	0,424	0,659	0,309
18	ОВ	ΣQ^- , ΣC_{ad} , MW, PSA _{ca} , PSA _{ca} , PSA _{cd} , TPSA _{tot}	0,695	0,343	0,705	0,402

Как видно из представленных данных, при использовании ЛДА среди одиночных дескрипторов наибольшим дискриминирующим эффектом обладает ΣC_{ad} , т. е. дескриптор, который характеризует суммарную Н-акцепторную и Н-донорную способность молекул. Уже только использование одного этого дескриптора позволило построить модель (1) с достаточно высокими статистическими характеристиками (ACC_{cv} = 0,785, MCC_{cv} = 0,546). Дальнейшее увеличение размера дескрипторного пространства немного улучшало статистику. Максимальные величины были получены для модели (3) из трех дескрипторов (ACC_{cv} = 0,819, MCC_{cv} = 0,620). Дальнейшее увеличение числа дескрипторов ухудшало статистические показатели.

Похожее общее поведение наблюдалось при последовательном конструировании моделей на основе СЛ. Однако при этом ведущее положение в качестве главного дискриминирующего фактора занимал топологический дескриптор TPSA_{tot}. Максимальный эффект наступал при числе дескрипторов равным 2 (модель 8, ACC_{cv} = 0,797, MCC_{cv} = 0,575). Как видно из представленных данных, статистические характеристики ЛДА- и СЛ-моделей близки между собой, однако оптимальные комбинации дескрипторов при этом отличались.

При создании классификационных ОВ-моделей наиболее «весомым» среди одиночных дескрипторов оказалась молекулярная масса MW (модель 13, ACC_{cv} = 0,729, MCC_{cv} = 0,437). Наилучшая модель (15) была получена на основе трех дескрипторов: ΣC_{ad} , MW, TPSA_{tot}. По своим статистическим

характеристикам ($ACC_{cv} = 0,740$, $MCC_{cv} = 0,467$) она несколько уступает оптимальным моделям ЛДА (3) и СЛ (8).

Интересно отметить разное поведение статистических показателей внешней тестовой выборки в зависимости от используемого метода и размерности пространства. Так, в случае ЛДА при переходе от одномерной к оптимальной многомерной классификации величины ACC и MCC резко уменьшаются: $ACC = 0,818$, $MCC = 0,639$ (модель 1) и $ACC = 0,636$, $MCC = 0,262$ (модель 3). Аналогичное поведение демонстрируют СЛ-модели: $ACC = 0,727$, $MCC = 0,467$ (модель 7) и $ACC = 0,591$, $MCC = 0,203$ (модель 8). Напротив, при использовании ОБ-классификации переход от одномерного (13) к трехмерному (15) дескрипторному пространству сопровождался увеличением степени общего правильного распознавания (ACC) с 65,9% до 72,7%.

Классификационные модели, полученные оптимизацией величин MCC тестовой выборки.

Другой метод выбора лучших классификационных моделей был основан на использовании максимальных величин MCC тестовой выборки (таблица 3).

Табл. 3. Статистические характеристики классификационных моделей, отобранных на основе величин MCC тестового ряда

Модель	Метод	Дескрипторы	Обучающий ряд		Тестовый ряд	
			ACC _{cv}	MCC _{cv}	ACC	MCC
19	ЛДА	PSA _{ca}	0,746	0,462	0,841	0,682
20	ЛДА	PSA _{ca} , PSA _{ed}	0,757	0,485	0,886	0,771
21	ЛДА	PSA _{ca} , PSA _{ca} , PSA _{ed}	0,763	0,498	0,886	0,771
22	ЛДА	PSA _{ca} , PSA _{ca} , PSA _{ed} , TPSA _{tot}	0,718	0,403	0,864	0,726
23	ЛДА	$\Sigma Q-$, ΣC_{ad} , MW, PSA _{ca} , TPSA _{tot}	0,791	0,559	0,705	0,427
24	СЛ	PSA _{ca}	0,644	0,264	0,750	0,494
25	СЛ	PSA _{ca} , PSA _{ed}	0,718	0,403	0,818	0,650
26	СЛ	PSA _{ca} , PSA _{ed} , TPSA _{tot}	0,701	0,366	0,886	0,781
27	СЛ	ΣC_{ad} , PSA _{ca} , PSA _{ed} , TPSA _{tot}	0,746	0,465	0,795	0,598
28	СЛ	ΣC_{ad} , PSA _{ca} , PSA _{ca} , PSA _{ed} , TPSA _{tot}	0,734	0,444	0,795	0,612
29	ОБ	$\Sigma Q-$	0,712	0,398	0,818	0,671
30	ОБ	$\Sigma Q-$, ΣC_{ad}	0,723	0,435	0,818	0,671

Модель	Метод	Дескрипторы	Обучающий ряд		Тестовый ряд	
			ACC _{cv}	MCC _{cv}	ACC	MCC
31	OB	$\Sigma Q-, \Sigma C_{ad}, PSA_{ca}$	0,712	0,398	0,818	0,671
32	OB	$\Sigma Q-, \Sigma C_{ad}, PSA_{ca}, PSA_{ca}$	0,695	0,351	0,864	0,748
33	OB	$\Sigma Q-, \Sigma C_{ad}, PSA_{ca}, PSA_{ca}, TPSA_{tot}$	0,689	0,337	0,864	0,748

При использовании для классификации методов ЛДА и СЛ можно отметить ряд общих моментов. В качестве основного одиночного дискриминирующего дескриптора выступал PSA_{ca} , т. е. дескриптор, связанный с описанием способности соединений к образованию водородной связи. Уже одно использование этого дескриптора приводило к получению моделей (19, 24) с высокой (84,1%, ЛДА) и хорошей (75,0%, СЛ) способностью к правильному распознаванию. Дальнейшее увеличение числа дескрипторов улучшало статистические характеристики. В случае ЛДА максимальные показатели для модели (20) из двух дескрипторов PSA_{ca} , PSA_{ed} составили $ACC = 0,886$, $MCC = 0,771$. При использовании СЛ лучшая классификационная модель (26) была получена на основе трех дескрипторов – PSA_{ca} , PSA_{ed} и $TPSA_{tot}$ – с аналогичными модели (20) высокими статистическими характеристиками: $ACC = 0,886$, $MCC = 0,781$.

Основным дискриминирующим фактором при конструировании OB-моделей (29)÷(33) проявила себя сумма отрицательных атомных зарядов $\Sigma Q-$. Уже одиночная модель (29) демонстрировала хорошую статистику: $ACC = 0,818$, $MCC = 0,671$. Интересно при этом отметить ступенчатый характер изменения статистических характеристик при изменении числа переменных. Модели с числом дескрипторов от одного до трех (29)÷(31) и с числом дескрипторов от четырех до пяти (32)÷(33) имеют одинаковые величины MCC и ACC . Лучшая модель (32) из четырех дескрипторов ($\Sigma Q-, \Sigma C_{ad}, PSA_{ca}, PSA_{ca}$) имеет показатели ($ACC = 0,864$, $MCC = 0,748$), сопоставимые с характеристиками ЛДА- и СЛ-моделей.

Необходимо отметить отсутствие общей корреляции в изменении величин MCC (ACC) и соответствующих величин MCC_{cv} (ACC_{cv}) в моделях (19)÷(33). Так, в отличие от моделей (1)÷(18), где рост MCC_{cv} сопровождался падением величин MCC тестовой выборки с коэффициентом корреляции Пирсона $r = -0,3$, изменение величин MCC в ряду моделей (19)÷(33) сопровождалось более слабой линейной связью с величиной $r = -0,2$.

Изменение критерия, положенного в основу отбора моделей, сильно влияет на состав оптимального дескрипторного пространства. Так, лучшая ЛДА-модель (3) сконструирована с использованием трех дескрипторов: $\Sigma Q-, \Sigma C_{ad}, MW$. Соответствующая ЛДА-модель (20) создана на основе со-

всем другим дескрипторам: PSA_{ca} и PSA_{cd} . При переходе от одной модели к другой также сильно меняются статистические характеристики. Например, в модели (3) $MCC_{cv} = 0,620$, $MCC = 0,262$, а в модели (20) $MCC_{cv} = 0,485$, $MCC = 0,771$.

Ранее в литературе уже были опубликованы результаты бинарной классификации 187 соединений (177 из которых использованы в настоящей работе в качестве обучающего ряда) [7]. При этом с применением методов СЛ, ОВ и к-ближайшего соседа было создано 23 классификационные модели на основе 30 различных дескрипторов. Лучшие результаты были получены с использованием 3D молекулярных дескрипторов VS+. При этом минимальное число дескрипторов было равно 4, и величины ACC_{cv} (ACC) и MCC_{cv} (MCC) составляли $\sim 0,80$ и $\sim 0,60$ соответственно. Разработанные в настоящей работе классификационные модели при сопоставимых статистических характеристиках выгодно отличаются тем, что они имеют меньшее число дескрипторов (например, модели (8) и (3) созданы на основе всего двух и трех дескрипторов). Кроме того, использованные дескрипторы являются двумерными, что снимает ряд вопросов, связанных с моделированием структуры молекул.

Заключение и выводы

С использованием методов линейного дискриминантного анализа, случайного леса и опорных векторов на основе различных комбинаций, содержащих от одного до семи молекулярных дескрипторов с ясной физико-химической интерпретацией, созданы классификационные модели, имеющие хорошие статистические характеристики. При этом для отбора статистически значимых QSAR-моделей применены два критерия: кросс-валидированное значение корреляционного коэффициента Мэтьюза MCC и MCC тестовой выборки. В результате установлено, что оптимальный набор дескрипторного пространства сильно зависит от используемого критерия. Практически во всех моделях присутствуют дескрипторы водородной связи, что подтверждает ее важную роль в процессах Рgp-транспорта лекарственных соединений через гематоэнцефалический барьер.

ЛИТЕРАТУРА

1. Li Y., Yuan H., Yang K., Xu W., Tang W., Li X. *Curr. Med. Chem.* 2010. 17(8), pp. 786–800. DOI:10.2174/092986710790514507.
2. Miller D.S., Bauer B., Hartz A.M.S. *Pharmacol. Rev.* 2008. 60, pp. 196–209. DOI:10.1124/pr.107.07109.
3. Higgins C.F., Linton K.J. *Nat. Struct. Mol. Biol.* 2004. 11 (10), pp. 918–926. DOI:10.1038/nsmb836.

4. *Rosenberg M.F., Callaghan R., Modok S., Higgins C.F., Ford R.C.* J. Biol. Chem. 2005. 280, pp. 2857–2862. DOI:10.1074/jbc.M410296200.

5. *Aller S.G., Yu J., Ward A., Weng Y., Chittaboina S., Zhuo R., Harrell P.M., Trinh Y.T., Zhang Q., Urbatsch I.L., Chang G.* Science. 2009. 323, pp. 1718–1722. DOI:10.1126/science.1168750.

6. *Раевский О.А., Солодова С.Л., Лагунин А.А., Поройков В.В.* Биомедицинская химия. 2014. 60 (2), с. 161–181. DOI:10.18097/PBMC20146002161.

7. *Broccatelli F.* J. Chem. Inf. Model. 2012. 52, pp. 2462–2470. DOI:10.1021/ci3002809.

8. *Crivori P., Reinach B., Pezzetta D., Poggese I.* Molecular Pharmaceutics. 2006. 3, pp. 33–44. DOI:10.1021/mp050071a.

9. *Раевский О.А., Григорьев В.Ю., Трепалин С.В.* Свидетельство об официальной регистрации программы для ЭВМ HYBOT (Hydrogen Bond Thermodynamics) № 990090 от 26 февраля 1999 г., Москва, Федеральная служба по интеллектуальной собственности, патентам и товарным знакам.

10. *Раевский О.А., Трепалин С.В., Раздольский А.Н.* Свидетельство об официальной регистрации программы для ЭВМ MOLTRA (Molecular Transform Analysis) № 990092 от 26 февраля 1999 г., Москва, Федеральная служба по интеллектуальной собственности, патентам и товарным знакам.

11. URL: <http://www.taletе.mi.it>.

12. *Murtagh F., Heck A.* Multivariate Data Analysis, D. Reidel Publ. Co., Dordrecht, 1987. pp. 128–133.

13. URL: http://www.stat.berkeley.edu/~breiman/RandomForests/cc_examples/prog.f.

14. URL: <https://github.com/jbcolme/fortran-ls-svm>.

15. *Veerasley R., Rajak H., Jain A., Sivadasan S., Varghese C.P., Agrawal R.K.* Int. J. Drug Des. Discov. 2011. 2 (3), pp. 511–519.

