

Многослойные перцептроны в исследовании зависимостей «структура—свойство» для органических соединений

И. И. Баскин, В. А. Палюлин, Н. С. Зефирова

ИГОРЬ ИОСИФОВИЧ БАСКИН — кандидат химических наук, ведущий научный сотрудник Химического факультета МГУ им. М.В. Ломоносова. Область научных интересов: исследование соотношений «структура—свойство», «структура—активность» (QSPR/QSAR) для органических соединений, искусственные нейронные сети, химическая информатика, молекулярное моделирование. E-mail baskin@org.chem.msu.ru

ВЛАДИМИР АЛЕКСАНДРОВИЧ ПАЛЮЛИН — кандидат химических наук, ведущий научный сотрудник Химического факультета МГУ им. М.В. Ломоносова. Область научных интересов: органическая химия, исследование соотношений «структура—свойство», «структура—активность» (QSPR/QSAR) для органических соединений, молекулярное моделирование. E-mail var@org.chem.msu.ru

НИКОЛАЙ СЕРАФИМОВИЧ ЗЕФИРОВ — академик РАН, профессор, заведующий кафедрой органической химии Химического факультета МГУ им. М.В. Ломоносова, директор Института физиологически активных веществ РАН. Область научных интересов: органическая и медицинская химия, исследование соотношений «структура—свойство», «структура—активность» (QSPR/QSAR) для органических соединений. E-mail zefirov@org.chem.msu.ru

119992 Москва, Ленинские горы, МГУ, Химический факультет, тел. (495)939-35-57, факс (495)939-02-90.

С накоплением огромной массы экспериментальных данных по физико-химическим свойствам, биологической активности, спектрам и реакционной способности органических соединений актуальной стала задача их обобщения и статистической обработки. Решение этой задачи позволяет перейти к созданию методов прогнозирования свойств новых химических соединений, равно как и к направленному молекулярному дизайну соединений с заданным набором свойств. В последние годы активно используются более сотни разнообразных статистических подходов и методов машинного обучения. Среди них особую и постоянно возрастающую роль играет группа методов, объединенных под общим названием «искусственные нейронные сети», «вычислительные нейронные сети» или просто «нейросети».

Биологические и искусственные нейроны

В основе всех нейросетевых методов лежат идеи, отражающие те или иные аспекты обработки информации в человеческом мозгу. Искусственные нейронные сети состоят из определенного количества искусственных нейронов, являющихся упрощенной математической моделью биологических нейронов, и связей между ними, соответствующих синаптическим контактам между аксонами и дендритами* биологических нейронов (впервые математическая модель искусственного нейрона предложена в 1943 году У.С. МакКаллоком и В. Питтсом [1]).

Подобно тому, как биологические нейроны вследствие деполяризации мембраны под действием нейромедиаторов способны возбуждаться и проявлять спайковую активность, так и их искусственные аналоги имеют определенный уровень активности, обычно в интервале от 0 (активность нейрона в состоянии по-

кой) до 1 (активность возбужденного нейрона). Этот уровень активности передается в виде сигнала на соседние искусственные нейроны, что имитирует биологические процессы распространения деполяризации мембраны по аксону, выделения молекул нейромедиатора, их диффузии через синаптические щели и воздействия на рецепторы, расположенные на мембранах дендритов соседних нейронов. Весь этот сложный процесс передачи сигнала от одного нейрона к другому описывается в методологии искусственных нейронных сетей одним числом, называемым весом связи, которое является аналогом понятия синаптической проводимости биологических нейронов. Обычно считается, что степень воздействия искусственного нейрона j на другой нейрон i равна произведению уровня активности первого нейрона o_j на вес связи (синаптическую проводимость) w_{ji} между ними. Положительное значение синаптической проводимости соответствует прохождению через синаптические контакты возбуждающих нейромедиаторов, например глутамата или ацетилхолина, а отрицательное — тормозящих, например γ -аминомасляной кислоты. Абсолютная величина этого числа отражает степень легкости передачи сигнала, что в случае биологических нейронов определяется количеством и разветвленностью синаптических контактов, уровнем экспрессии и активности постсинаптических рецепторов, легкостью выделения нейромедиаторов и многими другими факторами, управляемыми как генетически, так и при помощи разнообразных сигнальных систем.

В методе искусственных нейронных сетей функционирование отдельного нейрона обычно описывается следующим уравнением (см. рис. 1):

$$o_i = g\left(\sum_j w_{ji} o_j - t_i\right) \quad (1)$$

где o_i — результирующий выходной сигнал (равный уровню активности) нейрона i ; w_{ji} — вес связи (синап-

* Аксон — отросток нейрона, проводящий нервный импульс от клетки к другим нейронам; дендриты — отростки нейрона, воспринимающие нервный импульс от других нейронов.

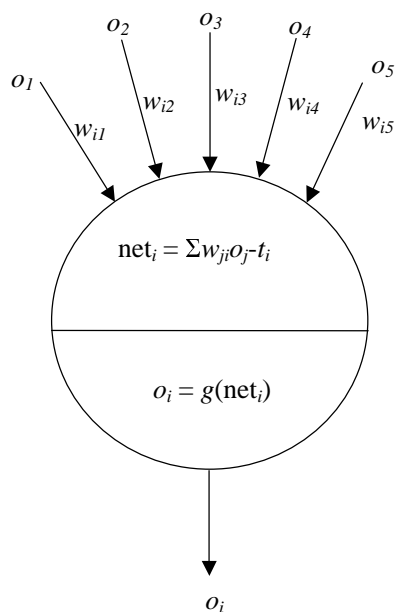


Рис. 1. Нейрон МакКаллока—Питтса

тическая проводимость) между нейронами j и i ; o_j — выходной сигнал (равный уровню активности) соседнего нейрона j , с которого пришел сигнал; t_i — порог активации нейрона i , по превышении которого суммой воздействий со стороны соседних нейронов он переходит в возбужденное состояние; $g(x)$ — так называемая функция активации нейрона, которая в простейшем случае может быть, например, определена как пороговая:

$$g(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (2)$$

Таким образом, уравнение (1) в сочетании с определением функции (2) описывает, хотя и очень упрощенно, функционирование биологического нейрона, находящегося, например, в коре головного мозга человека.

Подобно своему биологическому прототипу, нейроны МакКаллока—Питтса способны обучаться путем подстройки параметров w , описывающих синаптическую проводимость.

Перцептроны

На функционировании описанного выше искусственного нейрона были основаны разработанные более 40 лет назад первые типы искусственных нейронных сетей, получивших название «перцептроны» (термин «перцептрон» происходит от английского слова perception — восприятие) [2–4]. Тогда же были предприняты первые попытки создания искусственного интеллекта путем имитации работы головного мозга человека на клеточном уровне. Перцептроны предложены в 1958 г. Фрэнком Розенблаттом именно в связи с попытками имитировать с помощью нейронов МакКаллока—Питтса человеческое восприятие (прежде всего зрение) и распознавание объектов внешнего мира.

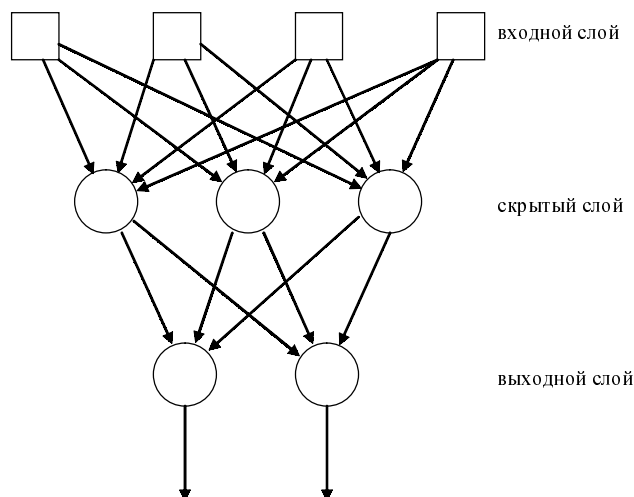


Рис. 2. Многослойный перцептрон

Перцептрон Розенблатта имеет многослойную архитектуру (рис. 2), причем только последний (выходной) слой содержит нейроны с подстраиваемыми параметрами синаптической проводимости. Формируемые ими выходные сигналы свидетельствуют о принадлежности анализируемого объекта к определенному классу. Само описание объекта в перцептронах Розенблатта формируется на входном слое нейронов, названном рецепторным полем по аналогии с биологическим прототипом. Сигналы с рецепторного поля поступают на необязательный скрытый слой нейронов по синапсам, проводимость которых иницируется случайными числами и в процессе обучения не меняется. Генерируемые на нейронах скрытого слоя сигналы в свою очередь поступают на выходной слой нейронов для дальнейшей обработки.

