

Прогноз спектра биологической активности органических соединений

Д. А. Филимонов, В. В. Поройков

ДМИТРИЙ АЛЕКСЕЕВИЧ ФИЛИМОНОВ — кандидат физико-математических наук, ведущий научный сотрудник лаборатории структурно-функционального конструирования лекарств ГУ НИИ биомедицинской химии им. В.Н. Ореховича РАМН. Область научных интересов: биоинформатика и компьютерное конструирование лекарств.

ВЛАДИМИР ВАСИЛЬЕВИЧ ПОРОЙКОВ — кандидат физико-математических наук, доктор биологических наук, профессор, заместитель директора и заведующий лабораторией структурно-функционального конструирования лекарств ГУ НИИ биомедицинской химии им. В.Н. Ореховича РАМН. Область научных интересов: биоинформатика и компьютерное конструирование лекарств.

119121 Москва, ул. Погодинская, 10, ГУ НИИ биомедицинской химии, тел. (495)247-30-29, факс (495)245-08-57, E-mail dmitry.filimonov@ibmc.msk.ru

В настоящее время преобладает направленный подход к поиску и созданию новых лекарств: химические соединения тестируются лишь на небольшое число требуемых видов биологической активности, и свойства выявленных базовых структур оптимизируются путем синтеза и исследования их аналогов. При этом многие виды биологической активности, присущие изучаемому веществу, но являющиеся побочными по отношению к избранному направлению исследований, остаются неизученными. Но каждое вещество способно проявлять несколько видов биологической активности. Некоторые из них обнаруживаются впоследствии как побочные токсические эффекты, другие становятся основанием для регистрации лечебного препарата по новому назначению. Так, например, ацетазоламид был предложен в качестве диуретика в 1954 г. и как противоэпилептическое средство в 1956 г.; левамизол — как антигельминтное средство в 1968 г. и как иммуностимулятор в 1980 г.; алпростадил — как антиагрегант в 1988 г. и как эректант в 1994 г.; анальгетическое действие аспирина было выявлено в 1899 г., а его антиагрегантные свойства были открыты лишь в 1971 г., и т.д. Таким образом, имеется определенное противоречие между жесткой направленностью исследования новых биологически активных соединений и множественностью биологических эффектов, которые каждое вещество потенциально способно вызывать.

Биологическая активность есть результат взаимодействия вещества с биологическим объектом. Она зависит от характеристик вещества (структуры его молекул и физико-химических свойств), биологического объекта (вида, пола, возраста) и способа воздействия (пути введения, дозы). Весь комплекс биологических эффектов, которые вещество способно вызывать при некоторых условиях взаимодействия с биологическими объектами, без учета особенностей конкретных экспериментов, будем называть спектром биологической активности вещества. Это качественное характеристическое свойство вещества, зависящее только от структуры его молекул. С другой стороны, большой массив данных можно собрать, лишь только используя много разных источников, поскольку информация из единичной публикации никогда не охватывает всех

аспектов биологического действия описываемого в нем вещества. Например, по справочнику [1] кофеин (номер по CAS 58-08-2) является стимулятором, аналептиком и диуретиком. В результате специально проведенного информационного поиска найдено, что кофеин обладает следующими видами биологической активности:

«Adenosine deaminase inhibitor (ингибитор аденозиндеаминазы)», «Analeptic (аналептик)», «Arhythmic (аритмогенный)», «ATP diphosphatase inhibitor (ингибитор АТФ дифосфотазы)», «Carcinogenic (канцерогенный)», «Carcinogenic, group 3 (канцерогенный, 3-я группа по классификации МАИР)», «Cardiotonic (кардиотоник)», «Convulsant (вызывающий конвульсии)», «Cyclic AMP phosphodiesterase inhibitor (ингибитор циклической АМФ фосфодиэстеразы)», «dATP(dGTP)-DNA purinetransferase inhibitor (ингибитор dATP(dGTP)-ДНК пуринтрансферазы)», «Diuretic (диуретик)», «Embryotoxic (эмбриотоксичный)», «Glycogen (starch) synthase inhibitor (ингибитор гликоген (крахмал) синтазы)», «Guanylate cyclase inhibitor (ингибитор гуанилатциклазы)», «Hydroxyacylglutathione hydrolase inhibitor (ингибитор гидроксиацилглутатионгидролазы)», «Hypertensive (гипертензивный)», «Immunosuppressant (иммуносупресант)», «Lactoylglutathione lyase inhibitor (ингибитор лактоилглутатионлиазы)», «Non mutagenic, Salmonella (не мутагенный, сальмонелла)», «Nucleotide metabolism regulator (регулятор нуклеотидного метаболизма)», «Phosphatidylinositol kinase inhibitor (ингибитор фосфатидилинозитолкиназы)», «Phosphodiesterase inhibitor (ингибитор фосфодиэстеразы)», «Phosphorylase inhibitor (ингибитор фосфорилазы)», «Psychostimulant (психостимулятор)», «Purine nucleosidase inhibitor (ингибитор пури-нуклеозидазы)», «Respiratory analeptic (аналептик респираторный)», «Saluretic (салуретик)», «Spasmogenic (спазмогенный)», «Spasmolytic (спазмолитик)», «Teratogen (тератоген)», «Thymidine kinase inhibitor (ингибитор тимидинкиназы)», «Toxic (токсичный)», «Urate oxidase inhibitor (ингибитор уратоксидазы)», а также является субстратом следующих семейств цитохромов P450: CYP1A, CYP1A1, CYP1A2, CYP2, CYP2B, CYP2B1, CYP2B2, CYP2D, CYP2D6, CYP2E, CYP2E1, CYP3A, CYP3A1, CYP3A4, CYP3A5.

Некоторые кажущиеся противоречия в представленных здесь видах биологической активности кофеина обусловлены противоположным действием его в разных дозировках.

Практически невозможно исследовать экспериментально на все известные виды активности ни одно химическое соединение [2]. Даже если принять во внимание возможности современного высокопроизводительного скрининга, который также осуществляется направленно по отношению к одной или нескольким биологическим мишеням действия будущих лекарств, рассматриваемых как перспективные в конкретный период времени [3]. Реальную возможность комплексного исследования биологической активности веществ может обеспечить развитие новых технологий компьютерного прогнозирования и их применение для оценки вероятных видов активности химических соединений с последующим тестированием изучаемых веществ в соответствии с результатами прогноза. Большая часть существующих в настоящее время компьютерных методов молекулярного моделирования и анализа связи «структура—активность» (Structure—Activity Relationships — SAR) применяется для исследования взаимодействия «лиганд—рецептор» (одна молекула—мишень) и оптимизации свойств базовых структур на основе анализа количественных соотношений «структура—активность» (Quantitative Structure—Activity Relationship — QSAR) в рамках одного узкого химического класса [4—9]. Методы молекулярного сходства [10, 11] и кластеризации [11, 12] также могут использоваться для разделения соединений на структурные группы [13], прогноза биологической активности и отбора веществ [14]. Существующие методы химической информатики практически исчерпывающе охватывает четырехтомное издание под редакцией Йохана Гастайгера «Handbooks of Cheminformatics: From Data to Knowledge» [9].

Преодоление разрыва между единичными активностями, которые принимаются во внимание на каждой стадии традиционных исследований, и многочисленностью видов биологической активности органических соединений возможно с помощью компьютерной системы, позволяющей прогнозировать большое число вероятных видов биологической активности вещества на основе его структурной формулы с использованием единогообразного описания химической структуры и универсального математического алгоритма установления зависимости «структура—активность». Попытки создания подобной системы неоднократно предпринимались ранее [15—24]. Возможности компьютерного прогнозирования биологической активности веществ на основе их структурных формул изучались в СССР в рамках Государственной системы регистрации всех вновь синтезированных химических соединений в течение многих лет [25]. Эта задача не была решена тогда полностью по ряду объективных и субъективных причин, однако выполненные при этом исследования обеспечили базу для создания такой компьютерной системы в будущем [26—29].

Нами разработана и развивается компьютерная система PASS (*Prediction of Activity Spectra for Substances — прогноз спектров биологической активности органических соединений*) [30—37], основанная на анализе взаимосвязей «структура—активность» с использованием обучающей выборки, содержащей большое ко-

личество разнородных химических соединений с различными видами биологической активности. Точность прогноза зависит от нескольких факторов [38], но в настоящее время наиболее важным в этом подходе представляется качество обучающей выборки. «Идеальная» обучающая выборка должна включать биологически активные вещества с исчерпывающей информацией об их биологической активности. Другими словами, для каждого вещества, входящего в выборку, должен быть полностью изучен весь спектр биологической активности. На самом деле не существует ни одной достаточно обширной базы данных химических соединений, которые были бы исследованы на все виды биологической активности: неполнота информации о биологической активности присуща всем базам данных. Более того, полное экспериментальное исследование принципиально невозможно [2]: для испытания на все известные в настоящее время виды биологической активности всех известных веществ (а их более 25 млн. — см. <http://www.cas.org>) не хватит ни ресурсов всей биосферы, ни ресурсов всего человечества — общее количество требуемых для этого исследований составит десятки миллиардов.