В.Ю. Григорьев, О.Е. Раевская, А.В. Ярков, О.А. Раевский

QSAR-МОДЕЛИРОВАНИЕ ОСТРОЙ НЕЙРОТОКСИЧНОСТИ РЯДА ОРГАНИЧЕСКИХ РАСТВОРИТЕЛЕЙ ПО ОТНОШЕНИЮ К ГРЫЗУНАМ

*Институт физиологически активных веществ РАН,
Черноголовка, Московская область*

Введение

Важным компонентом при разработке новых физиологически активных веществ является оценка их токсичности и, в частности, нейротоксичности [1]. Исследованию данного явления уделяется определенное внимание, в том числе с использованием такого междисциплинарного подхода, как анализ количественных взаимосвязей «структура – активность» (QSAR) [2–5]. В опубликованных источниках практически отсутствуют данные по острой нейротоксичности, которые по своему количественному и качественному составу соответствуют требованиям QSAR. Одним из немногих примеров является массив из 48 органических растворителей [6], на основе которого был создан ряд линейных регрессионных моделей [7, 8]. Однако этим разработкам присущи некоторые недостатки, связанные, главным образом, или с невысокими статистическими показателями, или с относительно большим числом неясно интерпретируемых параметров.

Целью настоящей работы было конструирование линейных и нелинейных QSAR-моделей нейротоксичности с удовлетворительными статистическими характеристиками на основе дескрипторов с ясным физико-химическим содержанием.

Материалы и методы

В литературе описаны различные нейротоксические эффекты и используемые для их изучения тесты, которые могут быть использованы для количественной оценки нейротоксичности соединений [9], включая величину ТД₅₀ (доза, вызывающая нейротоксический эффект у 50% животных) по тесту вращающегося стержня в исследованиях на мышах при внутрибрю-

шинном введении вещества [1]. В настоящей работе использовали литературные данные [6, 7] по острой сублетальной токсичности (EC_{30} , мкМ/л, ингаляционно) 47 органических растворителей по отношению к мышам и крысам. Оценку токсичности (по данным вышеуказанных авторов) проводили на основе установленного ранее нейротоксического эффекта, который проявляется в появлении электрически вызываемых судорог у животных в условиях эксперимента. Для количественной оценки токсичности исследовались зависимости «концентрация – эффект» (скорость удлинения конечностей для мышей и продолжительность удлинения конечностей для крыс) и определялась величина EC_{30} (эффективная концентрация, оказывающая нейротоксическое действие у 30% животных).

Для описания структуры соединений применили 15 физико-химических дескрипторов, включая поляризуемость, парциальные атомные заряды, дескрипторы водородной связи, рассчитанные на основе программы NYBOT [10], липофильность [11] и дипольный момент [12]. Названия соединений, CAS-номера, данные по нейротоксичности и физико-химические дескрипторы приведены в Приложении. При создании регрессионных моделей использовали следующие статистические методы: линейная регрессия (ЛР), случайный лес (СЛ), опорные вектора (ОВ) и соответствующие им компьютерные программы: SVD [13], rforest [14] и flssvm [15]. В качестве статистических характеристик моделей фигурировали: n – число молекул, m – число дескрипторов, r^2 – квадрат коэффициента линейной корреляции, s – стандартное отклонение, q^2 и s_{cv} – квадрат коэффициента линейной корреляции и стандартное отклонение в условиях перекрестного контроля по k -блокам ($k = 10, 100$ итераций), соответственно. Для оценки погрешности коэффициентов модели ЛР использовали стандартные ошибки. Расчет регрессионных моделей выполняли путем перебора всех возможных комбинаций из 1–5 дескрипторов. Лучшие модели определяли по минимальной величине s . При близких значениях s предпочтение отдавалось моделям с меньшей величиной m . Для определения области применимости (ОП) моделей использовали простой подход, который заключался в определении интервалов изменения зависимых и независимых переменных.

Результаты и обсуждение

В таблице 1 представлены результаты расчетов регрессионных моделей нейротоксичности соединений по отношению к крысам. При исследовании всех 47 соединений лучшие результаты показала модель случайного леса № 2. Регрессионные зависимости, полученные на основе методов ЛР и ОВ (№ 1 и № 3 соответственно), имели более низкие величины r^2 (q^2) и более высокие значения s (s_{cv}). Дальнейшее улучшение качества моделей было достигнуто путем анализа Y -выбросов. При этом из обучающей выборки удалялись соединения, у которых абсолютная разность между

экспериментальными и рассчитанными величинами превышала 2s. Полученные модели (№ 4 ÷ № 6) закономерно имели более высокие статистические характеристики. С помощью метода СЛ были сконструированы модели с наиболее высокими показателями описательной статистики ($r^2 = 0,851$; $s = 0,19$). Однако (что более важно) при этом была получена более плохая предсказательная статистика по сравнению с ЛР ($q^2 = 0,662/0,777$ и $s_{cv} = 0,28/0,26$ соответственно).

Анализируя состав дескрипторов, на основе которых получены регрессионные зависимости, можно отметить следующее. Всего для конструирования использовали семь молекулярных дескрипторов. При этом дескриптор $\Sigma(C_a)$, характеризующий Н-акцепторную способность молекул, встречался во всех моделях. Вторым по частоте использования являлся дипольный момент μ . Остальные молекулярные характеристики встречались реже.

Табл. 1. Статистические характеристики моделей нейротоксичности по отношению к крысам

№	Метод	n	M	r ²	s	q ²	s _{cv}	Дескрипторы*	Y-выбросы**
1	ЛР	47	4	0,740	0,28	0,560	0,37	α , $\Sigma(Q^+)$, $\Sigma(C_a)$, μ	-
2	СЛ	47	3	0,823	0,23	0,628	0,34	$\max(Q^+)$, $\Sigma(Q^+)/\alpha$, $\Sigma(C_a)$	-
3	ОВ	47	3	0,634	0,33	0,475	0,40	$\Sigma(C_a)$, $\Sigma(C_{ad})$, μ	-
4	ЛР	44	4	0,823	0,23	0,777	0,26	α , $\Sigma(Q^+)$, $\Sigma(C_a)$, μ	1, 16, 21
5	СЛ	44	3	0,851	0,19	0,662	0,28	$\max(Q^+)$, $\Sigma(Q^+)/\alpha$, $\Sigma(C_a)$	1, 3, 47
6	ОВ	43	3	0,688	0,26	0,541	0,32	$\Sigma(C_a)$, $\Sigma(C_{ad})$, μ	1, 3, 46, 47

Примечание. * α – молекулярная поляризуемость, Å³; $\Sigma(Q^+)$ – сумма положительных атомных зарядов; $\Sigma(C_a)$ – сумма свободноэнергетических Н-акцепторных дескрипторов; μ – дипольный момент, Д; $\max(Q^+)$ – максимальный положительный атомный заряд; $\Sigma(C_{ad})$ – сумма свободноэнергетических Н-акцепторных и Н-донорных дескрипторов; **1: н-пентан; 3: 1-пропанол; 16: диоксан; 21: сероуглерод; 46: 3-метилпиридин; 47: 1,1,2,2-тетрахлорэтан.

Относительно Y-выбросов можно констатировать, что н-пентан фигурировал в таком качестве во всех моделях. Это могло свидетельствовать о систематическом характере наблюдаемого явления, связанного, например, с несоответствием данного соединения области применимости моделей (учитывая, что н-пентан имеет минимальную токсичность $\log(1/EC_{30}) = -2,94$ и минимальную величину $\Sigma(C_a) = 0.00$ среди исследованных 47 соединений).

Таким образом, в качестве наилучшей была выбрана модель № 4, которая в явном виде выглядит следующим образом:

$$\log(1/EC_{30}) = -3,08(\pm 0,21) + 0,169(\pm 0,016)\alpha - 2,61(\pm 0,23)\Sigma(Q^+) + 0,238(\pm 0,085)\Sigma(C_a) + 0,204(\pm 0,072) \quad (1)$$

(n = 44; r² = 0,823; s = 0,23; q² = 0,777; s_{cv} = 0,26)

ОП уравнения (1): $\log(1/EC_{30})$: $-2,77 \div -0,57$; α : $6,43 \div 15,94 \text{ \AA}^3$;
 $\Sigma(Q^+)$: $0,040 \div 0,87$; $\Sigma(C_a)$: $0,00 \div 2,92$; μ : $0,00 \div 2,76 \text{ Д}$.

Разработанное регрессионное уравнение вполне соответствует требованиям, которые предъявляются в настоящее время к количественным моделям «структура – активность» [16, 17]. В частности, $r^2 > 0,6$ и $q^2 > 0,5$. Входящие в состав уравнения (1) дескрипторы имеют четкий физико-химический смысл. Они отражают способность исследованных соединений к межмолекулярному взаимодействию, которое лежит в основе проявления нейротоксичности. При этом увеличению нейротоксичности способствует рост поляризуемости, Н-акцепторной способности и дипольного момента молекул. И, наоборот, токсичность соединений падает с ростом суммы положительных атомных зарядов.

Как отмечалось выше, для исследуемого ряда органических соединений ранее уже был получен ряд линейных регрессионных уравнений. Так, в работе [7] была разработана модель на основе трех дескрипторов: температуры кипения, молекулярной связности нулевого и второго порядка. Однако при этом были получены достаточно низкие статистические характеристики ($n = 44$; $r^2 = 0,503$; $s = 0,388$). Улучшение статистики ($r^2 = 0,667$; $s = 0,280$) было достигнуто в основном путем значительного уменьшения числа анализируемых соединений ($n = 37$). Гораздо лучшие показатели ($n = 45$; $r^2 = 0,814$; $s = 0,252$; $q^2 = 0,814$; $s_{cv} = 0,273$) были достигнуты на основе уравнения с использованием шести дескрипторов, включая различные спектральные моменты, температуру кипения и число водородных связей [8]. С помощью уравнения (1) удалось получить аналогичные статистические характеристики, которые, однако, были достигнуты с использованием меньшего числа независимых переменных (4 против 6).

Представленные в таблице 2 результаты расчетов нейротоксичности исследованных соединений по отношению к мышам свидетельствуют о том, что метод случайного леса оказался лучшим по сравнению с методами линейной регрессии и опорных векторов в случае полного набора из 47 соединений. Однако после проведения анализа данных на Y-выбросы выявилась конкуренция между СЛ и ЛР. При этом было установлено преимущество по предсказательной способности (при одинаковых величинах s_{cv}) модели №10 ($q^2 = 0,670$) по сравнению с моделью №11 ($q^2 = 0,633$).

В состав полученных моделей входило от двух до четырех дескрипторов. Единственным дескриптором, который фигурировал во всех моделях, была сумма свободноэнергетических протоноакцепторных дескрипторов $\Sigma(C_a)$. Что касается Y-выбросов, то необходимо отметить, что соединение 23 (тетрагидрофуран) встречалось во всех моделях (10)÷(12). При этом нейротоксичность этого соединения ($\log(1/EC_{30}) = -1,08$) была близка к максимальному значению в исследуемом ряду ($\log(1/EC_{30}) = -0,82$; 1,1,2-трихлорэтан). В качестве гипотезы для объяснения этого Y-выброса можно предположить существование отклонения анализируемого соединения от ОП-моделей.

Табл. 2. Статистические характеристики моделей нейротоксичности по отношению к мышам

№	Метод	n	M	r ²	s	q ²	s _{cv}	Дескрипторы*	Y-выбросы**
7	ЛР	47	4	0,595	0,35	0,370	0,44	$\alpha, \Sigma(Q^+), \Sigma(C_a), \mu$	–
8	СЛ	47	2	0,786	0,25	0,536	0,37	$\max(C_a), \Sigma(C_a)$	–
9	ОВ	47	3	0,569	0,36	0,435	0,41	$\max(C_a), \Sigma(C_a), \Sigma(C_{ad})$	–
10	ЛР	45	4	0,741	0,28	0,670	0,32	$\alpha, \Sigma(Q^+), \Sigma(C_a), \mu$	21, 23
11	СЛ	45	2	0,835	0,22	0,633	0,32	$\max(C_a), \Sigma(C_a)$	10, 23
12	ОВ	45	3	0,623	0,31	0,491	0,36	$\max(C_a), \Sigma(C_a), \Sigma(C_{ad})$	1, 23

Примечание. * α — молекулярная поляризуемость, Å^3 ; $\Sigma(Q^+)$ — сумма положительных атомных зарядов; $\Sigma(C_a)$ — сумма свободнoэнергетических Н-акцепторных дескрипторов; μ — дипольный момент, Д; $\max(C_a)$ — максимальный свободнoэнергетический Н-акцепторный дескриптор; $\Sigma(C_{ad})$ — сумма свободнoэнергетических Н-акцепторных и Н-донорных дескрипторов; **1: н-пентан; 10: ацетон; 21: сероуглерод; 23: тетрагидрофуран.