Эти первые попытки имитации человеческого восприятия на нейронах МакКаллока—Питтса оказались не совсем удачными и не оправдали всех возлагавшихся на них надежд [4]. Поскольку в модель Розенблатта был заложен известный в то время способ настройки весов связей, идущих лишь к нейронам одного (выходного) слоя, то на практике перцептроны Розенблатта оказались неспособными обучаться распознаванию сложных образов, их реальная распознающая способность была не выше, чем у более простых и понятных стандартных методов дискриминантного анализа. Все это привело к прекращению практически всех проводившихся в этом направлении работ. Лишь к середине 1980-х годов стало ясно, что одна из причин неудач кроется в конкретном виде пороговой функции активации (2). Замена пороговой функции (2) на непрерывную, ограниченную и монотонно возрастающую, например сигмоидную функцию (3), привела к построению многослойных перцептронов, все веса связей которых способны эффективно обучаться при помощи алгоритма обратного распространения ошибки (error backpropagation) [5, 6]:

$$g(x) = 1/(1 + e^{-x}) \quad (3)$$

Кроме чисто математических причин, переход к подобным непрерывным дифференцируемым функци-

ям имеет и определенное нейрофизиологическое обоснование. С точки зрения способа передачи информации сигнал реальных биологических нейронов модулирован не по амплитуде, а по частоте, и к тому же является стохастическим, что вполне согласуется с уравнениями (1) и (3) при условии, что сигнал (уровень активации) o_i показывает, с какой вероятностью нейрон i переходит в возбужденное состояние.

С конца 1980-х годов начался этап активного развития и использования аппарата искусственных нейронных сетей в разных областях науки и техники (см., например, пособия [7–12]), а с начала 1990-х — в различных областях химии ([13]) и, в частности, в области исследования зависимостей «структура—свойство» для органических соединений [14].

Общая методика использования многослойных персептронов для поиска зависимостей «структура—свойство»

Прежде всего подготавливается база данных, содержащая структуры химических соединений и известные значения тех свойств, которые в дальнейшем предполагается прогнозировать с помощью обученной нейросети. Как правило, эта база разбивается на две части. По первой из них, называемой обучающей выборкой, путем многократного ее предъявления производится обучение нейросети. Вторая, называемая контрольной выборкой, используется для контроля прогнозирующей способности обученной нейросети. Как варианты могут использоваться две контрольные выборки, а также процедура скользящего контроля, когда каждое из химических соединений поочередно попадает в контрольную выборку.

На следующем этапе для всех химических соединений из выборок производится расчет дескрипторов, т.е. чисел, описывающих химическую структуру соединения. Как правило, эти числа инвариантны к перенумерации вершин молекулярного графа, которым может быть описана структура химического соединения, т.е. являются инвариантами графов. Дескрипторы могут быть фрагментными (подструктурными), топологическими индексами, физико-химическими, квантовохимическими, могут характеризовать пространственную структуру и т.д. Достаточно полный набор дескрипторов, используемых в современных исследованиях «структура—свойство», описан в книге Р. Тодескини и В. Консонни [15].

Далее, после необязательной стадии предварительного отбора либо преобразования дескрипторов следует этап построения персептрона. Число нейронов входного слоя (рецепторного поля) обычно устанавливается равным числу дескрипторов (часто во входной слой добавляют еще один псевдонейрон смещения с постоянным уровнем выходного сигнала, т.е. с постоянным уровнем активности), а уровень выходного сигнала каждого из них устанавливается равным числовому значению соответствующего дескриптора (обычно после нормализации или масштабирования). Число нейронов выходного слоя принимается равным числу прогнозируемых свойств, причем в качестве прогнозируемого значения каждого из свойств берется значение выходного сигнала соответствующего выходного нейрона (обычно после денормализации или демасштабирования). Нейроны скрытого слоя участ-

вуют при промежуточных вычислениях, и их число часто подбирается, исходя из критерия наилучшей прогнозирующей способности нейросети. Что касается псевдонейронов смещения, то они выполняют служебные функции и имеют постоянный выходной сигнал, равный единице.

В процессе обучения нейросети обучающая выборка предъявляется нейросети определенное число раз (обычно довольно большое). В процессе предъявления выборки значения дескрипторов каждого соединения последовательно вводятся (обычно после нормализации или масштабирования) в качестве активности соответствующих входных нейронов, инициируется процесс работы нейросети (см. выше) и с выходных нейронов снимаются прогнозируемые значения свойств. Полученные значения (после денормализации или демасштабирования) сравниваются с экспериментальными и на основании выявленной разницы по определенным алгоритмам производится подстройка весов связей между нейронами с целью уменьшения этой разницы. Таким образом, в процессе обучения происходит постепенное уменьшение ошибок прогнозирования свойств химических соединений, входящих в обучающую выборку.

Персептрон Розенблатта в исследованиях зависимости «структура—активность»

Первым применением искусственных нейронных сетей в исследованиях зависимостей «структура—свойство» для органических соединений следует считать пионерские работы Гиллера с соавт. (1971–1973 гг.) [16, 17], в которых была использована единственная известная к тому времени нейросеть — персептрон Розенблатта. В этих работах на входные нейроны персептрона подавались дескрипторы, описывающие наличие определенных фрагментов в структурах химических соединений, а исходя из значений выходного сигнала нейрона, получали качественные прогнозы — наличие или отсутствие определенной биологической активности у соответствующего химического соединения.

Надо сказать, что на фоне общего снижения интереса к персептронам эти исследования не получили в то время достойного продолжения. Работы по применению искусственных нейронных сетей в исследованиях «структура—свойство» были возобновлены лишь через 20 лет в начале 1990-х годов прошлого века.

Многослойный персептрон или многослойная нейронная сеть прямого распространения с обратным распространением ошибок

Попытки имитировать обработку сенсорной информации в многослойных структурах соответствующих отделов головного мозга привели к идее объединения искусственных нейронов в многослойные сети с последовательной передачей сигнала от одного слоя к другому. Как и персептроны Розенблатта, эти сети состоят из одного слоя входных нейронов, из которого сигналы поступают на один или несколько слоев скрытых нейронов и затем передаются на слой выходных нейронов (см. рис. 2). Благодаря использованию сигмоидной функции активации нейрона (3), для обучения нейросети возможно применение эффективной процедуры обратного распространения ошибок

(backpropagation) [6]. Соответственно эти сети получили название «многослойные сети прямого распространения с обратным распространением ошибок» (feed-forward neural networks with error backpropagation).

Популярности многослойной нейросетевой архитектуры способствовало также еще одно ее качество. В соответствии с теоремой Колмогорова о представлении непрерывных функций нескольких переменных суперпозицией непрерывных функций одной переменной и суммирования [18, 19], ее нейросетевой интерпретацией [20, 21], а также согласно весоному результату, полученному в работе [22], такие нейросети способны аппроксимировать непрерывную функцию многих переменных любой сложности и, как следствие, способны обучаться распознавать образы любой сложности.

Многослойные персептроны в прогнозировании биологической активности органических соединений

Метод многослойных персептронов является ценным инструментом в исследовании зависимостей «структура—свойство» для органических соединений. Это показали уже первые модельные исследования, начатые в 1990 годах. В работах Т. Аоямы с соавт. данный метод использовали для качественной классификации органических соединений на активные и неактивные по отношению к определенным видам биологической активности [23, 24] и для поиска количественных зависимостей между биологической активностью и молекулярными дескрипторами [25]. Почти одновременно в 1991 г. возможность применения многослойных персептронов для построения количественных корреляций «структура—биологическая активность» показали Т.А. Андреа с соавт. [26] на примере изучения ингибирования фермента дигидрофолатредуктазы, а также М. Визе [27], получавший количественные данные по гипертензивной активности.

В 1992 г. исследования по применению многослойных персептронов для качественной классификации органических соединений по уровню проявляемой биологической активности были продолжены в работах по изучению митомицинов [28], нафтохинонов [29], гербицидной активности производных пиридазина [30]. В том же году были опубликованы результаты исследований по построению количественных корреляций «структура—биологическая активность» для пестицидов [31], бензилпиримидиновых ингибиторов дигидрофолатредуктазы [32], а также работы методологического плана [33, 34].