Из изложенного выше важнейшие требования к прогнозу биологической активности химических соединений можно сформулировать следующим образом:

- большое количество и разнообразие прогнозируемых видов биологической активности;
- использование реальной обучающей выборки, далекой по полноте информации от идеальной;
- способность прогнозировать биологическую активность соединений различных химических классов с приемлемой точностью;
- использование минимальной информации о веществе, достаточной для прогноза спектров активности новых и даже еще не синтезированных, а только планируемых к синтезу химических соединений;
- быстрое действие, позволяющее за приемлемое время выполнять прогноз и анализ спектров биологической активности сотен тысяч и миллионов химических соединений.

К настоящему времени компьютерная система PASS в значительной степени удовлетворяет этим требованиям. Ее основными составляющими являются:

- представление биологической активности;
- описание структуры химических соединений;
- база данных и знаний о взаимосвязях «структура—активность» — SAR base;
- алгоритм прогноза спектров биологической активности.

Система PASS постоянно развивается. Актуальная на момент написания данной статьи версия 1.917, завершенная в сентябре 2005 года, имеет ряд существенных отличий от ранее описанных версий.

Представление биологической активности

Биологическая активность представлена в PASS качественно в виде спектра биологической активности химического соединения. Для каждого соединения имеется список тех видов активности, которые оно способно проявлять при соответствующих условиях. При этом мы следуем принципу «презумпции неви-

новности»: в PASS принимается, что вещество не обладает теми видами биологической активности, которые не указаны в его спектре. Хотя нельзя исключить ситуации, когда информация о какой-либо активности вещества не была найдена в доступных источниках, либо оно обладает некоторой биологической активностью, но на эту активность вещество еще не испытывалось. Это приближение не оказывает существенного влияния на результаты анализа взаимосвязей «структура—активность» и выполняемого на этой основе прогноза благодаря статистической устойчивости используемого в PASS алгоритма.

Необходимо подчеркнуть, что для прогноза с помощью PASS может быть использован любой способ объективной классификации органических соединений. Если соответствующие классы действительно определяются особенностями структуры молекул, то прогноз принадлежности к этим классам может быть вполне успешным. Например, интервал значений некоторой количественной величины можно рассматривать в PASS как «активность»: если значение величины принадлежит этому интервалу, то вещество «активно», и «неактивно» в иных случаях. Поэтому ясно, что применимость PASS гораздо шире прогноза только спектра биологической активности.

Описание химической структуры

Результат воздействия вещества на биологический объект, при прочих равных условиях, определяется структурой его молекул. В химической практике вещество описывают структурной формулой и набором его физико-химических параметров. При анализе взаимосвязи «структура—активность» используется множество различных характеристик химических соединений: структурные фрагменты, геометрические и топологические индексы, физико-химические параметры и т.д. Для разных видов биологической активности в SAR/QSAR для разных рядов химических соединений конкретные дескрипторы, описывающие структуру молекул, оказываются более или менее значимыми [4–9]. При прогнозировании одновременно большого количества разнообразных видов биологической активности для разнородных органических соединений важно использовать такое описание химической структуры, которое обеспечивало бы получение достаточно точных непротиворечивых результатов, но не было бы чувствительно к случайным зависимостям.