С учетом предсказательной статистики в качестве лучшей была выбрана модель № 10:

$$\log(1/EC_{30}) = -3,32(\pm 0,24) + 0,172(\pm 0,019)\alpha - 2,36(\pm 0,28)\Sigma(Q^+) + 0,394(\pm 0,086)\Sigma(C_a) + 0,131(\pm 0,071)\mu \quad (2)$$

(n = 45; r² = 0,741; s = 0,28; q² = 0,670; s_{cv} = 0,32)

ОП-уравнения (2): $\log(1/EC_{30})$: $-2,98 \div -0,82$; α : $6,43 \div 15,94 \text{ Å}^3$; $\Sigma(Q^+)$: $0,040 \div 0,87$; $\Sigma(C_a)$: $0,00 \div 2,96$; μ : $0,00 \div 2,76 \text{ Д}$.

Полученная модель по своему качественному и количественному составу мало отличается от уравнения (1). Поэтому она обладает всеми теми достоинствами, которые были отмечены ранее: высокими статистическими характеристиками, относительно небольшим числом хорошо интерпретируемых физико-химических параметров. Следует отметить важную роль водородной связи в формировании нейротоксичности исследованных соединений, а именно протоноакцепторную способность $\Sigma(C_a)$, которая встречается в качестве молекулярного дескриптора во всех сконструированных моделях (1)–(12). В ранее проведенном исследовании нейротоксичности 47 органических растворителей [8], также в качестве дескриптора фигурировало число образуемых соединением водородных связей.

Следует подчеркнуть, что одним из желательных требований, предъявляемым к современным QSAR-моделям, является их механистическая интерпретируемость [16, 17]. Токсический эффект является сложным по своей природе явлением. Это связано как с большим числом биологических мишеней в организме, так и с различными типами их взаимодействия с молекулами веществ. Существуют различные схемы классификации химических соединений по типу токсического действия [18]. Учет этих типов и

использование соответствующих дескрипторов приводит к получению наиболее качественных QSAR-моделей.

В отдельную обширную группу можно выделить соединения с неспецифической токсичностью (наркоз). В настоящее время этот эффект рассматривается как результат обратимого межмолекулярного взаимодействия токсикантов с биологическими мембранами клеток [19]. В литературе опубликовано большое количество QSAR-моделей неспецифической токсичности, где в качестве механистического (т. е. отражающего механизм действия) молекулярного дескриптора выступает коэффициент распределения вещества в системе «н-октанол – вода» ($\log P$) [18, 20]. Этот дескриптор был использован и при конструировании регрессионной модели для изучаемого ряда органических растворителей [7]. Ранее [21] нами было показано, что $\log P$ является сложным составным дескриптором, который может быть заменен на комбинацию двух других молекулярных характеристик: α и $\Sigma(C_a)$. Использование этих и других ($\Sigma(Q^+)$ и μ) дескрипторов позволило, с одной стороны, расширить ряд анализируемых соединений по сравнению с ранее опубликованной моделью [7], а с другой стороны – подтвердить неспецифический характер токсичности исследованных органических растворителей и выявить количественный вклад различных типов межмолекулярных взаимодействий.

В качестве основного метода валидации разработанных в настоящей работе регрессионных моделей выступает перекрестный контроль. Другим методом проверки предсказательной способности моделей является валидация на основе прогноза для внешней выборки. Для ее реализации необходимо наличие независимых данных, которые не были использованы в обучающей выборке. В качестве иллюстративного примера мы использовали этиловый спирт, который обладает выраженным нейротоксическим эффектом [22]. Для оценки нейротоксичности использовали модель (4) со следующими характеристиками этанола: $\alpha = 5,08 \text{ \AA}^3$; $\Sigma(Q^+) = 0,32$; $\Sigma(C_a) = 1,66$; $\mu = 1,69$ Д. Полученная величина $\log(1/EC_{30}) = -2,32$ вполне соответствует величинам нейротоксичности других спиртов, например, 1-пропанола ($\log(1/EC_{30}) = -2,71$) и 1-бутанола ($\log(1/EC_{30}) = -2,15$).

Таким образом, в настоящей работе установлено, что при моделировании острой нейротоксичности ряда органических растворителей по отношению к крысам и мышам с помощью трех статистических методов самые лучшие результаты показывает линейная регрессия. Близкие показатели демонстрирует метод случайного леса. Полученные линейные регрессионные уравнения имеют более высокие статистические показатели и/или меньшее число хорошо интерпретируемых физико-химических дескрипторов по сравнению с описанными в литературе моделями. Выявлена важная роль протоакцепторной способности исследованных соединений в формировании их нейротоксичности. Следует отметить, что использованный в настоящей работе простой метод расчета области применимости моделей не гаранти-

рует возможность оценки активности новых тестируемых соединений. Этот вопрос требует дальнейших исследований. Кроме того, такая оценка может быть осуществлена только на основе величин физико-химических дескрипторов, рассчитанных с помощью компьютерной программы НУВОТ. Авторы настоящей работы способны провести необходимые расчеты для интересующихся исследователей. С учетом вышесказанного, разработанные статистические модели могут быть использованы для количественной оценки нейротоксического эффекта новых неисследованных соединений.

ЛИТЕРАТУРА

1. *Миронов, А.Н., Бунатян, Н.Д., Васильев, А.Н.* Руководство по проведению доклинических исследований лекарственных средств, Часть 1. Гриф и К, Москва, 2012. сс. 237–238. ISBN 978-5-9901447-1-2
2. *Makhaeva G.F., Radchenko E.V., Baskin I.I., Palyulin V.A., Richardson R.J., Zefirov N.S.* Combined QSAR studies of inhibitor properties of O-phosphorylated oximes toward serine esterases involved in neurotoxicity, drug metabolism and Alzheimer's disease. 2012. SAR QSAR Environ. Res., 23 (7–8), pp. 627–647. <https://doi.org/10.1080/1062936X.2012.679690>
3. *Mager P.P.* Structure neurotoxicity relationships applied to organophosphorus pesticides. 1982. Toxicol. Lett., 11 (1–2), pp. 67–71. [https://doi.org/10.1016/0378-4274\(82\)90107-2](https://doi.org/10.1016/0378-4274(82)90107-2)
4. *Sanz P., Flores I.C., Soriano T., Repetto G., Repetto M.* In vitro quantitative structure-activity relationship assessment of pyrrole adducts production by γ -diketone-forming neurotoxic solvents. Toxicology in Vitro. 1995. 9 (5), pp. 783–787. [https://doi.org/10.1016/0887-2333\(95\)00017-3](https://doi.org/10.1016/0887-2333(95)00017-3)
5. *Lapenna S., Fuart-Gatnik M. and Worth A.* Review of QSAR Models and Software Tools for predicting Acute and Chronic Systemic Toxicity. EUR – Scientific and Technical Research series. 2010. DOI: 10.2788/60766
6. *Frantik E., Hornychova M. and Horvath M.* Relative Acute Neurotoxicity of Solvents: Isoeffective Air Concentration of 48 Compounds Evaluated in Rats and Mice. 1994. Environ. Res., 66 (2), pp. 173–185. <https://doi.org/10.1006/enrs.1994.1053>
7. *Cronin M.T.D.* Quantitative structure-activity relationship (QSAR) analysis of the acute sublethal neurotoxicity of solvents. 1996. Toxicology in Vitro, 10 (2), pp. 103–110. [https://doi.org/10.1016/0887-2333\(95\)00109-3](https://doi.org/10.1016/0887-2333(95)00109-3)
8. *Estrada E., Molina E. and Uriarte E.* Quantitative Structure-Toxicity Relationships Using Tops-Mode. 2. Neurotoxicity of a Non-Congeneric Series of Solvents. 2001. SAR QSAR Environ. Res., 12 (5), pp. 445–459. <https://dx.doi.org/10.1080/10629360108035384>
9. *Moser V.C.* Functional Assays for Neurotoxicity Testing. 2011. Toxicol. Pathol., 39 (1), pp. 36–45. <https://doi.org/10.1177/0192623310385255>

10. *Raevsky O.A., Grigor'ev V.Yu., Trepalin S.V.* (1999) HYBOT (Hydrogen Bond Thermodynamics). Russian State Patent Department (registration number № 990090).

11. *Raevskii O.A., Trepalina E.P. and Trepalin S.V.* Slipper: A new program for calculating solubility, lipophilicity, and liposome penetrability. 2000. *Pharm. Chem. J.*, 34 (1), pp. 33–36. <https://doi.org/10.1007/BF02524557>

12. *Осипов О.А., Минкин В.И., Гарновский А.Д.* Справочник по дипольным моментам. Москва. 1971. Высшая школа, с. 123–350.

13. *Форсайт Дж., Малькольм М., Моулер К.* Машинные методы математических вычислений. Москва. 1980. Мир, с. 86–248.

14. Program RRforest. Retrieved May 05, 2018 from http://www.stat.berkeley.edu/~breiman/RandomForests/reg_examples/RFR.f

15. R Support Vector Machine application written in fortran 90. Retrieved May 05, 2018 from <https://github.com/jbcolme/fortran-ls-svm>

16. OECD principles for the validation, for regulatory purposes, of (quantitative) structure-activity relationship models. Retrieved May 05, 2018 from <http://www.oecd.org/dataoecd/33/37/37849783.pdf>

17. *Tropsha, A., Gramatica, P. and Gombar, V.K.* The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models. *QSAR Comb. Sci.* 2003. 21 (1), pp. 69–77. <https://doi.org/10.1002/qsar.200390007>

18. *Netzeva, T.I., Pavan, M., and Worth, A.P.* Review of (Quantitative) Structure – Activity Relationships for Acute Aquatic Toxicity. *QSAR Comb. Sci.* 2008. 27 (1), pp. 77–90. <https://doi.org/10.1002/qsar.200710099>

19. *Escher B.I. and Hermens J.L.M.* Modes of Action in Ecotoxicology: Their Role in Body Burdens, Species Sensitivity, QSARs, and Mixture Effects. *Environ. Sci. Technol.* 2002. 36 (20), pp. 4201–4217. <https://doi.org/10.1021/es015848h>

20. *Devillers J. and Devillers H.* Prediction of acute mammalian toxicity from QSARs and interspecies correlations. *SAR QSAR Environ. Res.* 2009. 20 (5–6), pp. 467–500. <https://doi.org/10.1080/10629360903278651>

21. *Raevskii O.A. and Grigor'ev V.Yu.* Lipophilicity of organic compounds: a quantitative description based on the concepts of polarizability and acceptor ability to hydrogen bond formation. *Pharm. Chem. J.* 1999. 33 (5), pp. 274–277. <https://doi.org/10.1007/BF02510052>

22. *Markizova N.F., Grebenyuk A.N., Bašarin V.A., & Bonitenko, E.J.* Spirty: Toksikologija dlja vračej. 2004. Voenno-Medicinskaja Akademije, Sankt-Peterburg.

В.Ю. Григорьев, О.А. Раевский

QSAR-МОДЕЛИРОВАНИЕ БЛОКАДЫ НМДА-РЕЦЕПТОРА ПОЛИФАРМАКОФОРНЫМИ СОЕДИНЕНИЯМИ НА ОСНОВЕ ПРОИЗВОДНЫХ КАРБАЗОЛА И 1-АМИНОАДАМАНТАНА

*Институт физиологически активных веществ РАН,
Черноголовка, Московская область*

Введение

Одним из вызовов современности, который требует быстрого и адекватного ответа, является необходимость разработки медицинских препаратов для лечения нейродегенеративных заболеваний, в частности, болезни Альцгеймера (БА) [1]. В настоящее время выявлен ряд биологических мишеней, связанных с этой болезнью [2]. Одной из них является НМДА-рецептор [3], играющий важную роль в функционировании центральной нервной системы. Одним из перспективных направлений при разработке лекарств для БА является использование соединений с несколькими фармакофорными группами, действующими на несколько биомишеней одновременно [4].

Целью настоящей работы является конструирование регрессионных моделей блокады НМДА-рецептора (сайт связывания $[N^3]$ МК-801) полифармакофорными соединениями на основе производных карбазола и 1-аминоадамантана.