В последующие годы список видов биологической активности, для изучения которых во взаимосвязи со структурами органических соединений были использованы многослойные персептроны, значительно расширился, и к настоящему времени количество опубликованных в этом направлении работ исчисляется многими сотнями. Перечислим лишь ферменты, ингибирование которых было исследовано с использованием нейросетей данного типа (для каждого фермента ссылка дается на первую опубликованную работу): обратная транскриптаза HIV-1 [35], интеграза HIV-1 [36], протеаза HIV-1 [37], цитохромы P450 3A4 [38], 2D6 [39], 1A2 [40], ацил-КоА:холестерол О-ацилтрансфераза [41], изоформы UDP-глюкуронозилтрансферазы [42], фарнезилтрансфераза [43], циклоок-

сигеназа-2 [44], моноаминоксидаза [45], $\text{Na}^+\text{-K}^+\text{-ATPase}$ [46] (в этой работе была применена особая модификация многослойного персептрона без скрытого слоя, называемая сетью функциональных связей), $\text{Na}^+\text{-H}^+\text{-ATPase}$ [47], ацетолактат-синтаза [48], тромбосан-синтаза [49], α -химотрипсин [50], p56lck тирозин-киназа [51], трипаносомальная нуклеозид-гидролаза [52], эпоксид-гидролаза [53], дигидроптеридинредуктаза [54], человеческая 5α -редуктаза типа 1 [55] (в статье рассматривается также селективность ингибирования по отношению к 3β -гидрокси-D5-стероид дегидрогеназе/3-кето-D5-стероид-изомеразе), ангиотензин-превращающий фермент [56], ацетилхолинэстераза [57], фосфорилаза [57], термоллизин [56], тромбин [56], α -химотрипсин [58], дофамин- β -гидроксилаза [57].

Таким образом, многослойные нейросети данного типа успели себя хорошо зарекомендовать при изучении связи «структура—ингибирующая способность» для очень большого числа ферментов.

Имеется также целый ряд работ по использованию многослойных персептронов для изучения связи между структурой органического соединения и его способностью оказывать действие на рецептор. В первой работе [59] в этом направлении рассматривается связывание стероидов с прогестероновым рецептором. В дальнейшем разнообразие исследованных рецепторов значительно расширилось (приводим ссылку на первую опубликованную работу): опиоидные рецепторы (агонистическое действие производных 3-метилфентанила) [60], бензодиазепиновый сайт ГАМК_A-рецептора (агонистическое действие бензодиазепинов) [61], ССК панкреатические рецепторы (антагонистическое действие бензодиазепинов) [62], АсН-рецептор (действие полихлордибензофуранов) [63], мускариновые рецепторы [64], дофаминовый D₂-рецептор (антагонистическое действие тетрагидропротоверберинов) [65], четыре подтипа (D₁, D₂, D₃ и D₄) дофаминового рецептора (антагонистическое действие) [75], серотониновый 5-HT_{2A}-рецептор (антагонистическое действие 2- и 4-(4-метилпиперазино)пиримидинов [66], серотониновый 5-HT_{1A}-рецептор и α_1 -адренорецептор (связывание с арилпиперазинами) [67], разные подтипы α -адренорецептора (агонистическое действие) [69], девять подтипов сопряженных с G-белками рецепторов (GPCR) [68], эстрогеновый рецептор (связывание 4-замещенных производных деоксигексэстрола) [70], глициновый сайт NMDA-рецептора (антагонистическое действие) [71, 72], тромбосановый A₂-рецептор (антагонистическое действие производных бензимидазола и имидазолпиридина) [73], аденозиновые рецепторы (неселективное агонистическое действие бензодиазепинов) [74].

Близкими к действию на рецепторы являются блокировка и открытие ионных каналов. Здесь можно упомянуть работу [76], где на примере блокирования 1,4-дигидропиридинами потенциал-зависимых кальциевых каналов L-типа была развита методология учета симметрии расположения заместителей при построении моделей «структура—активность» [76], работу [77], в которой для прогнозирования биологической активности этих же объектов был использован подход, сочетающий метод многослойных нейронных сетей с процедурой анализа главных компонент (име-

ется и еще несколько более поздних работ для этих же объектов). Отметим также модельные исследования блокаторов hERG-подтипа калиевых каналов [78] и соединений, вызывающих открытие калиевых каналов и обладающих вазорелаксирующим действием [79].

Многослойные перцептроны используются для прогнозирования еще одной важной разновидности биологической активности — токсических свойств, в частности, различных видов цитотоксичности, действия на кожу, а также генотоксичности — мутагенности и канцерогенности.

В работе [80] токсичность фенолов ранжировалась с помощью многослойного перцептрона, принимающего на свой вход значение липофильности и одного из топологических индексов молекулярной связности, причем показатели работы нейросети оказались лучше, чем у процедуры пошагового линейного дискриминантного анализа. В работе [81] этот же тип нейросети был использован для построения корреляции между дескрипторами, описывающими галогензамещенные фенолы, и количественными характеристиками их токсичности, причем статистические показатели нейросетевых моделей были сопоставлены с аналогичными показателями регрессионного анализа.

Многослойные перцептроны весьма широко используются в прогнозировании физиологического действия разнообразных органических токсикантов, проводимого по широкому набору токсикологических тестов. Перечислим изученные с их помощью разновидности токсичности и соответствующие токсиканты (для каждого типа биотеста и группы соединений приводятся только первые опубликованные работы): токсичность люминесцентных бактерий (тест Микро-токс) [82], токсичность производных бензола [83], цитотоксичность антиоксидантов (каротиноидов и липофильных витаминов) [84], токсичность солей бензотиазолия [85] (в этой работе были использованы те же структурные дескрипторы, что и в аддитивном методе Фри—Вильсона [86], причем статистические показатели моделей, построенных с применением обоих подходов, оказались близкими), токсичность ароматических производных серы по отношению к фотобактериям (на этом примере было также продемонстрировано превосходство нейросетевых моделей по сравнению с линейно-регрессионным методом) [87], острая токсичность по отношению к млекопитающим замещенных анилинов [88] и фосфорорганических пестицидов [89], токсичность, измеренная по отношению к домашней мухе, аналогов ДДТ, диарилнитропропанов, диарилтрихлорэтана и др. (также было продемонстрировано преимущество использования нейросетей по сравнению со стандартным регрессионным анализом) [90], токсичность аминов, измеренная в тестах на крысах [91], цитотоксичность фторхинолонов [92], токсичность нитробензолов [93], острая токсичность разнообразных органических соединений, измеренная на рыбах *Pimephales promelas* (толстоголовый гольян) [94], острая токсичность хлорированных углеводородов [95], альдегидов [96] и циклических эфиров [97], токсическое действие пестицидов, измеренное на рыбах *Oncorhynchus mykiss*, [98] и *Lepomis macrochirus* [99] (особенность последних двух работ заключается в том, что построенные нейросетевые модели токсичности включают зависимость ее не

только от структуры органического соединения, но и от «переменных среды», таких как масса рыбы, температура, жесткость и pH воды, а также продолжительность действия токсиканта), токсичность ароматических соединений, измеренная по ингибированию роста инфузории *Tetrahymena pyriformis* [100], цитотоксичность производных НЕРТ (препарата для лечения СПИД) [101], токсичность пестицидов по отношению к пчелам [102].

Таким образом, можно сказать, что многослойные перцептроны активно вошли в практику прогнозирования токсичности органических соединений. Отметим, что применению нейросетей для прогнозирования токсичности органических соединений посвящено несколько недавно опубликованных обзорных статей [103—105].

Важной разновидностью токсичности является генотоксичность, включающая мутагенную и канцерогенную активности. Первые работы по использованию многослойных перцептронов для прогнозирования мутагенной активности датируются 1993 годом, когда в работе [106] с помощью этого типа нейросети провели качественное прогнозирование наличия мутагенной активности у органического соединения, исходя из наличия в его структурной формуле определенных подструктурных фрагментов, оптимальный набор которых подбирался автоматически. В опубликованных в этом же году работах [107—109] многослойные перцептроны были использованы уже при поиске количественной зависимости внутри классов соединений (главным образом ароматические и гетероароматические нитросоединения) между мутагенной активностью и такими дескрипторами молекулярной структуры, как липофильность, энергия низшей свободной молекулярной орбитали и др. Было продемонстрировано преимущество применения для этой цели нейросетей перед линейным регрессионным анализом. Дальнейшие исследования мутагенности с использованием нейросетевого подхода были распространены на другие классы органических соединений (для каждого класса указываем первую публикацию): нитропроизводные полиароматических углеводородов [110], гетероциклические амины (с использованием подструктурных дескрипторов) [111], гетероароматические амины (с использованием квантовохимических дескрипторов) [112], замещенные полициклические соединения (с использованием фрагментных, квантовохимических дескрипторов, а также гидрофобности) [113], аминоазобензольные красители [114] и др.