Структурная формула, традиционно записываемая в соответствии с номенклатурными правилами в химии, отражает атомный состав и взаимное расположение атомов в молекуле. Однако реально вещество представляет собой смесь компонент различного характера: молекулы и ионы в разных конформационных и электронных состояниях, таутомеры, комплексы молекул разного вида и степени устойчивости, и т.д. Эта смесь находится в динамическом квазиравновесии и ее состав существенно зависит от внешних условий. Структурная же формула характеризует лишь одну компоненту, условно принимаемую за основную. Но, с другой стороны, все состояния молекулы однозначно связаны между собой, хотя эти зависимости могут быть и довольно сложными. Поэтому, в принципе, любое состояние молекулы может быть использовано для описания ее структуры. Основное состояние электронной подсистемы свободной молекулы в

вакууме однозначно определяется зарядами и положениями ядер атомов, а равновесные координаты ядер соответствуют минимуму полной энергии молекулы [39, 40]. Поэтому с учетом сказанного выше о взаимосвязи различных состояний молекулы можно предложить следующую формулировку: *взаиморасположение в пространстве атомов свободной молекулы в основном состоянии в вакууме — необходимая и достаточная характеристика ее структуры.*

В менее идеализированных условиях, при учете взаимодействия с другими молекулами — мишенями, рецепторами, ферментами, — вместо координат атомов надо рассматривать функцию распределения координат атомов, которую не сразу ясно не только, как вычислять, но и в каком виде представить. В такой ситуации более удобным для практического использования описанием структуры является выборка из соответствующей генеральной совокупности достаточно большого набора координат атомов, характеризующая разнообразие конформационных состояний молекулы.

Однако при использовании такого способа описания структуры молекул требуются значительные вычислительные ресурсы для проведения соответствующих квантовохимических расчетов, и возникает непростая и еще не решенная проблема инвариантного к преобразованиям координат полного описания. Но вместе с тем в основе таких расчетов всегда лежит традиционная структурная формула. И в этом смысле можно сказать, что структурная формула однозначно определяет свойства молекулы.

Поэтому в PASS в настоящее время в качестве основы для описания структуры органических соединений выбрана именно структурная формула. Это соответствует задаче прогноза спектров активности новых органических соединений, причем не только экспериментально не исследованных, но даже еще не синтезированных, а только планируемых к синтезу. Стереохимические особенности молекул не учитываются, несмотря на очевидную зависимость от них биологической активности. Это обусловлено невозможностью обеспечить полноту информации о стереохимии молекул для достаточно большой выборки, а также тем, что единой универсальной стереохимической номенклатуры нет, и наиболее адекватным способом представления стереохимической информации является только пространственная структура молекул.

Первоначально в PASS [26–29] использовался предложенный Виктором Вениаминовичем Авидоном фрагментарный код суперпозиции подструктур (ФКСП) [15, 16, 18, 19, 41]. Суть его состоит в том, что в структуре молекулы выделяются значимые с точки зрения биологической активности дескрипторные центры: гетероатомы с учетом заряда и гибридизации орбиталей, метильные группы и т.п., из которых образуются фрагменты в виде пары дескрипторных центров и кратчайшей цепи между ними. В языке ФКСП имеются также циклические дескрипторы в виде нотации циклов или полициклических систем и входящих в них гетероатомов [18, 19, 41]. Но это описание имеет ряд недостатков. Как и многие другие дескрипторы, язык ФКСП является эвристическим, отражающим скорее абстрагирование структурной химической информации в сознании человека, чем явление биологической активности. Список дескрипторных центров соответствует конкретным представле-

нием его авторов о лиганд-рецепторных взаимодействиях, которые к тому же не исчерпывают все многообразие взаимодействий химических соединений с биологическими объектами. Многие дескрипторы ФКСП оказываются уникальными — встречаются в единственной структуре, — и поэтому, по существу, бесполезны для прогноза. Некоторые молекулы описываются двумя—тремя кодами ФКСП, тогда как другие, незначительно отличающиеся от них, — сотнями кодов ФКСП.

Недостатки языка ФКСП, выявившиеся в ходе его использования, было предложено преодолеть с помощью разработки новых структурных дескрипторов. Сначала эти дескрипторы получили название «молекулярный базис атомных окрестностей», теперь же они известны под названием «MNA-дескрипторы» (Multilevel Neighborhoods of Atoms) — многоуровневые атомные окрестности [30—32, 37]. MNA-дескрипторы основаны на таком представлении структурной формулы, в котором, согласно валентностям и зарядам атомов, явно указаны все атомы водорода и не учитываются типы связей: природа не знает, что такое «стертые водороды», а кратность связей во многих случаях на самом деле должна быть дробной — например, в ароматическом кольце или в группе —NO₂, — можно лишь утверждать, имеется ли между данными двумя атомами достаточно устойчивая химическая связь или нет. В таком виде структурная формула становится однозначной даже формально — она не зависит, например, от альтернативных способов изображения ароматических систем.