Материалы и методы

Выборка, содержащая информацию о структуре и биологической активности соединений, была извлечена из базы данных [5]. Объектами исследования служили 14 соединений, представляющих собой конъюгаты замещенных карбазолов (тип I) и тетрагидрокарбазолов (тип II) с 1-аминоадамантанами (рисунок 1, таблица 1):

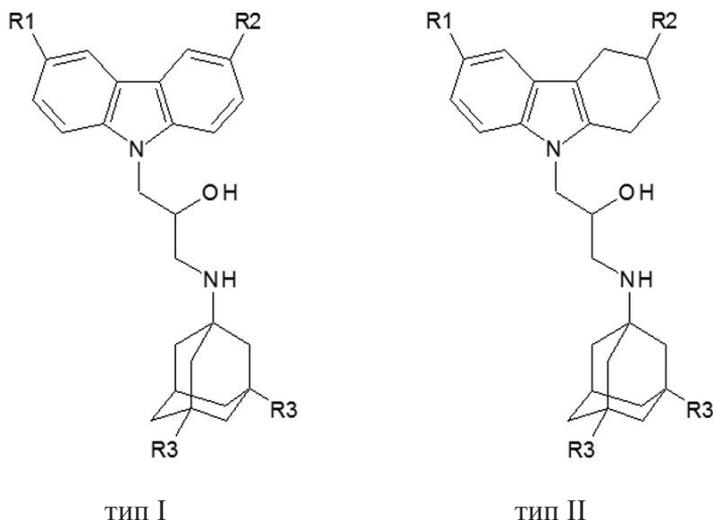


Рис. 1. Структурные формулы полифармакофорных соединений на основе производных карбазолов (тип I), тетрагидрокарбазолов (тип II) и 1-аминоадамантанов

В качестве первоначальной количественной оценки биологической активности использовали величину IC_{50} (мкМ) (концентрация, вызывающая 50% блокаду НМДА-рецептора, сайт связывания $[H^3]$ МК-801). Для QSAR-моделирования эти величины были конвертированы в $-\lg(IC_{50})$. При этом минимальное, максимальное и среднее значение активности в анализируемой выборке составили $-2,217$, $-1,190$ и $-1,929$ соответственно.

Для количественного описания структуры соединений рассчитали несколько сот физико-химических, топологических, электронных дескрипторов с использованием компьютерных программ HYBOT[6] и DRAGON [7]. В качестве следующего шага была применена процедура сокращения дескрипторного пространства на основе анализа парных коэффициентов корреляции (r_{ij}) при критическом значении $r = 0,8$. В результате число дескрипторов было уменьшено до 34. В состав финальных QSAR-моделей вошли всего семь дескрипторов: BAC (центрированный индекс Балабана), SIC4 (комплементарное информационное содержание четвертого порядка), MATS7v (автокорреляция Морана с лагом 7, взвешенная атомными ван-дер-ваальсовыми объемами), MATS8v (автокорреляция Морана с лагом 8, взвешенная атомными ван-дер-ваальсовыми объемами), EEig10x (10-е собственное значение матрицы связности, взвешенное по степеням вершин), VENe4 (4-е высшее собственное значение матрицы Бурдена, взвешенное атомными электроотрицательностями Сандерсона), Psy-80 (индекс Ghose-Viswanadhan-Wendoloski антипсихотического типа при 80%). Таблица 1 содержит данные, которые были использованы для проведения QSAR-моделирования.

Табл. 1. Тип и заместители (рисунок 1), биологическая активность ($-\lg(\text{IC}_{50})$, мкМ) и дескрипторы (D1 ÷ D7) соединений

№	Тип	R ₁	R ₂	R ₃	$-\lg(\text{IC}_{50})$	D1	D2	D3	D4	D5	D6	D7
1	I	Cl	Cl	H	-2.217	10	1.277	-0.304	-0.126	3.000	3.551	0
2	II	CH ₃	H	H	-2.137	5	1.028	-0.344	-0.091	2.574	3.545	0
3	II	F	H	H	-2.114	5	1.001	-0.351	0.297	2.574	3.545	1
4	II	H	H	H	-2.083	2	1.001	-0.375	-0.050	2.537	3.544	1
5	II	CH ₃	H	CH ₃	-2.076	17	0.849	-0.378	-0.140	2.574	3.551	0
6	I	H	H	H	-2.064	2	1.277	-0.375	-0.050	2.817	3.549	1
7	II	CH ₃	CH ₃	H	-2.049	10	1.023	-0.304	-0.126	3.000	3.558	0
8	II	H	H	CH ₃	-2.033	10	0.816	-0.407	-0.106	2.537	3.550	0
9	I	Br	Br	CH ₃	-2.021	26	1.055	0.146	-0.525	3.137	3.556	0
10	II	F	H	CH ₃	-2.021	17	0.816	-0.371	0.254	2.574	3.550	0
11	I	Br	Br	H	-1.926	10	1.277	0.165	-0.516	3.000	3.550	0
12	I	H	H	CH ₃	-1.862	10	1.055	-0.407	-0.106	2.817	3.555	0
13	I	Cl	Cl	CH ₃	-1.215	26	1.055	-0.341	-0.170	3.137	3.557	0
14	II	CH ₃	CH ₃	CH ₃	-1.190	26	0.852	-0.341	-0.170	3.107	3.564	0

Примечание. D1 (BAC); D2 (CIC4); D3 (MATS7v); D4 (MATS8v); D5 (EEig10x); D6 (BENe4); D7 (Psy-80)

QSAR-моделирование проводили с использованием трех методов: линейной регрессии (MLR) [8], случайного леса (RF) [9] и гауссовского процесса (GP) [10]. Для оценки предсказательной способности регрессионных моделей применяли перекрестный контроль с выбором по 5. В качестве статистических характеристик моделей использовали: n – число соединений; m – число дескрипторов; R^2 – квадрат коэффициента линейной корреляции; rmse – среднеквадратичное отклонение; R^2_{cv} – квадрат коэффициента линейной корреляции в условиях перекрестного контроля; rmse_{cv} – среднеквадратичное отклонение в условиях перекрестного контроля; R^2_p – метрика для оценки случайной корреляции [11]. Область применимости (ОП) моделей оценивали на основе анализа интервалов величин дескрипторов.

Результаты и обсуждение

Табл. 2. Статистические характеристики QSAR-моделей блокады НМДА-рецептора

Модель	Дескрипторы	R^2	rmse	R^2_{cv}	rmse_{cv}	R^2_p
MLR	CIC4, EEig10x, Psy-80	0,587	0,20	0,305	0,26	0,493
RF	BAC, MATS8v, BENe4	0,703	0,17	0,301	0,26	0,466
GP	MATS7v, EEig10x	0,950	0,07	0,764	0,15	0,755

В таблице 2 приведены результаты проведенного QSAR-моделирования блокады НМДА-рецептора. Из представленных данных ясно, что регрессионные модели MLR и RF не могут быть использованы для предсказания активности новых соединений из-за своих низких статистических показателей. Но модель, созданная на основе Гауссовского процесса, выглядит вполне адекватной. В состав модели входят два дескриптора, что является вполне приемлемым, так как $(n/m) > 5$. Величина $R^2 > 0,6$, $R_{cv}^2 > 0,5$ и $R_p^2 > 0,5$, что соответствует требованиям, предъявляемым к QSAR-моделям [11, 12]. ОП модели (интервальная оценка): MATS7v: $-0,407 \div 0,165$, EEig10x: $2,537 \div 3,137$. Что касается физико-химического содержания дескрипторов GP-модели, то достаточно трудно дать им однозначное толкование. Можно лишь отметить, что они являются топологическими дескрипторами и легко могут быть рассчитаны для различных молекул.

Таким образом, анализ полученных результатов свидетельствует о том, что способность исследуемых соединений влиять на связывание [НЗ]МК-801 с НМДА-рецептором может быть удовлетворительно описана с помощью регрессионной модели на основе Гауссовского процесса с небольшим числом дескрипторов. Эта статистическая модель может быть использована для количественной оценки биоактивности новых полифармакофорных соединений на основе производных карбазолов и тетрагидрокарбазолов с 1-аминоадамантами.

Благодарности

Работа выполнена в рамках государственного задания на 2018 год (тема № 0090-2017-0020).

ЛИТЕРАТУРА

1. Pangalos M.N., Schechter L.E., & Hurko O. Drug development for CNS disorders: strategies for balancing risk and reducing attrition. *Nat. Rev. Drug Discov.* 2007. 6 (7), pp. 521–532. DOI: 10.1038/nrd2094
2. Silva T., Reis J., Teixeira J., Borges F. Alzheimer's disease, enzyme targets and drug discovery struggles: from natural products to drug prototypes. *Ageing Res. Rev.* 2014. 15, pp. 116–145. DOI: 10.1016/j.arr.2014.03.008
3. Gitto R., Luca L.D., Ferro S.C., Sarro G.D., Costa L., Ciranna L., Chimirri A. Development of 3-substituted-1H-indole derivatives as NR2B/NMDA receptor antagonists. *Bioorg. Med. Chem.* 2009. 17, pp. 1640–1647. DOI: 10.1016/j.bmc.2008.12.058
4. Fang J., Li Y., Liu R., Pang X., Li C., Yang R., He Y., Lian W., Liu A., Du G. Discovery of multi-target-directed ligands against Alzheimer's disease through

systematic prediction of chemical-protein interactions. *J. Chem. Inf. Model.* 2015. 55 (1), pp. 149–164. DOI: 10.1021/ci500574n

5. Раевский О.А., Трепалин С.В., Григорьев В.Ю., Ярков А.В., Бачурин С.О. 2017. Свидетельство о государственной регистрации базы данных «Мультицелевые органические соединения с потенциальным действием на центральную нервную систему» № 2017620020.

6. HYBOT. Retrieved August 24, 2018, from <http://molpro.ipac.ac.ru/hybot.html>

7. DRAGON. Retrieved August 24, 2018, from http://www.taletе.mi.it/products/dragon_projects.htm

8. Forsythe G.E., Malcolm M.A., Moler C.B. *Computer Methods for Mathematical Computations*, Englewood Cliffs. NJ. 1977. Prentice-Hall, pp. 227–235.

9. RF. Retrieved August 24, 2018, from http://www.stat.berkeley.edu/~breiman/RandomForests/cc_examples/prog.f

10. Obrezanova O., Csányi G., Gola J.M.R., Segall M.D. Gaussian processes: a method for automatic QSAR modeling of ADME properties. *J. Chem. Inf. Model.* 2007. 47, pp. 1847–1857. DOI: 10.1021/ci7000633

11. Mitra I., Saha A., Roy K. 2010. Exploring quantitative structure–activity relationship studies of antioxidant phenolic compounds obtained from traditional Chinese medicinal plants. *Mol. Simul.*, 36 (13), pp. 1067–1079. DOI: 10.1080/08927022.2010.503326

12. Tropsha A., Gramatica P., Gombar V.K. 2003. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Comb. Sci.*, 21 (1), pp. 69–77. DOI: 10.1002/qsar.200390007

*В.Ю. Григорьев¹, С.Л. Солодова¹, Д.Е. Полианчик¹,
Дж. С. Дьерден², О.А. Раевский¹*

КОМПЬЮТЕРНЫЕ РЕГРЕССИОННЫЕ МОДЕЛИ Р-ГЛИКОПРОТЕИНОВОГО ТРАНСПОРТА ЛЕКАРСТВЕННЫХ СОЕДИНЕНИЙ

*¹Институт физиологически активных веществ РАН,
Черноголовка, Московская область*

*²School of Pharmacy and Biomolecular Sciences,
Liverpool John Moores University, Liverpool, U.K.*

Из литературных данных [1] известно, что белок Р-гликопротеин (Pgp) выполняет различные функции в организме: барьерную функцию (защита от ксенобиотиков); участие в стероидном обмене; влияние на обмен холестерина; участие в работе иммунной системы; участие в гибели клеток и в клеточной дифференциации. В частности, Р-гликопротеин играет ключевую роль в качестве транспортера в процессах, связанных с преодолением гематоэнцефалического барьера [2]. Известно [3], что Pgp представляет собой белок с молекулярной массой 170 кДа, состоящий из 1280 аминокислотных остатков. При этом взаимодействие низкомолекулярных соединений с Pgp происходит по весьма сложному механизму, который обусловлен тем, что этот белок содержит несколько сайтов связывания, а также тем, что в результате связывания происходит изменение конформации Р-гликопротеина [4].

Трехмерная структура этого белка в настоящее время изучена недостаточно. Так, в работе [5] была представлена пространственная структура Pgp с разрешением $\sim 8 \text{ \AA}$ в комплексе с AMP-PNP, а в публикации [6] – структура Pgp мыши с разрешением около 4 \AA . В последнее время появились данные о структуре Pgp с более высоким разрешением ($3,40 \text{ \AA}$) [7], что открывает новые возможности для молекулярного моделирования взаимодействия этого белка с лигандами.