Первая работа по применению многослойных перцептронов для прогнозирования канцерогенной активности была опубликована также в 1993 году [115]. Был осуществлен нейросетевой прогноз канцерогенной активности третичных ароматических аминов с использованием подструктурных дескрипторов. Далее (1994 г.) этот подход был распространен на полициклические ароматические углеводороды [116]. Предсказанию канцерогенности именно внутри этих двух классов органических соединений посвящено большинство последующих работ, и лишь в последние два года начали появляться публикации по использованию многослойных перцептронов для прогнозирования канцерогенной активности широкого круга разнообразных органических соединений [117, 118].

В работе [117] прогноз осуществлялся с использованием фрагментных дескрипторов, оптимальный набор которых был получен с помощью генетического алгоритма, а в работе [118] в качестве дескрипторов использовался набор главных компонент, рассчитанных на основе первоначального набора большого числа разнообразных молекулярных дескрипторов.

Метод многослойных перцептронов нашел применение в экотоксикологии для изучения кинетики дезактивации токсикантов под воздействием физических факторов окружающей среды и микроорганизмов. Способность к биодegradации была впервые спрогнозирована с применением многослойных перцептронов для производных бензола в [119]. Впоследствии аналогичные исследования были проведены и в отношении ряда других классов органических экотоксикантов.

Отдельный цикл исследований посвящен прогнозированию «полезной» разновидности токсичности, проявляемой пестицидами. Так, в первой в этом направлении работе [30] был проведен анализ гербицидной активности в ряду производных пиридазина. Впоследствии гербицидная активность была изучена у фосфорорганических соединений (с использованием модификации многослойного перцептрона — сети функциональных связей) [120], амидов [121], производных фенилацетамида [122–125] (работа [124] еще интересна и тем, что в ней, по-видимому, впервые был применен метод «роя частиц» для обучения нейросетей, а в работе [125] для прогноза гербицидной активности применялась не отдельная нейросеть, а их «комитет»), N-фенилацетамидных гербицидов, ингибирующих фотосистему II [126], тиопиримидильных эфиров [127], бензоилциклогександионов [128]. Обратим внимание, что практически все работы по изучению гербицидной активности были выполнены исключительно китайскими и румынскими исследователями, что, очевидно, отражает определенные тенденции в сельском хозяйстве этих стран.

Первая работа по выявлению связи между структурой органического соединения и его инсектицидной способностью с использованием многослойных перцептронов была опубликована в 1993 г. [129]. В этой работе, посвященной изучению активности производных пиридина, в качестве дескрипторов были использованы специально сконструированные инварианты молекулярного графа, описывающие пространственную форму заместителей пиридина. На основе построенной нейросетевой модели был осуществлен дизайн новых инсектицидов, которые были синтезированы и при экспериментальной проверке показали высокую инсектицидную активность. В последующих работах в этом направлении объектом исследования были в основном фосфорорганические инсектициды (первая публикация [130]), но имеется также работа по инсектицидной активности спинозинов (семейство макроциклических лактонов) по отношению к бабочкам и двукрылым насекомым [131].

В рамках исследований по прогнозированию активности пестицидов были проведены работы по изучению кинетической устойчивости, в частности, времени полураспада. В первой работе в этом направлении (1993 г.) был применен многослойный перцептрон, использующий в качестве дескрипторов числа встречаемости определенных структурных фрагментов

в молекуле [132]. Следует упомянуть также исследование по предсказанию (с использованием квантовохимических дескрипторов) коэффициентов абсорбции в почве гербицидов из группы фенилмочевин [133]. В обеих этих работах отмечается существенное преимущество использования нейросетей по сравнению со стандартными статистическими процедурами.

Из других свойств, которые условно можно считать биологической активностью и для прогнозирования которых применялся многослойный перцептрон, отметим запах. В первой работе в этом направлении (1991 г.) нейросеть такого типа была использована для предсказания наличия мускусного запаха у производных нитробензола [134]. В качестве дескрипторов были использованы параметры заместителей, описывающие объем и электроотрицательность. Впоследствии этот подход был распространен на прогнозирование фруктового и камфорного запаха у алифатических спиртов [135], орехового и двух типов перцового запаха у пиразинов и триазолов [136], сандалового запаха у группы органических соединений, содержащих алифатические циклы и полициклические фрагменты [137], мускусного запаха у производных тетралина и индана [138] и др.

Несомненно важным приложением метода многослойных перцептронов, примыкающим к токсикологии, является изучение фармакокинетических свойств веществ: кинетика всасывания химических препаратов в организм, перераспределения между тканями, метаболизм и вывод из организма (так называемые ADME свойства). Наиболее изучена из этих свойств способность к преодолению гематоэнцефалического барьера, зависимость которой от структуры впервые была смоделирована с применением многослойных перцептронов в работе [139]. С использованием подобного подхода изучалась оральная абсорбция [140]. К этой же категории можно отнести и работу [141], в которой многослойный перцептрон в сочетании с фрагментными дескрипторами был использован для прогнозирования воздействия органических соединений на центральную нервную систему.

Многослойные перцептроны в прогнозировании физико-химических свойств органических соединений

Работы по применению метода многослойных перцептронов к моделированию и прогнозированию физико-химических свойств активно начались в 1990-х годах. Представленный в данной статье обзор опубликованных статей, отражающий развитие работ по этому направлению, показывает широкие практические возможности и перспективность рассматриваемого метода.

В одной из первых опубликованных работ (1991 г.), посвященных исследованию зависимости между структурами органических соединений и их физико-химическими свойствами методом многослойных перцептронов [142], прогнозировалась растворимость в воде органических соединений с использованием квантовохимических дескрипторов. В работе [144] многослойные перцептроны были впервые применены для прогнозирования физико-химических свойств алканов (температур кипения и плавления, молярного объема, молярной рефракции, теплоты испарения, критической температуры, критического давления, поверхностного натяжения) и октанового числа угле-

водородов. Было показано, что в большинстве случаев многослойные перцептроны обеспечивают более высокую прогнозирующую способность по сравнению с множественной линейной регрессией. Кроме того установлено, что использование чисел встречаемости подструктурных фрагментов в качестве дескрипторов обеспечивает построение лучших моделей, чем на основании топологических индексов. В [145] многослойные перцептроны были применены для прогнозирования температуры кипения гетероциклических соединений с использованием набора разнообразных дескрипторов (топологических индексов, пространственных, физико-химических и квантовохимических дескрипторов), и снова был сделан вывод о преимуществе использования многослойных перцептронов по сравнению с множественной линейной регрессией.

Есть сообщение [146] о применении многослойных перцептронов для нахождения коэффициентов активности химических соединений в жидкой фазе, что дает возможность рассчитать равновесия между жидкой фазой и парами. Здесь в качестве дескрипторов были использованы групповые вклады (в сущности фрагментные дескрипторы), а результаты прогнозирования были сопоставлены с результатами метода UNIFAC, обычно применяемого для этой же цели.

Начиная с 1994 года, многослойные перцептроны стали широко применяться для исследования зависимостей «структура—физико-химические свойства» для органических соединений без ограничений на класс. Построенные модели пригодны для прогнозирования свойств произвольных органических соединений. В рамках этого подхода в работе [147] были построены нейросетевые модели, позволяющие прогнозировать температуру кипения (при нормальных условиях) и критическую температуру произвольных органических соединений, причем по достигнутой точности прогнозирования модель превзошла ранее используемый для этой цели метод групповых вкладов Джобака. Методом многослойных перцептронов осуществлено прогнозирование коэффициента распределения органических молекул в системе октанол-вода (модель липофильности органических соединений) с использованием квантовохимических дескрипторов [143].

На основе многослойных перцептронов разработана нейросетевая модель, позволяющая прогнозировать вязкость жидких органических соединений при 20 °С [148]. Модель базируется на известных значениях других физических свойств, таких как молярная рефракция, критическая температура, молярная магнитная восприимчивость и энергия когезии (испарения), а также нескольких индикаторных переменных, показывающих наличие определенных структурных фрагментов в молекуле. Первые нейросетевые модели, позволяющие прогнозировать вязкость на основе теоретически вычисляемых дескрипторов, были предложены в работах [149, 150] (с использованием фрагментных дескрипторов) и в [151] (с использованием набора разнообразных дескрипторов). Работа [149] была, по-видимому, также первой, в которой описывается нейросетевая модель для расчета плотности произвольных жидких органических соединений, а в работе [151] дана модель для прогнозирования их поверхностного натяжения.