Следует отметить, что в литературе можно найти много публикаций, обзор которых представлен в работе [8], по QSAR-моделированию Р-гликопротеинового транспорта соединений. Из анализа этих публикаций, в частности, следует, что для конструирования QSAR-моделей использует-

ся большое количество разнообразных дескрипторов, что затрудняет физико-химическую интерпретацию полученных результатов. Кроме того, нужно отметить, что в большинстве случаев для описания способности химических соединений к образованию водородной связи используются полуквантитативные дескрипторы, которые не соответствуют в достаточной мере современным представлениям о роли Н-связи в межмолекулярных взаимодействиях вообще и в субстрат-рецепторных комплексах в частности.

Целью настоящего исследования является создание прогностических регрессионных QSAR-моделей взаимодействия химических соединений и лекарств с Pgp на основе небольшого числа физико-химических дескрипторов с включением в эти модели дескрипторов водородной связи.

Экспериментальная часть

В работе были использованы литературные данные по определению тестовой субстратной специфичности соединений к Р-гликопротеину, основанные на изучении клеточной проницаемости. При этом в качестве основной характеристики были использованы величины ER (Efflux Ratio), которые представляют собой отношение коэффициентов проницаемости соединений в двух противоположных направлениях. В обучающую выборку были включены данные по субстратной специфичности соединений, полученные на основе линии клеток MDCK–MDR1 [2], а в тестовую выборку вошли данные из работы [9], основанные на использовании клеток Caco-2, исключая общие соединения из этих двух публикаций. В результате размер обучающей выборки составил 177 соединений, а в состав тестового ряда вошло 44 соединения.

Для представления молекулярной структуры соединений были рассчитаны 15 физико-химических дескрипторов, заметная роль которых в транспорте веществ в живых организмах была уже ранее установлена [8]. Это молекулярные дескрипторы, связанные в основном с описанием межмолекулярного взаимодействия, размера молекул, липофильности. Для их расчета были использованы компьютерные программы HYBOT [10], MOLTRA [11], DRAGON [12]. Для уменьшения линейной зависимости между переменными был проведен отбор дескрипторов путем анализа корреляционной матрицы с использованием итерационной процедуры, которая состояла из ряда шагов: 1) выбор наиболее информативного дескриптора с максимальной величиной коэффициента вариации; 2) формирование вокруг этого выбранного дескриптора кластера родственных дескрипторов на основе граничной величины коэффициента корреляции (0,95); 3) удаление из матрицы кластера за исключением центрального дескриптора; 4) повторение шагов 1–3 до остановки процедуры. В результате было отобрано семь дескрипторов: ΣQ^- (сумма отрицательных атомных зарядов, HYBOT), ΣC_{ad} (сумма свободноэнергетических Н-акцепторных и Н-донорных дескрипторов,

HYBOT), MW (молекулярная масса, а. е. м., HYBOT), PSA_{ca} (парциальный поверхностный дескриптор суммарной энтальпийной Н-акцепторной способности, Å^2 , MOLTRA), PSA_{ca} (парциальный поверхностный дескриптор суммарной свободноэнергетической Н-акцепторной способности, Å^2 , MOLTRA), PSA_{ed} (парциальный поверхностный дескриптор суммарной энтальпийной Н-донорной способности, Å^2 , MOLTRA), $TPSA_{tot}$ (общая поверхность полярных взаимодействий, Å^2 , DRAGON). Данные об использованных лекарственных соединениях, их активности и свойствах представлены в Приложении.

При конструировании регрессионных моделей использовали три метода: линейная регрессия (ЛР), случайный лес (СЛ), опорные векторы (ОВ) и соответствующие им модифицированные нами компьютерные программы SVD [13], rforest [14] и flssvm [15]. При использовании СЛ параметры метода имели следующие значения: число деревьев ($jbt = 500$), число случайно выбираемых переменных ($mtry = m/3$, где m – общее число дескрипторов в модели). Для создания ОВ-моделей использовали автошкалированные дескрипторы и радиальную базисную функцию с заданными параметрами (без оптимизации).

В качестве статистических характеристик моделей использовали: n – число молекул; m – число дескрипторов; r – коэффициент линейной корреляции, который рассчитывали по формуле: $\Sigma((Y_{exp,i} - \hat{Y}_{exp}) * (Y_{calc,i} - \hat{Y}_{calc})) / ((\Sigma(Y_{exp,i} - \hat{Y}_{exp})^2)^{0,5} * (\Sigma(Y_{calc,i} - \hat{Y}_{calc})^2)^{0,5})$, где $Y_{exp,i}$ – экспериментальное значение активности для i -й молекулы, $Y_{calc,i}$ – рассчитанное значение активности для i -й молекулы, \hat{Y}_{exp} – средняя величина экспериментальных величин активности; \hat{Y}_{calc} – средняя величина рассчитанных величин активности; s – среднеквадратичная ошибка (СКО), которую рассчитывали по формуле: $s = ((\Sigma(Y_{exp,i} - Y_{calc,i})^2/n)^{0,5}$; r_{cv} и s_{cv} – коэффициент корреляции и среднеквадратичная ошибка в условиях перекрестного контроля (CV) с выбором по k -блокам ($k = 5$), соответственно.

Для оценки структурного сходства молекул использовали коэффициенты Танимото [16] T_c , которые рассчитывали с помощью компьютерной программы MOLDIVS [17] на основе структурных фрагментов с применением следующей формулы: $T_c = NAB / (NA + NB - NAB)$, где NA – число структурных фрагментов (ненулевых битов) в молекуле А, NB – число структурных фрагментов (ненулевых битов) в молекуле В, NAB – число общих структурных фрагментов в молекулах А и В.

Учитывая относительно небольшое число отобранных дескрипторов, для расчета моделей был использован метод полного перебора всех возможных комбинаций дескрипторов. В качестве критерия для отбора наиболее значимых моделей использовали среднеквадратичную ошибку. При этом применяли два подхода. В первом случае отбирали модели с минимальным значением s_{cv} . Во втором случае выбор осуществляли на основе минимальных величин s тестовой выборки.

Для улучшения качества регрессионных моделей проводился анализ Y-выбросов. При этом из выборки удалялись соединения, у которых абсолютная разность между экспериментальными и рассчитанными величинами превышала $3s$.

Результаты и обсуждение

Необходимость использования для отбора лучших регрессионных моделей двух вышеназванных подходов обусловлена тем, что среди специалистов уже достаточно длительное время ведется дискуссия о предпочтительности внутренней или внешней валидации для оценки предсказательной способности моделей [18]. Результаты, полученные с использованием внутренней валидации, представлены в таблице 1.

Табл. 1. Статистические характеристики регрессионных моделей, отобранных на основе величин СКО (s_{cv}) обучающего ряда

№	Метод	Дескрипторы	Обуч. ряд		Тест. ряд		Y-выбросы**
			r^2_{cv}	s_{cv}	r^2	s	
1	ЛР	ΣC_{ad}	0,476	0,39	0,529	0,47	65,126,152
2	ЛР	$\Sigma C_{ad}, PSA_{ed}$	0,492	0,38	0,519	0,47	65,126,152
3	ЛР	$\Sigma Q^-, \Sigma C_{ad}, TPSA_{tot}$	0,510	0,37	0,464	0,55	65,126,152
4	ЛР	$\Sigma Q^-, \Sigma C_{ad}, PSA_{ca}, PSA_{ed}$	0,517	0,37	0,438	0,59	65,126,152
5	ЛР	$\Sigma Q^-, \Sigma C_{ad}, PSA_{ca}, PSA_{ed}, TPSA_{tot}$	0,505	0,38	0,378	0,63	65,152
6*	ЛР	$\Sigma Q^-, \Sigma C_{ad}, MW, PSA_{ca}, PSA_{ca}, PSA_{ed}, TPSA_{tot}$	–	–	–	–	–
7	СЛ	$TPSA_{tot}$	0,340	0,43	0,270	0,53	65,152,157
8	СЛ	$MW, TPSA_{tot}$	0,430	0,40	0,664	0,39	65,152
9	СЛ	$MW, PSA_{ed}, TPSA_{tot}$	0,443	0,39	0,675	0,42	65,152,167
10	СЛ	$\Sigma C_{ad}, MW, PSA_{ca}, TPSA_{tot}$	0,463	0,39	0,658	0,40	65,152,167
11	СЛ	$\Sigma Q^-, \Sigma C_{ad}, MW, PSA_{ca}, TPSA_{tot}$	0,451	0,40	0,624	0,42	65,167
12	СЛ	$\Sigma Q^-, \Sigma C_{ad}, MW, PSA_{ca}, PSA_{ca}, PSA_{ed}, TPSA_{tot}$	0,460	0,40	0,657	0,42	65
13	ОВ	ΣC_{ad}	0,463	0,39	0,595	0,43	65,126,152
14	ОВ	$\Sigma C_{ad}, MW$	0,488	0,38	0,567	0,43	65,152
15	ОВ	$\Sigma C_{ad}, MW, TPSA_{tot}$	0,508	0,38	0,472	0,47	65,167

№	Метод	Дескрипторы	Обуч. ряд		Тест. ряд		Y-выбросы**
			r^2_{cv}	s_{cv}	r^2	s	
16	ОВ	ΣQ^- , ΣC_{ad} , MW, $TPSA_{tot}$	0,453	0,40	0,457	0,47	65,174
17	ОВ	ΣQ^- , ΣC_{ad} , MW, PSA_{ca} , $TPSA_{tot}$	0,421	0,41	0,430	0,49	65,174
18	ОВ	ΣQ^- , ΣC_{ad} , MW, PSA_{ca} , PSA_{ca} , PSA_{ed} , $TPSA_{tot}$	0,322	0,44	0,351	0,52	65,174

Примечание. * Сингулярная матрица дескрипторов; ** соединения обучающего ряда: 65 – Rhodamine 123, 126 – Methotrexate, 152 – Hoechst 33342, 157 – Paclitaxel, 167 – Eletriptan, 174 – Saquinavir.

Как видно из представленных данных, при использовании ЛР среди одиночных дескрипторов наибольшим эффектом обладает ΣC_{ad} , т. е. дескриптор, который характеризует суммарную H-акцепторную и H-донорную способность молекул. Использование этого дескриптора позволило построить модель (1) со следующими статистическими характеристиками ($r^2_{cv} = 0,476$, $s_{cv} = 0,39$). Дальнейшее увеличение размера дескрипторного пространства немного улучшало статистику. Максимальные величины были получены для модели (4) из четырех дескрипторов ($r^2_{cv} = 0,517$, $s_{cv} = 0,37$). Дальнейшее увеличение числа дескрипторов ухудшало статистические показатели.

Похожее общее поведение наблюдалось при последовательном конструировании моделей на основе СЛ. Однако при этом ведущее положение в качестве главного регрессионного фактора занимал топологический дескриптор $TPSA_{tot}$. Максимальный эффект наступал при числе дескрипторов, равном 4 (модель 10, $r^2_{cv} = 0,463$, $s_{cv} = 0,39$). Как видно из представленных данных, статистические характеристики ЛР- и СЛ-моделей достаточно близки между собой, однако оптимальные комбинации дескрипторов при этом отличались.

При создании регрессионных ОВ-моделей наиболее «весомым» среди одиночных дескрипторов, так же как и в случае ЛР-моделей, оказался дескриптор ΣC_{ad} (модель 13, $r^2_{cv} = 0,463$, $s_{cv} = 0,39$). Наилучшая модель (15) была получена на основе трех дескрипторов: ΣC_{ad} , MW, $TPSA_{tot}$. По своим статистическим характеристикам ($r^2_{cv} = 0,508$, $s_{cv} = 0,38$) она вполне соответствует оптимальным моделям ЛР (4) и СЛ (10).

Интересно отметить разное поведение статистических показателей внешней тестовой выборки в зависимости от используемого метода и размерности пространства. Так, в случае ЛР увеличение числа дескрипторов в модели от 1 до 5 сопровождалось увеличением СКО от 0,47 до 0,63. Аналогичное поведение демонстрируют ОВ-модели: $s = 0,43$ (модель 13) и $s = 0,52$

(модель 18). Напротив, при использовании СЛ-регрессии максимальное значение ($s = 0,53$) наблюдается для модели (7) с одним дескриптором. При дальнейшем увеличении числа дескрипторов величины СКО уменьшались до $0,39 \div 0,42$.

Другой метод выбора лучших регрессионных моделей был основан на использовании минимальных величин s тестовой выборки (таблица 2).