Предложена нейросетевая модель для расчета давления насыщенных паров [152]. Хроматографические индексы удерживания для органических соединений внутри одного класса (стероидов) были связаны с помощью многослойных перцептронов с квантовохимическими дескрипторами [153]. В первой нейросетевой модели, позволяющей прогнозировать хроматографические свойства произвольных органических соединений, [154] в качестве дескрипторов использовался набор топологических индексов.

Многослойный перцептрон, способный классифицировать органические соединения по четырем уровням их чувствительности к удару на основе значений фрагментных дескрипторов, описан в статье [155]. Позднее этими же авторами была построена на фрагментных дескрипторах количественная нейросетевая модель [156]. Описан многослойный перцептрон, обученный прогнозировать энтальпию сублимации твердых органических соединений [157]. Многослойный перцептрон был использован для расчета температуры вспышки [158], магнитной восприимчивости [159] разнообразных органических соединений, а также средства азокрасителей и антрахиноновых красителей к целлюлозе [160]. В работе [161] описан метод прогнозирования температуры просветления нематических жидких кристаллов при помощи многослойного перцептрона. Опубликовано несколько статей, описывающих использование многослойных перцептронов для прогнозирования температуры стеклования полимерных [162, 163] и неполимерных соединений [164].

Разработана нейросетевая модель для прогнозирования диэлектрической проницаемости [165]. Сольватохромный параметр полярности (поляризуемости) был впервые смоделирован с использованием этой методологии в работе [166]. Б. Митчел и П. Джурс применили многослойный перцептрон для прогнозирования температуры самовоспламенения [167]. Константы закона Генри были впервые спрогнозированы при помощи многослойных перцептронов в работе [168].

Многослойные перцептроны как нельзя лучше подходят для моделирования физико-химических свойств, зависящих в значительной мере от параметров состояния (температура, давление, а также состав растворов и смесей), благодаря своей способности аппроксимировать нелинейные зависимости произвольного вида. В этом случае при формировании входных сигналов нейросети к набору значений дескрипторов добавляются значения одного или нескольких параметров состояния. В определенном смысле нейросети представляют собой альтернативу традиционному для физической химии подходу, предполагающему использование в явном виде формул для аппроксимации зависимости подобных свойств от значений параметров состояния.

Применение нейросетей позволяет исследовать зависимость «структура—условия—свойство» для органических соединений. В рамках этой методологии были построены многослойные перцептроны, обученные прогнозировать: (а) давление насыщенных паров углеводородов [169] и произвольных органических соединений [170] в зависимости от температуры, (б) вязкость в зависимости от температуры [171] (при использовании в качестве дескрипторов других физико-химических характеристик), (в) температуру кипе-

ния углеводов в зависимости от давления, (ρ) плотность и динамическую вязкость углеводов в зависимости от температуры [172] (при использовании фрагментных дескрипторов).

Еще одна характерная особенность физико-химических свойств — возможность их оценки в том или ином приближении с помощью квантовохимических расчетов. Применение многослойных перцептронов позволяет значительно повысить точность предсказаний свойств химических соединений за счет использования информации из экспериментальной базы данных. В этом случае при формировании входных сигналов нейросети к набору дескрипторов добавляются значения либо прогнозируемого свойства, либо других свойств, тесно с ним связанных, полученные в результате молекулярного моделирования и квантовохимических расчетов. В качестве примеров использования этой методологии можно привести работу [173] по оценке энергии окисления замещенных дигидропиридинов и аналогичных гетероциклов, работу [174] по прогнозированию положения длинноволновой полосы поглощения цианиновых красителей в спиртовом растворе и работу [175], в которой за счет использования многослойных перцептронов для постобработки результатов неэмпирических квантовохимических расчетов достигнуто многократное увеличение их точности. К этой же категории можно отнести работу [176], в которой с помощью многослойного перцептрона было достигнуто восьмикратное уменьшение ошибок при прогнозировании положений полос поглощения ИК-спектров с использованием полумпирических квантовохимических расчетов.

В последние годы многослойные перцептроны нашли применение при изучении супрамолекулярных явлений, таких как комплексообразование и образование соединений включения. Наибольшее число работ посвящено образованию комплексов включения между β -циклодекстрином и производными бензола. Первой подобной публикацией является работа [177] (1999 г.) по прогнозированию константы связывания (ассоциации) в указанной системе с использованием особой модификации многослойного перцептрона (так называемая вейвлетная нейронная сеть), в которой сигмоидная передаточная функция нейронов заменена специальным вейвлетным преобразованием. В статье [178] описывается прогнозирование константы ассоциации барбитуратов с α - и β -циклодекстринами с помощью стандартного многослойного перцептрона. Многослойный перцептрон был применен для прогнозирования констант устойчивости комплексов Gd(III) [179].

Метод перцептронов позволяет прогнозировать локальные свойства молекулы, относящиеся к ее определенной части, например, к определенному атому или заместителю. Исторически первыми локальными свойствами, для прогнозирования которых были использованы многослойные перцептроны, явились химические сдвиги ядер атомов в ЯМР-спектрах. Наибольшее число работ в этом направлении посвящено ЯМР-спектрам на ядрах ^{13}C . Одной из первых публикаций из этой серии работ является статья [180] (1993 г.). Здесь химические сдвиги ^{13}C в алканах предсказывались с помощью многослойного перцептрона, использующего в качестве входных сигналов надструктурные дескрипторы заселенности, которые показы-

вают заселенность определенных вершин суперграфа при вписывании в него ближайшего окружения текущего атома. Первоначально большинство опубликованных работ в этом направлении были ограничены одним структурным классом органических соединений. Но уже в 2000 г. появилась работа [181], в которой многослойный перцептрон был обучен воспроизводить химические сдвиги ^{13}C по очень большой разнородной выборке, состоящей из полумиллиона вариантов окружения атомов углерода. Благодаря этому обученная нейросеть оказалась способной точно предсказывать химические сдвиги для произвольного органического соединения без ограничений на его класс. Аналогичный подход был применен к расчету химических сдвигов на ядрах ^{31}P [182] и на ядрах ^1H [183].

Из неспектроскопических локальных свойств одним из наиболее важных является значение кислотности pK_a . Многослойные перцептроны были применены для прогнозирования pK_a замещенных фенолов [184], производных бензойной кислоты [185] и пиридина [186]. Другим важным локальным свойством являются параметры Абрахама, описывающие способность атомов в молекуле выступать в роли донора либо акцептора водородной связи. Многослойный перцептрон был применен для прогнозирования четырех параметров Абрахама (E , S , A и B), характеризующих различные вклады в свободную энергию взаимодействия растворенного вещества с растворителем [187]. К локальным свойствам органических соединений можно также отнести константы заместителей, значения которых впервые были спрогнозированы с применением многослойного перцептрона и фрагментных дескрипторов в работе [188].

Заключение

Многослойные перцептроны являются универсальным инструментом, позволяющим исследовать зависимости между структурами органических соединений и широким спектром их физико-химических свойств и биологической активности. Вместе с тем потенциал этого типа нейросетей еще далеко не исчерпан, о чем свидетельствует не только все возрастающее из года в год число публикаций по их применению для решения разнообразных задач, но и постоянно возникающие архитектуры нейронных сетей (такие как, например, нейронные устройства для построения прямых корреляций «структура—свойство» [189, 190], либо ассоциативные нейронные сети [191]), включающие многослойные перцептроны в качестве составных элементов.

Кроме многослойных перцептронов, существуют и другие нейросетевые архитектуры, нашедшие широкое применение в исследованиях «структура—свойство» для органических соединений, из которых наиболее популярными являются самоорганизующиеся карты Кохонена. Вместе с тем разные типы нейросетей, как правило, ориентированы на решение различных задач, и поэтому они могут успешно дополнять друг друга. В качестве примера подобного сочетания можно привести подход, согласно которому многослойный перцептрон обучается по выборке, сформированной из исходной базы данных с использованием процедуры кластеризации, реализованной с помощью карт Кохонена либо сетей адаптивного резонанса. Многослой-

ные перцептроны также хорошо сочетаются и со многими нейросетевыми методами и алгоритмами машинного обучения (такими, например, как генетические алгоритмы, анализ главных либо независимых компонент, методы муравьиных колоний, роя частиц и т.д.), которые могут, в частности, использоваться для отбора дескрипторов либо понижения размерности их пространства.

Одна из причин популярности перцептронов кроется в их универсальности. Хотя перцептроны, разработанные в рамках проекта создания искусственного интеллекта, первоначально предназначались для моделирования человеческого зрения, они оказались очень хорошо приспособленными и для работы в других областях, в частности, для проведения исследований соотношения «структура—свойство» для органических соединений. В зависимости от типа свойства они могут решать задачи как классификации (при дискретном наборе значений моделируемого свойства), так и регрессии (при непрерывных значениях моделируемого свойства). Другая причина популярности перцептронов в прикладных исследованиях заключается в их наглядности и отсутствии необходимости в глубоком знании математической статистики для их эффективного использования.

ЛИТЕРАТУРА

1. McCulloch W.S., Pitts W. Bull. Math. Biophys., 1943, v. 5, p. 115.
2. Розенблатт Ф. Принципы нейродинамики. М.: Мир, 1964.
3. Нильсен Н. Обучающиеся машины. М.: Мир, 1967.
4. Минский М., Пейперт С. Перцептроны. М.: Мир, 1971.
5. Rumelhart D.E., McClelland J.L. Parallel Distributed Processing. V. 1, 2, MIT Press, 1986.
6. Rumelhart D.E., Hinton G.E., Williams R.J. In: Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Eds. J.L. McClelland, D.E. Rumelhart. MIT Press, Cambridge MA, 1986, v. 1, p. 318.
7. Горбань А.Н. Обучение нейронных сетей. М.: ПараГраф, 1990, 160 с.
8. Уоссерман Ф. Нейрокомпьютерная техника. М.: Мир, 1992, 240 с.
9. Горбань А.Н., Россиев Д.А. Нейронные сети на персональном компьютере. Новосибирск: Наука, 1996, 276 с.
10. Ежов А.А., Шумский С.А. Нейрокомпьютинг и его приложения в экономике и бизнесе. М.: МИФИ, 1998, 224 с.
11. Головкин В.А. Нейронные сети: обучение, организация и применение. М.: ИПРЖР, 2001, 256 с.
12. Круглов В.В., Борисов В.В. Искусственные нейронные сети. Теория и практика. М.: Горячая линия — Телеком, 2001, 382 с.
13. Zupan J., Gasteiger J. Neural Networks in Chemistry. Weinheim: Wiley-VCH, 1999.
14. Гальберштам Н.М., Баскин И.И., Палюлин В.А., Зефирова Н.С. Успехи химии, 2003, т. 72, № 7, с. 706.
15. Todeschini R., Consonni V. Handbook of Molecular Descriptors. Weinheim: Wiley-VCH, 2000.
16. Гиллер С.А., Глаз А.Б., Растринин Л.А., Розенблит А.Б. Докл. АН СССР, 1971, т. 199, с. 851.
17. Hiller S.A., Golender V.E., Rozenblit A.B., Glaz A.B., Rastriugin L.A. Comput. Biomed. Res., 1973, v. 6, p. 411.
18. Колмогоров А.Н. Докл. АН СССР, 1957, т. 114, с. 953.
19. Kolmogorov A.N. Am. Math. Soc. Trans., 1963, v. 28, p. 55.
20. Hecht-Nielsen R. In: IEEE First Annual Int. Conf. on Neural Networks, San Diego, 1987, v. 3, p.11.
21. Funahashi K. Neural Networks, 1989, v. 2, p. 183.
22. Kůrková V. Ibid., 1992, v. 5, p. 501.
23. Aoyama T., Suzuki Y., Ichikawa H. J. Med. Chem., 1990, v. 33, p. 905.
24. Aoyama T., Suzuki Y., Ichikawa H. Ibid., 1990, v. 33, p. 2583.
25. Aoyama T., Ichikawa H. Chem. Pharm. Bull., 1991, v. 39, p. 358.
26. Andrea T.A., Kalayeh H. J. Med. Chem., 1991, v. 34, p. 2824.
27. Wiese M. QSAR, 1991, v. 10, p. 369.
28. Aoyama T., Ichikawa H. J. Chem. Inf. Comput. Sci., 1992, v. 32, p. 492.
29. Manallack D.T., Livingstone D.J. Med. Chem. Res., 1992, v. 2, p. 181.
30. Shi L., Zhou J., Liu X., Sun D., Xu Z. Huaxue Tongbao, 1992, issue 6, p. 46.
31. Salt D.W., Yildiz N., Livingstone D.J., Tinsley C.J. Pesticide Science, 1992, v. 36, p. 161.
32. So S.S., Richards W.G. J. Med. Chem., 1992, v. 35, p. 3201.
33. Livingstone D.J., Salt D.W. Bioorg. Med. Chem. Lett., 1992, v. 2, p. 213.
34. Maggiora G.M., Elrod D.W., Trenary R.G. J. Chem. Inf. Comput. Sci., 1992, v. 32, p. 732.
35. Tetko I.V., Tanchuk V.Yu., Chentsova N.P. e. a. J. Med. Chem. 1994, v. 37, p. 2520.
36. Huuskonen J.J., Vuorela H., Hitunen R. In: Molecular Modeling and Prediction of Bioactivity. Eds. K. Gundertofte, F.S. Jorgensen. N. Y.: Kluwer Academic / Plenum Publishers, 2000, p. 377.
37. Milac A.-L., Avram S., Petrescu A.-J. Romanian Journal of Biochemistry, 2004, v. 40, p. 35.
38. Molnar L., Keseru G.M. Bioorg. Med. Chem. Lett., 2002, v. 12, p. 419.
39. Balakin K.V., Ekins S., Bugrim A. e. a. Drug Metabolism and Disposition: Biological Fate of Chemicals, 2004, v. 32, p. 1111.
40. Chohan K.K., Paine S.W., Mistry J., Barton P., Davis A.M. J. Med. Chem., 2005, v. 48, p. 5154.
41. Patankar S.J., Jurs P.C. J. Chem. Inf. Comput. Sci., 2000, v. 40, p. 706.
42. Soric M.J., Miners J.O., McKinnon R.A. e. a. Ibid., 2003, v. 43, p. 2019.
43. Polley M.J., Winkler D.A., Burden F.R. J. Med. Chem., 2004, v. 47, p. 6230.
44. Kauffman G.W., Jurs P.C. J. Chem. Inf. Comput. Sci., 2001, v. 41, p. 1553.
45. Cherqaoui D., Esseffar M., Zakarya D., Mesbah A., Villemin D. ACH-Models in Chemistry, 1998, v. 135, p. 79.
46. Yin C., Gao W., Wu K. e. a. Jishou Daxue Xuebao, Ziran Kexueban, 1999, v. 20, iss. 4, p. 1.
47. Kauffman G.W., Jurs P.C. J. Chem. Inf. Comput. Sci., 2000, v. 40, p. 753.
48. Andrea T.A. ACS Symposium Series, 1995, v. 606 (Classical and Three-Dimensional QSAR in Agrochemistry), p. 282.
49. Mager P.P. Molecular Simulation, 2002, v. 29, p. 71.
50. Beck B., Glen R.C., Clark T. Journal of Molecular Graphics, 1996, v. 14, p. 130.
51. Novic M., Nikolovska-Coleska Z., Solmajer T. J. Chem. Inf. Comput. Sci., 1997, v. 37, p. 990.
52. Braunheim B.B., Miles R.W., Schramm V.L., Schwartz S.D. Biochemistry, 1999, v. 38, p. 16076.
53. McElroy N.R., Jurs P.C. In: Abstracts of Papers, 222nd ACS National Meeting, Chicago, IL, United States, August 26–30, 2001, p. COMP-144.