Табл. 2. Статистические характеристики регрессионных моделей, отобранных на основе величин СКО (s) тестового ряда

№	Метод	Дескрипторы	Обуч. ряд		Тест. ряд		Y-выбросы**
			r^2_{cv}	s_{cv}	r^2	s	
19	ЛР	PSA_{ca}	0,263	0,48	0,526	0,38	25
20	ЛР	$PSA_{ca}, TPSA_{tot}$	0,264	0,48	0,585	0,39	25
21	ЛР	$PSA_{ca}, PSA_{cd}, TPSA_{tot}$	0,354	0,45	0,512	0,44	–
22	ЛР	$PSA_{ca}, PSA_{ca}, PSA_{cd}, TPSA_{tot}$	0,345	0,45	0,503	0,45	–
23	ЛР	$\Sigma Q^-, PSA_{ca}, PSA_{ca}, PSA_{cd}, TPSA_{tot}$	0,375	0,44	0,493	0,54	–
24*	ЛР	$\Sigma Q^-, \Sigma C_{ad}, MW, PSA_{ca}, PSA_{ca}, PSA_{cd}, TPSA_{tot}$	–	–	–	–	–
25	СЛ	PSA_{ca}	0,222	0,50	0,477	0,37	25
26	СЛ	$PSA_{ca}, TPSA_{tot}$	0,301	0,47	0,675	0,37	–
27	СЛ	$PSA_{ca}, PSA_{ca}, TPSA_{tot}$	0,333	0,46	0,663	0,34	25
28	СЛ	$\Sigma Q^-, MW, PSA_{ca}, TPSA_{tot}$	0,332	0,46	0,709	0,38	–
29	СЛ	$Q^-, MW, PSA_{ca}, PSA_{ca}, TPSA_{tot}$	0,388	0,44	0,680	0,40	–
30	СЛ	$\Sigma Q^-, \Sigma C_{ad}, MW, PSA_{ca}, PSA_{ca}, PSA_{cd}, TPSA_{tot}$	0,381	0,44	0,613	0,44	–
31	ОВ	PSA_{ca}	0,251	0,48	0,604	0,35	25
32	ОВ	$PSA_{ca}, TPSA_{tot}$	0,259	0,48	0,692	0,38	–
33	ОВ	$PSA_{ca}, PSA_{ca}, TPSA_{tot}$	0,319	0,46	0,613	0,41	–
34	ОВ	$\Sigma Q^-, \Sigma C_{ad}, MW, PSA_{cd}$	0,409	0,43	0,548	0,45	–
35	ОВ	$\Sigma Q^-, MW, PSA_{ca}, PSA_{ca}, TPSA_{tot}$	0,327	0,46	0,539	0,47	–
36	ОВ	$\Sigma Q^-, \Sigma C_{ad}, MW, PSA_{ca}, PSA_{ca}, PSA_{cd}, TPSA_{tot}$	0,356	0,46	0,357	0,53	–

Примечание. * Сингулярная матрица дескрипторов; ** соединение тестового ряда: 25 – Vinblastine.

При использовании регрессионных методов можно отметить ряд общих моментов. В качестве основного одиночного ведущего дескриптора выступал PSA_{Aea} , т. е. дескриптор, связанный с описанием способности соединений к образованию водородной связи. Применение этого дескриптора приводило к получению моделей (19, 25, 31) с достаточно низкими величинами СКО: 0,38, 0,37 и 0,35 соответственно. Дальнейшее увеличение числа дескрипторов в случае ЛР- и ОБ-моделей ухудшало статистические характеристики. Однако при использовании СЛ лучшая регрессионная модель (27) была получена на основе трех дескрипторов: PSA_{ca} , PSA_{ca} и $TPSA_{tot}$ ($r^2 = 0,663$, $s = 0,34$).

Необходимо отметить существование некоторых тенденций при сопоставлении величин s и scv . Увеличение числа дескрипторов в моделях (19) ÷ (36) в целом сопровождается увеличением значений s и уменьшением значений scv . В моделях (1) ÷ (18) эта тенденция менее выражена. В первом случае коэффициент корреляции $r = -0,57$, а во втором случае $r = -0,10$.

Изменение критерия, положенного в основу отбора моделей, сильно влияет на состав оптимального дескрипторного пространства. Так, лучшая ОБ-модель (15) сконструирована с использованием трех дескрипторов ΣC_{ad} , MW , $TPSA_{tot}$. Соответствующая ОБ-модель (31) создана на основе совсем другого дескриптора PSA_{ca} . При переходе от одной модели к другой также сильно меняются статистические характеристики. Например, в модели (15) ($r_{cv}^2 = 0,508$, $s_{cv} = 0,38$, $r^2 = 0,472$, $s = 0,47$), а в модели (31) ($r_{cv}^2 = 0,251$, $s_{cv} = 0,48$, $r^2 = 0,604$, $s = 0,35$).

Необходимым элементом при проведении QSAR-моделирования является исследование области применимости созданных моделей. В настоящей работе для этого был использован метод, основанный на анализе близости молекул тестового ряда по отношению к молекулам обучающего ряда [19]. В качестве меры структурного сходства применяли коэффициенты Танимото. Для отображения полученных данных использовали интегральную функцию распределения T_c . (рисунок 1).

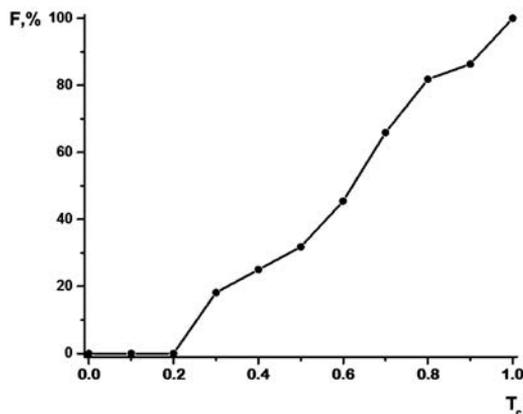


Рис. 1. Интегральная функция распределения коэффициентов Танимото (T_c) между молекулами тестового ряда и ближайшими соседями обучающего ряда

Анализ этой функции позволяет определить долю молекул тестовой выборки, которые соответствуют выбранному порогу молекулярного сходства. Например, из рисунка 1 можно сделать вывод о том, что в тестовой выборке отсутствуют молекулы с $T_c < 0,2$, или о том, что доля молекул с $T_c > 0,5$ составляет 68% ($100\% (T_c = 1,0) - 32\% (T_c = 0,5) = 68\%$).

Ранее в литературе уже были опубликованы результаты бинарной классификации 187 соединений (177 из которых использованы в настоящей работе в качестве обучающего ряда) [2]. При этом с применением методов СЛ, ОВ и к-ближайшего соседа были созданы классификационные модели на основе различных дескрипторов. Лучшие модели были получены с использованием трех или четырех трехмерных молекулярных дескрипторов. Разработанные в настоящей работе регрессионные модели отличаются тем, что при сопоставимом числе использованные дескрипторы являются двумерными, что снимает ряд вопросов, связанных с моделированием структуры молекул.

Таким образом, с использованием методов линейной регрессии, случайного леса и опорных векторов на основе различных комбинаций, содержащих от одного до семи молекулярных дескрипторов с ясной физико-химической интерпретацией, созданы регрессионные модели, имеющие хорошие статистические характеристики. При этом для отбора статистически значимых QSAR-моделей применено два критерия: СКО в условиях перекрестного контроля и СКО тестовой выборки. В результате установлено, что оптимальный набор дескрипторного пространства сильно зависит от используемого критерия. Практически во всех моделях присутствуют дескрипторы водородной связи, что подтверждает ее важную роль в процессах P_{gp} транспорта лекарственных соединений через гематоэнцефалический барьер. Разработанные статистические модели могут быть использованы для оценки Р-гликопротеинового транспорта новых неисследованных соединений.

ЛИТЕРАТУРА

1. Раменская Г.В., Скуридина Е.А., Сычев Д.А., Кукес В.Г. Клиническая фармакокинетика. 2005. 2 (3), с. 18–23.
2. Broccatelli F. J. Chem. Inf. Model. 2012. 52, pp. 2462–2470.
3. Li Y., Yuan H, Yang K. et al. Curr. Med. Chem. 2010. 17 (8), pp. 786–800.
4. Higgins C.F. and Linton K.J. Nat. Struct. Mol. Biol. 2004. 11 (10), pp. 918–926.
5. Rosenberg M.F., Callaghan R., Modok S. et al. J. Biol. Chem. 2005. 280, pp. 2857–2862.
6. Aller S.G., Yu J., Ward A. et al. Science. 2009. 323, pp. 1718–1722.
7. Szweczyk P., Tao H., McGrath A.P. et al. Acta Cryst. 2015. D71, pp. 732–741.

8. *Раевский О.А., Солодова С.Л., Лагунин А.А., Поройков В.В.* Биомедицинская химия. 2014. 60 (2), с. 161–181.

9. *Crivori P., Reinach B., Pezzetta D. et al.* Molecular Pharmaceutics. 2006. 3, pp. 33–44.

10. *Раевский О.А., Григорьев В.Ю., Трепалин С.В.* (1999) HYBOT (Hydrogen Bond Thermodynamics). Свидетельство об официальной регистрации программы для ЭВМ № 990090, Москва, Федеральная служба по интеллектуальной собственности.

11. *Раевский О.А., Трепалин С.В., Раздольский А.Н.* (1999) MOLTRA (Molecular Transform Analysis). Свидетельство об официальной регистрации программы для ЭВМ № 990092, Москва, Федеральная служба по интеллектуальной собственности.

12. <http://www.taletе.mi.it>

13. *Форсайт Дж., Малькольм М., Моулер К.* Машинные методы математических вычислений. Москва, Мир, 1980, с. 248–256.

14. http://www.stat.berkeley.edu/~breiman/RandomForests/reg_examples/RFR.f

15. <https://github.com/jbcolme/fortran-ls-svm>

16. *Rogers D.J. and Tanimoto T.T.* Science. 1960. 132 (3434), 1115–1118.

17. *Gerasimenko V.A., Trepalin S.V. and Raevsky O.A.* In: Molecular Modeling and Prediction of Bioactivity, K. Gundertofte and F.S. Jorgensen (eds.), Kluwer Academic/Plenum Publishers. New York. 2000. pp. 423–424.

18. *Veerasamy R., Rajak H., Jain A. et al.* Int. J. Drug Des. Discov. 2011. 2 (3), pp. 511–519.

19. *R.P. Sheridan, B.P. Feuston, V.N. Maiorov, and S.K. Kearsley.* J. Chem. Inf. Comput. Sci. 2004. 44, pp. 1912–1928.

*В.Ю. Григорьев¹, К.А. Щербаков², Д.Е. Полианчик¹, А.Н. Раздольский¹,
А.В. Веселовский², В.В. Григорьев¹, А.В. Ярков¹, О.А. Раевский¹*

ИССЛЕДОВАНИЕ КАНАЛЬНЫХ БЛОКАТОРОВ NMDA-РЕЦЕПТОРА В РЯДУ КОНЬЮГАТОВ МЕТИЛЕНОВОГО СИНЕГО С ИСПОЛЬЗОВАНИЕМ QSAR И МОЛЕКУЛЯРНОГО МОДЕЛИРОВАНИЯ

*¹Институт физиологически активных веществ РАН,
Черноголовка, Московская область*

*²Научно-исследовательский институт биомедицинской химии
имени В.Н. Ореховича, Москва*

Введение

В настоящее время в мире различные формы деменции диагностированы у 30 миллионов человек, и опубликованы прогнозы о том, что их число может достичь 100 миллионов к 2050 году [1], при этом экономические потери в 2018 году могут превысить 1 триллион долларов [2]. Поэтому одной из важных задач в медицинской химии является создание новых препаратов для лечения нейродегенеративных заболеваний, в частности болезни Альцгеймера (AD). При этом помимо общих проблем, связанных с разработкой новых лекарств (оптимизация связывания молекул с биомишенями и ADMET-свойств), возникают специфические проблемы. В первую очередь, это обусловлено сложным характером патогенеза AD. Учитывая мультифакторную природу этой болезни, весьма перспективным выглядит поиск новых соединений, действующих одновременно на несколько биомишеней [3–5]. Важное место среди этих мишеней занимают ионотропные глутаматные NMDA-рецепторы. В качестве перспективных подходов для исследований в этой области можно рассматривать компьютерные технологии [6]. В частности, опубликован ряд работ, связанных с изучением связывания лигандов с NMDA-рецептором с использованием QSAR и молекулярного моделирования [7–9]. В литературе имеются сведения о нейропротекторной активности димебона и его аналогов [10], а также метиленового синего [11]. Целью настоящего исследования яв-

ляется создание на основе QSAR и молекулярного докинга моделей, описывающих связывание бинарных структур, состоящих из этих фармакофоров, с NMDA-рецептором (сайт связывания МК-801).