54. Song X., Yu R. *Chemometrics and Intelligent Laboratory Systems*, 1993, v. 19, p. 101.
55. Bakken G.A., Jurs P.C. *J. Chem. Inf. Comput. Sci.*, 2001, v. 41, p. 1255.
56. Sutherland J.J., O'Brien L.A., Weaver D.F. *J. Med. Chem.*, 2004, v. 47, p. 5541.
57. So S.-S., Karplus M. *Ibid.*, 1997, v. 40, p. 4360.
58. Beck B., Glen R.C., Clark T. *Journal of Molecular Graphics*, 1996, v. 14, p. 130.
59. Ajay A. *J. Med. Chem.*, 1993, v. 36, p. 3565.
60. Tang Y., Wang H.-V., Chen K.-X., Ji R.Y. *Zhongguo Yaoli Xuebao*, 1995, v. 16, p. 26.
61. So S.-S., Karplus M. *J. Med. Chem.*, 1996, v. 39, p. 5246.
62. Huche M., Legendre J.J. *QSAR*, 1997, v. 16, p. 435.
63. Doucet J.P., Panaye A. *SAR and QSAR in Environmental Research*, 1998, v. 8, p. 249.
64. Burden F.R., Winkler D.A. *J. Med. Chem.*, 1999, v. 42, p. 3183.
65. Schaper K.-J. *QSAR*, 1999, v. 18, p. 354.
66. Borowski T., Krol M., Broclawik E. e. a. *J. Med. Chem.*, 2000, v. 43, p. 1901.
67. Lopez-Rodriguez M.L., Morcillo M.J., Fernandez E. e. a. *Ibid.*, 2001, v. 44, p. 198.
68. Balakin K.V., Lang S.A., Skorenko A.V. e. a. *J. Chem. Inf. Comput. Sci.*, 2003, v. 43, p. 1553.
69. Gonzalez-Arjona D., Lopez-Perez G., Gustavo-Gonzalez A. *Talanta*, 2002, v. 56, p. 79.
70. Elkhoul K., Afifi A.I., Kabbaj M., Villemain D., Cherqaoui D. *ACH-Models in Chemistry*, 2000, v. 137, p. 633.
71. Patankar S.J., Jurs P.C. In: *Abstracts of Papers, 220nd ACS National Meeting, Washington, DC, United States, August 20–24, 2000*, p. COMP-081.
72. Patankar S.J., Jurs P.C. *J. Chem. Inf. Comput. Sci.*, 2002, v. 42, p. 1053.
73. Mager P.P. *Molecular Simulation*, 2002, v. 29, p. 159.
74. Turner J.V., Cutler D.J., Spence I., Maddalena D.J. *J. Comp. Chem.*, 2003, v. 24, p. 891.
75. Fujishima S., Takahashi Y. *J. Chem. Inf. Comput. Sci.*, 2004, v. 44, p. 1006.
76. Baskin I.I., Halberstam N.M., Mukhina T.V., Palyulin V.A., Zefirov N.S. *SAR and QSAR in Environmental Research*, 2001, v. 12, p. 401.
77. Viswanadhan V.N., Mueller G.A., Basak S.C., Weinstein J.N. *J. Chem. Inf. Comput. Sci.*, 2001, v. 41, p. 505.
78. Roche O., Trube G., Zuegge J. e. a. *ChemBioChem.*, 2002, v. 3, p. 455.
79. Martinez A., Castro A., Stud M. e. a. *Med. Chem. Res.*, 1998, v. 8, p. 171.
80. Sun L., Sheng X., Yu R. *Zhongguo Huanjing Kexue*, 1994, v. 14, p. 259.
81. Xu L., Ball J.W., Dixon S.L., Jurs P.C. *Environ. Toxicology and Chemistry*, 1994, v. 13, p. 841.
82. Devillers J., Bintein S., Domine D., Karcher W. *SAR and QSAR in Environmental Research*, 1995, v. 4, p. 29.
83. Sutter J.M., Jurs P.C. *Data Handling in Science and Technology*, 1995, v. 15, p. 111.
84. Suzuki T., Ishida M. *Pharm. Sci.*, 1995, v. 1, p. 297.
85. Hatrik S., Zahradnik P. *J. Chem. Inf. Comput. Sci.*, 1996, v. 36, p. 992.
86. Free S.M., Wilson J.W. *J. Med. Chem.*, 1964, v. 7, p. 395.
87. Xie Q., Sun H., Zhou J. *Jisuanji Yu Yingyong Huaxue*, 1996, v. 13, p. 20.
88. Johnson S.R., Jurs P.C. In: *Computer-Assisted Lead Finding and Optimization: Current Tools in Medicinal Chemistry, (11th European Symposium on Quantitative Structure-Activity Relationships, Lausanne, Sept. 1–6, 1996)*, 1997, p. 31.
89. Eldred D.V., Jurs P.C. *SAR and QSAR in Environmental Research*, 1999, v. 10, p. 75.
90. Zakarya D., Boulaamail A., Larfaoui E.M., Lakhlifi T. *Ibid.*, 1997, v. 6, p. 183.
91. Sun X., Lu S. *Huanjing Kexue*, 1998, v. 19, p. 41.
92. Tang Y., Chen K.-X., Jiang H.-L., Ji R.-Y. *European Journal of Medicinal Chemistry*, 1998, v. 33, p. 647.
93. Duan Q., Cao Y., Lu S. *Tongji Yike Daxue Xuebao*, 1999, v. 28, p. 117.
94. Eldred D.V., Weikel C.L., Jurs P.C., Kaiser K.L.E. *Chemical Research in Toxicology*, 1999, v. 12, p. 670.
95. Tang G., Bai N. *Zhongguo Huanjing Kexue*, 1999, v. 19, p. 539.
96. Tang G., Bai N. *Huanjing Kexue*, 2000, v. 21, p. 89.
97. Bai N., Du Y., Xie F., Wang X. *Zhongguo Huanjing Kexue*, 2001, v. 21, p. 330.
98. Devillers J., Flatin J. *SAR and QSAR in Environmental Research*, 2000, v. 11, p. 25.
99. Devillers J. *Ibid.*, 2001, v. 11, p. 397.
100. Cronin M.T.D., Schultz T.W. *Chemical Research in Toxicology*, 2001, v. 14, p. 1284.
101. Bazoui H., Zahouily M., Sebti S., Boulajaaj S., Zakarya D. *Journal of Molecular Modeling*, 2002, v. 8, p. 1.
102. Devillers J., Pham-Delegue M.H., Decourtye A. e. a. *SAR and QSAR in Environmental Research*, 2002, v. 13, p. 641.
103. Moore D.R.J., Breton R.L., MacDonald D.B. *Environmental Toxicology and Chemistry*, 2003, v. 22, p. 1799.
104. Kaiser K.L.E. *Theochem.*, 2003, v. 622, p. 85.
105. Shoji R. *Current Computer-Aided Drug Design*, 2005, v. 1, p. 65.
106. Brinn M., Walsh P.T., Payne M.P., Bott B. *SAR and QSAR in Environmental Research*, 1993, v. 1, p. 169.
107. Ghoshal N., Mukhopadhyay S.N., Ghoshal T.K., Achari B. *Indian Journal of Chemistry, Section B: Organic Chemistry Including Medicinal Chemistry*, 1993, v. 32B, p. 1045.
108. Ghoshal N., Mukhopadhyay S.N., Ghoshal T.K., Achari B. *Bioorganic & Medicinal Chemistry Letters*, 1993, v. 3, p. 329.
109. Villemain D., Cherqaoui D., Cense J.M. *J. Chim. Phys.-chim. Biol.*, 1993, v. 90, p. 1505.
110. Qin H., Bai N. *Beijing Gongye Daxue Xuebao*, 1994, v. 20, p. 91.
111. Bai N. *Zhongguo Huanjing Kexue*, 1995, v. 15, p. 447.
112. Bai N. *Jisuanji Yu Yingyong Huaxue*, 1995, v. 12, p. 180.
113. Любимова И.К., Абилов С.К., Гальберштам Н.М. и др. *Изв. РАН, Сер. биол.*, 2001, с. 164.
114. Garg A., Bhat K.L., Bock C.W. *Dyes and Pigments*, 2002, v. 55, p. 35.
115. Adler B., Eitner B., Winterstein M. *Chemische Technik (Leipzig, Germany)*, 1993, v. 45, p. 94.
116. Villemain D., Cherqaoui D., Mesbah A. *J. Chem. Inf. Comput. Sci.*, 1994, v. 34, p. 1288.
117. Klopman G., Chakravarti S.K., Zhu H., Ivanov J.M., Saiakhov R.D. *Ibid.*, 2004, v. 44, p. 704.
118. Hemmateenejad B., Safarpour M.A., Miri R., Nesari N. *Ibid.*, 2005, v. 45, p. 190.
119. Devillers J. *SAR and QSAR in Environmental Research*, 1993, v. 1, p. 161.
120. Gu L., Wang L., Huang W., Yang H. *Chinese Science Bulletin*, 1995, v. 40, p. 871.
121. Zakarya D., Larfaoui E.M., Boulaamail A., Lakhlifi T. *SAR and QSAR in Environmental Research*, 1996, v. 5, p. 269.

122. Sun W., Chen D., Chen Y., Hu S. Gaodeng Xuexiao Huexue Xuebao, 1998, v. 19, p. 871.
123. Chen Y., Chen D., He C., Hu S. Chemometrics and Intelligent Laboratory Systems, 1999, v. 45, p. 267.
124. Zhang L., Yu H., Chen D., Hu S. Fenxi Huexue, 2004, v. 32, p. 1590.
125. Shi Y., Huang C., Hou C. Jisuanji Yu Yingyong Huexue, 2005, v. 22, p. 153.
126. Zakarya D., Larfaoui E.M., Boulaatail A., Tollabi M., Lakh-lifi T. Chemosphere, 1998, v. 36, p. 2809.
127. Yang G., Yang H. Huazhong Shifan Daxue Xuebao Ziranhexueban, 2001, v. 35, p. 40.
128. Huang M., Ning B., Su L., Shang Z. Jisuanji Yu Yingyong Huexue, 2002, v. 19, p. 519.
129. Sumi K., Kikkawa N., Yorifuji T. e. a. In: Comput. Aided Innovation New Mater. 2, Proc. 2nd Int. Conf. Exhib. Comput. Appl. Mater. Mol. Sci. Eng., 1993, Pt. 2, p. 875.
130. He L., Qian X. Jisuanji Yu Yingyong Huexue, 1996, v. 13, p. 97.
131. Sparks T.C., Crouse G.D., Durst G. Pest. Management Science, 2001, v. 57, p. 896.
132. Domine D., Devillers J., Chastrette M., Karcher W. SAR and QSAR in Environmental Research, 1993, v. 1, p. 211.
133. Liu G., Song X., Yu J., Qian X. Huanjing Kexue Xuebao, 2002, v. 22, p. 359.
134. Chastrette M., De Saint Laumer J.Y. European Journal of Medicinal Chemistry, 1991, v. 26, p. 829.
135. Chastrette M., Cretin D., El Aidi C. J. Chem. Inf. Comput. Sci., 1996, v. 36, p. 108.
136. Chastrette M., El Aidi C., Cretin D. SAR and QSAR in Environmental Research, 1997, v. 7, p. 233.
137. Zakarya D., Cherqaoui D., Esseffar M., Villemin D., Cense J.-M. J. Phys. Org. Chem., 1997, v. 10, p. 612.
138. Cherqaoui D., Esseffar M., Villemin D. e. a. New Journal of Chemistry, 1998, v. 22, p. 839.
139. Liu R., Sun H., So S.-S. J. Chem. Inf. Comput. Sci., 2001, v. 41, p. 1623.
140. Dorransoro I., Chana A., Abasolo M.I. e. a. QSAR & Combinatorial Science, 2004, v. 23, p. 89.
141. Engkvist O., Wrede P., Rester U. J. Chem. Inf. Comput. Sci., 2003, v. 43, p. 155.
142. Bodor N., Harget A., Huang M.-J. J. Am. Chem. Soc., 1991, v. 113, p. 9480.
143. Bodor N., Huang M.-J., Harget A. Theochem., 1994, v. 115, p. 259.
144. Баскин И.И., Палюлин В.А., Зефирова Н.С. Докл. АН, 1993, т. 332, с. 713.
145. Egoľf L.M., Jurs P.C. J. Chem. Inf. Comput. Sci., 1993, v. 33, p. 616.
146. Petersen R., Fredenslund A., Rasmussen P. Computers & Chemical Engineering, 1993, v. 18, p. S63.
147. Egoľf L.M., Wessel M.D., Jurs P.C. J. Chem. Inf. Comput. Sci., 1994, v. 34, p. 947.
148. Suzuki T., Ebert R.-U., Schüürmann G. Ibid., 1997, v. 37, p. 1122.
149. Артеменко Н.В., Баскин И.И., Палюлин В.А., Зефирова Н.С. Докл. АН, 2001, т. 381, с. 203.
150. Артеменко Н.В., Баскин И.И., Палюлин В.А., Зефирова Н.С. Изв. РАН, Сер. хим., 2003, с. 19.
151. Kauffman G.W., Jurs P.C. J. Chem. Inf. Comput. Sci., 2001, v. 41, p. 408.
152. Liang C., Gallagher D.A. Ibid., 1998, v. 38, p. 321.
153. Grunenberg J., Herges R. Ibid., 1995, v. 35, p. 905.
154. Pompe M., Novic M. Ibid., 1999, v. 39, p. 59.
155. Nefati H., Diawara B., Legendre J.J. SAR and QSAR in Environmental Research, 1993, v. 1, p. 131.
156. Nefati H., Cense J.-M., Legendre J.-J. J. Chem. Inf. Comput. Sci., 1996, v. 36, p. 804.
157. Charlton M.H., Docherty R., Hutchings M.G. J. Chem. Soc. Perkin Trans. 2, 1995, p. 2023.
158. Жохова Н.И., Баскин И.И., Палюлин В.А., Зефирова А.Н., Зефирова Н.С. Изв. РАН, Сер. хим., 2003, с. 1787.
159. Жохова Н.И., Баскин И.И., Палюлин В.А., Зефирова А.Н., Зефирова Н.С. Ж. структурной химии, 2004, т. 45, с. 660.
160. Жохова Н.И., Баскин И.И., Палюлин В.А., Зефирова А.Н., Зефирова Н.С. Ж. прикладной химии, 2005, т. 78, с. 1034.
161. Kraenz H., Vill V., Meyer B. J. Chem. Inf. Comput. Sci., 1996, v. 36, p. 1173.
162. Mattioni B.E., Jurs P.C. Ibid., 2002, v. 42, p. 232.
163. Sun H., Tang Y., Wu G. Macromolecular Research, 2002, v. 10, p. 13.
164. Kim Y.S., Kim J.H., Kim J.S., No K.T. J. Chem. Inf. Comput. Sci., 2002, v. 42, p. 75.
165. Schweitzer R.C., Morris J.B. Analytica Chimica Acta, 1999, v. 384, p. 285.
166. Svozil D., Sevcik J.G.K., Kvasnicka V. J. Chem. Inf. Comput. Sci., 1997, v. 37, p. 338.
167. Mitchell B.E., Jurs P.C. Ibid., 1997, v. 37, p. 538.
168. English N.J., Carroll D.G. Ibid., 2001, v. 41, p. 1150.
169. Yaffe D., Cohen Y. Ibid., 2001, v. 41, p. 463.
170. Chalk A.J., Beck B., Clark T. Ibid., 2001, v. 41, p. 1053.
171. Suzuki T., Ebert R.-U., Schüürmann G. Ibid., 2001, v. 41, p. 776.
172. Гальберштам Н.М., Баскин И.И., Палюлин В.А., Зефирова Н.С. Докл. АН, 2002, т. 384, с. 202.
173. Brewster M.E., Huang M.-J., Harget A., Bodor N. Tetrahedron, 1992, v. 48, p. 3463.
174. Баскин И.И., Аїт А.О., Гальберштам Н.М. и др. Докл. АН, 1997, т. 357, с. 57.
175. Duan X.-M., Li Z.-H., Wang W.-N. e. a. Chemical Physics Letters, 2005, v. 410, p. 125.
176. Barden C.J., Boysworth M.K., Palocsay F.A. J. Chem. Inf. Comput. Sci., 1998, v. 38, p. 483.
177. Liu L., Guo Q.-X. Ibid., 1999, v. 39, p. 133.
178. Loukas Y.L. International Journal of Pharmaceutics, 2001, v. 226, p. 207.
179. Qi Y.-H., Zhang Q.-Y., Xu L. J. Chem. Inf. Comput. Sci., 2002, v. 42, p. 1471.
180. Doucet J.P., Panaye A., Feuilleau Bois E., Ladd P. Ibid., 1993, v. 33, p. 320.
181. Meiler J., Meusinger R., Will M. Ibid., 2000, v. 40, p. 1169.
182. West G.M.J. Ibid., 1993, v. 33, p. 577.
183. Binev Y., Corvo M., Aires-de-Sousa J. Ibid., 2004, v. 44, p. 946.
184. Zhang X. Wuli Huexue Xuebao, 1996, v. 12, p. 845.
185. Qi Y.-H., Xu L. Yingyong Huaxue, 2002, v. 19, p. 1054.
186. Qi Y.-H., Zhang Q.-Y., Luo C.-C., Wang J., Xu L. Gaodeng Xuexiao Huaxue Xuebao, 2004, v. 25, p. 1100.
187. Jover J., Bosque R., Sales J. J. Chem. Inf. Comput. Sci., 2004, v. 44, p. 1098.
188. Kvasnicka V., Sklenak S., Pospichal J. J. Am. Chem. Soc., 1993, v. 115, p. 1495.
189. Баскин И.И., Палюлин В.А., Зефирова Н.С. Докл. АН, 1993, т. 333, с. 176.
190. Baskin I.I., Palyulin V.A., Zefirov N.S. J. Chem. Inf. Comput. Sci., 1997, v. 37, p. 715.
191. Tetko I.V. Ibid., 2002, v. 42, p. 717.