Материалы и методы

Сведения о структуре и биологической активности соединений были взяты из базы данных [12].

В работе были исследованы 29 конъюгатов восстановленной формы метиленового синего (МС) и четырех типов фармакофоров, включая производные карбазола (I), тетрагидрокарбазола (II), замещенных индолов (III) и γ -карболина (IV), объединенных 1-оксопропиленовым спейсером (рисунок 1).

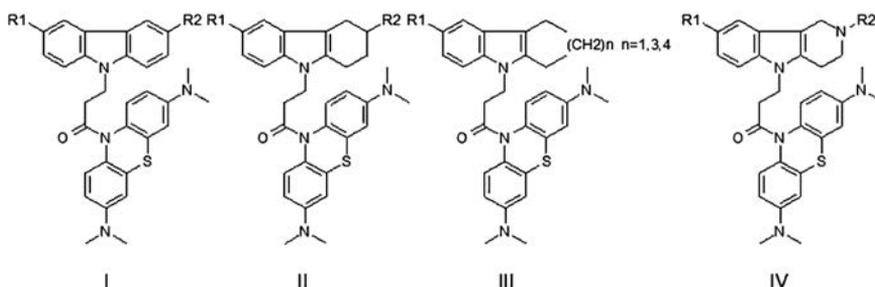


Рис. 1. Структурные формулы конъюгатов метиленового синего

В качестве меры биологической активности соединений служила величина IC_{50} (мкМ), которая рассчитывалась с использованием радиолигандного метода на основе изучения влияния соединений на связывание $[H^3]$ МК-801 с NMDA-рецепторами, полученными из мембран гиппокампа крыс [13].

Для описания структуры исследуемых соединений была использована компьютерная программа HUBOT, в основе которой лежит расчет количественных характеристик водородной связи [14]. В состав конечных регрессионных моделей вошло суммарно четыре дескриптора, включая: α (молекулярная поляризуемость), $\Sigma(C_a)$ (сумма свободноэнергетических Н-акцепторных дескрипторов), $\Sigma(C_a)/\alpha$ (композиционный дескриптор), E_a^{max} (максимальный энтальпийный Н-акцепторный дескриптор). Данные, использованные для расчетов, приведены в таблице 1.

Табл. 1. Формулы, биологическая активность (IC_{50} , мкМ) и дескрипторы (α (\AA^3), $\Sigma(C_a)$, $\Sigma(C_a)/\alpha$, E_a^{max}) исследуемых соединений

№	Тип	R ₁	R ₂	A	α	$\Sigma(C_a)$	$\Sigma(C_a)/\alpha$	E_a^{max}
1	I	Cl	Cl	-2,258	65,51	8,14	0,124	2,16
2	I	Br	Br	-2,243	66,90	8,21	0,123	2,16

Окончание таблицы 1

№	Type	R ₁	R ₂	A	α	Σ(C _a)	Σ(C _a)/α	E _a ^{max}
3	I	CH ₃	CH ₃	-2,220	65,32	7,80	0,119	2,16
4*	III, n=3	CH ₃	—	-2,212	64,62	7,33	0,113	2,16
5	III, n=4	CH ₃	—	-2,201	66,45	7,33	0,110	2,16
6	II	OCF ₃	CH ₃	-2,179	64,62	7,70	0,119	2,16
7	I	H	H	-2,173	61,65	7,92	0,128	2,16
8*	III, n=1	CH ₃	—	-2,158	60,95	7,33	0,120	2,16
9	II	OCF ₃	H	-2,137	62,78	7,70	0,123	2,16
10	II	OCH ₃	CH ₃	-2,117	64,89	8,31	0,128	2,16
11	II	F	H	-2,104	60,85	7,53	0,124	2,16
12	II	CH ₃	H	-2,104	62,78	7,33	0,117	2,16
13*	II	CF ₃	CH ₃	-2,104	64,34	7,36	0,114	2,16
14*	III, n=3	OCF ₃	—	-2,097	64,62	7,70	0,119	2,16
15*	III, n=1	H	—	-2,086	59,11	7,41	0,125	2,16
16*	II	H	H	-2,083	60,95	7,41	0,122	2,16
17*	II	CH ₃	CH ₃	-2,079	64,62	7,33	0,113	2,16
18*	II	OCH ₃	H	-2,079	63,05	8,31	0,132	2,16
19*	III, n=4	H	—	-2,076	64,62	7,41	0,115	2,16
20*	II	F	CH ₃	-2,072	62,69	7,53	0,120	2,16
21*	II	H	CH ₃	-2,057	62,78	7,41	0,118	2,16
22*	III, n=3	H	—	-2,053	62,78	7,41	0,118	2,16
23	IV	CH ₃	CH(CH ₃) ₂	-1,834	67,80	9,54	0,141	2,16
24*	IV	F	CH ₃	-1,798	62,21	9,75	0,157	2,37
25	IV	CH ₃	(CH ₂) ₂ CH ₃	-1,674	67,80	9,31	0,137	2,21
26	IV	H	CH ₂ CH ₃	-1,627	64,13	9,38	0,146	2,21
27*	IV	H	CH ₃	-1,627	62,30	9,63	0,155	2,37
28	IV	CH ₃	CH ₂ CH ₃	-1,560	65,97	9,31	0,141	2,21
29	IV	CH ₃	CH ₃	-1,303	64,13	9,55	0,149	2,37

Примечание: * Гидрохлориды.

При создании QSAR-моделей биологической активности использовали четыре метода: линейную регрессию (MLR), случайный лес (RF), опорные

векторы (SVM) и гауссовский процесс (GP). Оценку регрессионных коэффициентов и определение степени коллинеарности переменных проводили на основе компьютерной программы SVD [15]. Для анализа деревьев решений в методе RF использовали оригинальную программу автора метода Бреймана [16] с исходными параметрами и числом деревьев, равным 500. В качестве машины опорных векторов применяли алгоритм *fssvm* [17], особенностью которого является использование метода наименьших квадратов. Конструирование GP-моделей выполняли с применением алгоритма, описанного в работе [18].

Оценку предсказательной способности QSAR-моделей проводили на основе внутреннего (кросс-валидация) и внешнего тестирования. Кросс-валидацию осуществляли путем разбиения анализируемой выборки на пять равных частей, используя четыре части для конструирования модели и одну часть для ее тестирования. Тестовую выборку формировали благодаря упорядочиванию исследуемых соединений по активности и выбору каждого 5 соединения. Таким образом, была создана обучающая выборка (24 соединения) и тестовая выборка (5 соединений).

Для описания статистических свойств регрессионных моделей использовали: r^2 – квадрат коэффициента линейной корреляции (обучающая выборка); r_{mse} – среднеквадратичное отклонение (обучающая выборка); r_{cv}^2 – квадрат коэффициента линейной корреляции в условиях кросс-валидации; $r_{mse_{cv}}$ – среднеквадратичное отклонение в условиях кросс-валидации; r_{test}^2 – квадрат коэффициента линейной корреляции (тестовая выборка); $r_{mse_{test}}$ – среднеквадратичное отклонение (тестовая выборка); r_p^2 – метрика для оценки случайной корреляции, которую вычисляли по формуле [19]: $r_p^2 = r^2 * (r^2 - r_{rand}^2) / 0,5$, где r_{rand}^2 – квадрат рандомизированного коэффициента линейной корреляции.

Молекулярный докинг

Докинг соединений был проведен в структуру NMDA-рецептора (NMDAR) *XenopusLaevis* (код PDB 5uow). Данный рецептор состоит из четырех субъединиц: 2 GluN1 (цепи A, C), GluN2A (цепь B) и GluN2B (цепь D). Несмотря на то, что измерение активности исследованных соединений было проведено на NMDAR крысы, докинг был проведен в NMDAR лягушки по следующим причинам:

1) пространственная структура NMDAR *XenopusLaevis* была получена в комплексе с МК-801;

2) выравнивание аминокислотных последовательностей участков, образующих пору ионного канала, показало высокую степень их идентичности, что позволяет использовать NMDAR *X. Laevis* вместо *R. Norvegicus*.

Первоначально в модели 5uow были достроены недостающие аминокислотные остатки с использованием пакета SYBYL 8.1. Структура полученного комплекса была оптимизирована с использованием процедуры минимизации энергии.

Докинг осуществляли в пакете VinaAutoDock [20]. Подготовка белка и лигандов – в пакете AutoDockTools [21]. Докинг производился направленно в место связывания МК-801 в боксе размером 20x20x40Å. Для каждого лиганда были получены девять вероятных поз докинга.

Результаты и обсуждение

В таблицах 2, 3 представлены результаты QSAR-моделирования блокады NMDA-рецептора (сайт связывания МК-801). Линейная модель (MLR) показывает худшие результаты по сравнению с нелинейными моделями (RF, SVM, GP). При этом нелинейные QSAR-модели удовлетворяют минимальным требованиям, которые к ним предъявляются [22], в частности, $r^2 > 0,6$; $r_{cv}^2 > 0,5$; $r_{test}^2 > 0,5$; $r_p^2 > 0,5$.

Табл. 2. Статистические характеристики QSAR моделей блокады NMDA рецептора

Model	Descriptors	r^2	r_{mse}	r_{cv}^2	$rmse_{cv}$	r_{test}^2	$rmse_{test}$	r_p^2
MLR	$\Sigma(C_a)/\alpha$	0,708	0,13	0,627	0,14	0,377	0,18	0,656
RF	$\Sigma(C_a), E_a^{max}$	0,924	0,06	0,837	0,09	0,684	0,13	0,688
SVM	$\Sigma(C_a)$	0,990	0,02	0,858	0,09	0,751	0,12	0,699
GP	$\alpha, \Sigma(C_a), E_a^{max}$	0,971	0,04	0,748	0,12	0,815	0,10	0,747

Табл. 3. Номера, экспериментальные (A_{exp}), рассчитанные (A_{calc}) величины активности и их разность ($\Delta = A_{exp} - A_{calc}$) для соединений тестовой выборки

№	A_{exp}	MLR		RF		SVM		GP	
		A_{calc}	Δ	A_{calc}	Δ	A_{calc}	Δ	A_{calc}	Δ
6	-2,179	-2,128	-0,051	-2,112	-0,067	-2,127	-0,052	-2,157	-0,022
12	-2,104	-2,161	0,057	-2,152	0,048	-2,122	0,018	-2,071	-0,033
18	-2,079	-1,911	-0,168	-2,189	0,110	-2,199	0,120	-2,119	0,040
24	-1,798	-1,494	-0,304	-1,558	-0,240	-1,598	-0,200	-1,595	-0,203
28	-1,560	-1,761	0,201	-1,647	0,087	-1,657	0,097	-1,632	-0,129

Примечание. $A = -\log(IC_{50})$, IC_{50} мкМ.

Об удовлетворительной предсказательной способности анализируемых моделей свидетельствуют данные таблицы 3. Во всех случаях, за исключением соединения 24, разность между экспериментальными и рассчитанными величинами активности не превышает двух стандартных отклонений по отношению к $rmse_{cv}$. В случае соединения 24 наблюдаются систематические

отклонения величин Δ . Это может быть связано с тем, что в анализируемой выборке соединений присутствуют соединения как в нейтральной форме, так и в солевой форме, что не принимается во внимание при расчете дескрипторов.

Следует отметить, что современные QSAR-модели должны не только иметь хорошие статистические характеристики, но и обладать (по возможности) способностью к физико-химической интерпретации [23]. Из данных таблицы 2 следует, что в качестве независимых переменных фигурируют дескрипторы водородной связи. При этом наиболее значимой величиной является Н-акцепторная способность исследуемых соединений. Полученный результат хорошо согласуется с имеющимися представлениями о том, что Н-связь играет важную роль при блокаде ионных каналов NMDA-рецептора [24]. Сконструированные нами статистические модели содержат минимальное число дескрипторов, в то время как представленные в литературе QSAR-модели, описывающие блокаду NMDA-рецептора на основе количественных данных по вытеснению [3H]МК-801 из его сайта связывания, включают в свой состав 10 и более дескрипторов [8, 25] с неясной физико-химической интерпретацией.

Анализ полученных результатов свидетельствует о том, что между активностью исследуемых соединений и их структурой обнаруживается простая линейная тенденция, которая заключается в том, что с увеличением протоноакцепторной способности соединений также растет их способность к связыванию с NMDA-рецептором. Это становится очевидным при сопоставлении величин активности соединений I, II, и III типа с соединениями IV типа: появление в структуре дополнительного атома азота приводит к увеличению Н-акцепторной способности и, как следствие, ведет к увеличению активности.

Для исследования механизма связывания этих соединений с NMDAR был проведен их докинг в ионный канал рецептора. Результаты докинга показали, что исследуемые соединения располагаются в ионном канале NMDAR в месте связывания известного антагониста МК-801 (дизоцилпин). Исследуемые лиганды взаимодействуют с ионным каналом в основном за счет гидрофобных взаимодействий. В связывании также принимают участие водородные связи, вероятнее всего, они необходимы для специфического узнавания лигандов.

Основной водородной связью, наблюдаемой у всех лигандов, является связь между карбонильным кислородом лиганда и ОН-группой Thr646 А-цепи GluN1 (рисунок 2). Это отличает их от ранее известных антагонистов, связывающихся с Asn614 субъединицы GluN1, Asn602/Asn603 субъединицы GluN2A и Asn608/Asn609 субъединицы GluN2B [26].

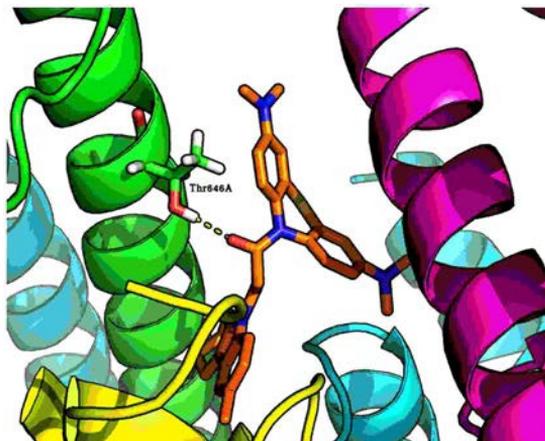


Рис. 2. Водородная связь между карбонильной группой лиганда и боковой цепью Thr646A-GluN1 на примере соединения 17

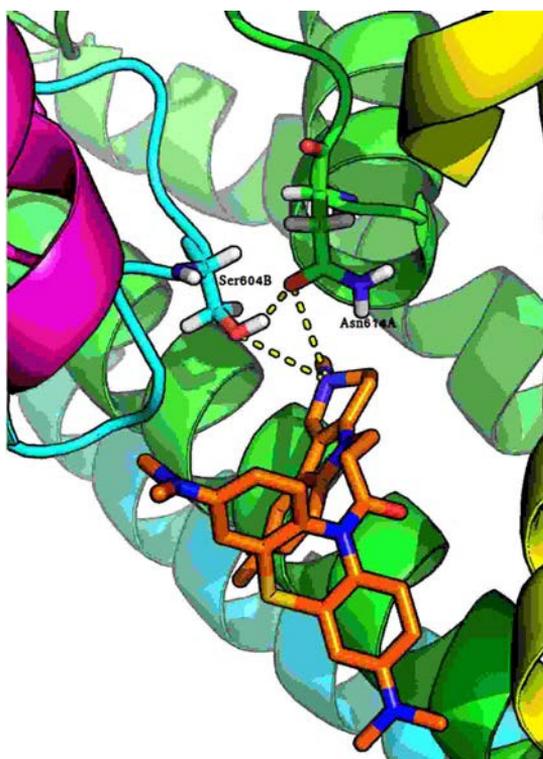


Рис. 3. Водородная связь между соединением 29 и боковой цепью Asn609D-GluN2B

Для лигандов, содержащих пять атомов азота (соединения 23–29), наблюдали дополнительный паттерн связывания. Дополнительный атом азота

образовывал две водородные связи: с гидроксилом Ser604B-GluN2A (рисунок 3) и NH₂-группой Asn614A-GluN1. Наличие двух Н-связей между белком и лигандами с пятью N-атомами может объяснить их более низкое IC₅₀ по сравнению с лигандами, содержащими четыре атома азота.

Таким образом, результаты молекулярного докинга хорошо согласуются с результатами, полученными при QSAR-моделировании. Установлено, что водородные связи играют важную роль в связывании конъюгатов метилевого синего с NMDAR, а появление в структуре дополнительного атома азота приводит к увеличению активности таких соединений.

Работа выполнена в рамках государственного задания ИФАВ РАН на 2019 год (тема № 0090-2019-0004).

ЛИТЕРАТУРА

1. Wimo, A., Jonsson, L., Bond, J., Prince, M., Winblad, B. The worldwide economic impact of dementia 2010. *Alzheimer's Dement.* 2013. 9 (1), pp. 1–11. e3. <https://doi.org/10.1016/j.jalz.2012.11.006>
2. Wimo A., Guerchet M., Ali G-C., Wu Y-T., Prina A.M., Winblad B., Jonsson L., Liu Z., Prince M. The worldwide costs of dementia 2015 and comparisons with 2010. *Alzheimer's Dement.* 2017. 13 (1), pp. 1–7. <https://doi.org/10.1016/j.jalz.2016.07.150>
3. Querfurth H.W., LaFerla F.M. Alzheimer's disease. *New Engl. J. Med.*, 362, pp. 329–344. <https://doi.org/10.1056/NEJMra0909142>
4. Carreiras M.C., Mendes E., Perry M.J., Francisco A.P., Marco-Contelles J. The multifactorial nature of Alzheimer's disease for developing potential therapeutics. *Curr. Top Med. Chem.* 2013. 13, pp. 1745–1770. <https://doi.org/10.2174/15680266113139990135>
5. Liu H., Wang L., Lv M., Pei R., Li P., Pei Z., Wang Y., Su W., Xie X.-Q. AlzPlatform: An Alzheimer's Disease Domain-Specific Chemogenomics Knowledgebase for Polypharmacology and Target Identification Research. 2014. *J. Chem. Inf. Model.*, 54, pp. 1050–1060. <https://doi.org/10.1021/ci500004h>
6. Raevsky O.A., Mukhametov A., Grigorev V.Y., Ustyugov A., Tsay S-C., Hwu R.-J.-R., Yarla N.S., Barreto G.E., Aliev G., Bachurin S.O. Applications of Multi-Target Computer-Aided Methodologies in Molecular Design of CNS Drugs. *Curr. Med. Chem.* 2018. 25 (39), pp. 5293–5314. <https://dx.doi.org/10.2174/0929867324666170920154111>
7. Zambre V.P., Hambarde V.A., Petkar N.N., Patel C.N., Sawan S.D. Structural investigations by in silico modeling for designing NR2B subunit selective NMDA receptor antagonists. 2015. *RSC Adv.*, 5, pp. 23922–23940. <https://doi.org/10.1039/c5ra01098e>

8. *Chtita S., Larif M., Ghamali M., Bouachrine M., Lakhlifi, T.* 2015. Quantitative structure – activity relationship studies of dibenzo[a,d]cycloalkenimine derivatives for non-competitive antagonists of N-methyl-d-aspartate based on density functional theory with electronic and topological descriptors. *J. Taibah Univ. Sci.*, 9 (2), pp.143–154. <https://dx.doi.org/10.1016/j.jtusci.2014.10.006>
9. *Elmbarki C., Elhallaoui M.* 2017. Quantum calculations to construct a 3D-QSAR model based on PCP-TCP derivatives and Molecular Docking with NMDA receptor. *J. Mater. Environ. Sci.*, 8 (4), pp. 1391–1400.
10. *Peters O.M., Connor-Robson, N., Sokolov, V.B., Aksinenko, A.Yu., Kukharsky, M.S., Bachurin, S.O., Ninkina, N., Buchman, V.L.* 2013. Chronic Administration of Dimebon Ameliorates Pathology in TauP301S Transgenic Mice. *J. Alzheimer's Dis.*, 33 (4), pp. 1041–1049. <https://doi.org/10.3233/JAD-2012-121732>
11. *Zhao, M., Liang, F., Xu, H., Yan, W., Zhang, J.* 2016. Methylene blue exerts a neuroprotective effect against traumatic brain injury by promoting autophagy and inhibiting microglial activation. *Mol. Med. Rep.*, 13, pp. 13–20. <https://doi.org/10.3892/mmr.2015.4551>
12. *Раевский О.А., Трепалин С.В., Григорьев В.Ю., Ярков А.В., Бачурин С.О.* Мультитаргетные органические соединения с потенциальным действием на центральную нервную систему. 2017. Свидетельство о государственной регистрации базы данных № 2017620020 от 09.01.2017, Москва.
13. *Coughenour L.L., Barr B.M.* 2001. Use of Trifluoroperazine Isolates a [3H]Ifenprodil Binding Site in Rat Brain Membranes with the Pharmacology of the Voltage-Independent Ifenprodil Site on N-Methyl-D-aspartate Receptors Containing NR2B Subunits. *J. Pharmacol. Exp. Ther.* 296, pp. 150–159.
14. *Raevsky O.A., Grigor'ev V.Yu., Kireev D.B., Zefirov N.S.* 1992. Complete Thermodynamic Description of H-Bonding in the Framework of Multiplicative Approach. *Quant. Struct.-Act. Relat.* 11, pp. 49–63. <https://doi.org/10.1002/qsar.19920110109>
15. *Forsythe G.E., Malcolm M.A., Moler C.B.* 1977. *Computer Methods for Mathematical Computations.* Englewood Cliffs, NJ: Prentice-Hall, 270.
16. URL: http://www.stat.berkeley.edu/~breiman/RandomForests/cc_examples/prog.f
17. URL: <https://github.com/jbcolme/fortran-ls-svm>
18. *Obrezanova O., Csányi G., Gola J.M.R., Segall M.D.* Gaussian Processes: A Method for Automatic QSAR Modeling of ADME Properties. *J. Chem. Inf. Model.* 2007. 47, pp. 1847–1857. <https://doi.org/10.1021/ci7000633>
19. *Mitra I., Saha A., Roy K.* Exploring quantitative structure–activity relationship studies of antioxidant phenolic compounds obtained from traditional Chinese medicinal plants. 2010. *Mol. Simul.* 36 (13), pp. 1067–1079. <https://doi.org/10.1080/08927022.2010.503326>
20. *Trott O., Olson A.J.* AutoDockVina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithread-

ing. *J. Comput. Chem.* 2010. 31 (2), pp. 455–461. <https://doi.org/10.1002/jcc.21334>

21. *Morris G.M., Huey R., Lindstrom W., Sanner M.F., Belew R.K., Goodsell D.S., Olson A.J.* AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* 2009. 30 (16), pp. 2785–2791. <https://doi.org/10.1002/jcc.21256>

22. *Kiralj R., Ferreira M.M.C.* Basic Validation Procedures for Regression Models in QSAR and QSPR Studies: Theory and Application. *J. Braz. Chem. Soc.* 20 (4). 2009. pp. 770–787. <https://doi.org/10.1590/S0103-50532009000400021>

23. *Gramatica P.* Principles of QSAR models validation: internal and external. *QSAR. Comb. Sci.* 2007. 26 (5), pp. 694–701. <https://doi.org/10.1002/qsar.200610151>

24. *Bolshakov K.V., Gmiro V.E., Tikhonov D.B., Magazani, L.G.* Determinants of trapping block of N-methyl- D -aspartate receptor channels. *J. Neurochem.* 2003. 87, pp. 56–65. <https://doi.org/10.1046/j.1471-4159.2003.01956.x>

25. *Beteringhe A., Filip P., Tarko L.* QSAR study for diarylguanidines, noncompetitive NMDA receptor antagonists. A new topological index AAd derived from local invariants of the chemical graphs of diarylguanidines. *Arkivoc.* 2005. pp. 45–62. <https://doi.org/10.3998/ark.5550190.0006.a05>

26. *Huettner, J.E., Bean, B.P.* Block of N-methyl-D-aspartate-activated current by the anticonvulsant MK-801: Selective binding to open channels. *Proc. Natl. Acad. Sci. USA.* 1988. 85, pp. 1307–1311. <https://doi.org/10.1073/pnas.85.4.1307>

**ИЗБРАННЫЕ ТРУДЫ
ДОКТОРА ХИМИЧЕСКИХ НАУК,
ПРОФЕССОРА О.А. РАЕВСКОГО
В ОБЛАСТИ МОЛЕКУЛЯРНОГО
КОМПЬЮТЕРНОГО ДИЗАЙНА**

Подписано в печать ??????
Формат 70x100 1/16
Гарнитура Times
Усл.-п. л. 18,5. Уч.-изд. л. 11,2
Тираж 300 экз.

Издатель – Российская академия наук

Публикуется в авторской редакции

Корректурa и верстка – ООО «ВИН»

Отпечатано в экспериментальной цифровой типографии РАН

Издается по решению Научно-издательского совета
Российской академии наук (НИСО РАН) от 12.02.2021 г.
и распространяется бесплатно