На основе описанного представления структурной формулы MNA-дескрипторы для каждого атома молекулы строятся рекурсивно следующим образом:

MNA-дескриптор 0-го уровня — метка *A* самого атома;

MNA-дескриптор любого следующего уровня — условное обозначение структурного фрагмента $A(D_1 D_2 \dots D_i \dots)$, где D_i — MNA-дескриптор предыдущего уровня для *i*-го непосредственного соседа данного атома с меткой *A*. Дескрипторы соседей $D_1 D_2 \dots D_i \dots$ записываются в каком-нибудь однозначном порядке, например лексикографическом.

Эта итерационная процедура может быть продолжена до любого уровня. Важно подчеркнуть, что метки атомов могут не только соответствовать общепринятым символам химических элементов, но и включать любую дополнительную информацию, например, о принадлежности атома к цепи или к какой-либо циклической системе, или что он является реакционным центром биотрансформации.

Структура молекулы в PASS версии 1.917 представлена как неповторное множество MNA-дескрипторов 1-го и 2-го уровней. В дескрипторах 2-го уровня используется индикатор «—» для обозначения атомов в цепях.

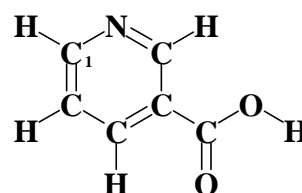
На рис. 1 представлена структура никотиновой кислоты. Для атома C₁ получается следующая последовательность MNA-дескрипторов:

- уровень 0: C
- уровень 1: C(CN—H)
- уровень 2: C(C(CC—H)N(CC)—H(C))
- уровень 3: C(C(C(CC—H)C(CN—H)—H(C))N(C(CN—H)C(CN—H))—H(C(CN—H)))

На первый взгляд в MNA-дескрипторах нет ничего нового — они похожи на очень многие уже известные структурные дескрипторы, особенно описанные В.М. Татевским [42]. Это не удивительно, поскольку, благодаря многолетнему широкому интересу к SAR/QSAR, уже предложены многие десятки, если не сотни, классов структурных дескрипторов, и процесс «придумывания» новых дескрипторов едва ли может быть хотя бы замедлен. История разработки MNA-дескрипторов такова, что поиск основополагающих идей для их конструирования привел к квантовой химии и далее — к теории квантованных полей: каждый MNA-дескриптор строится так же, как соответствующий член разложения в ряд квантовополевой функции Грина молекулы по функциям Грина составляющих ее атомов, но отличается тем, что создается лишь последовательность символов, а не сложный математический объект. Поэтому MNA-дескрипторы имеют фундаментальную основу и ясны пути их развития в направлении более полного учета квантовохимического описания структуры молекул, например, переход к пространственной структуре молекулы, при наличии необходимых для этого вычислительных и информационных ресурсов.

Важной особенностью MNA-дескрипторов является их открытость — дескрипторы порождаются на основе самой структурной формулы, а не на основе какого-либо заранее составленного списка структурных фрагментов. Другая их особенность заключается в сохранении целостности фрагментов структуры в том смысле, что для каждого MNA-дескриптора можно, при некотором навыке, изобразить соответствующий ему фрагмент.

В PASS особое значение имеет понятие эквивалентности структур: структуры считаются эквивалентными, если они описываются одним и тем же множе-



HC	C(C(CC—H)C(CC—C)—H(C))
HO	C(C(CC—H)C(CN—H)—H(C))
CHCC	C(C(CC—H)C(CN—H)—C(C—O—O))
CHCN	C(C(CC—H)N(CC)—H(C))
CCCC	C(C(CC—C)N(CC)—H(C))
CCOO	N(C(CN—H)C(CN—H))
NCC	—H(C(CC—H))
ОНС	—H(C(CN—H))
ОС	—H(—O(—H—C))
	—C(C(CC—C)—O(—H—C)—O(—C))
	—O(—H(—O)—C(C—O—O))
	—O(—C(C—O—O))

Рис. 1. Структурная формула никотиновой кислоты и используемое в PASS представление в виде множества MNA-дескрипторов 1-го и 2-го уровней

ством MNA-дескрипторов. В SAR base включаются только уникальные структуры. Поскольку MNA-дескрипторы не отражают стереохимических особенностей молекулы, структуры, которые имеют только стереохимические различия, рассматриваются как эквивалентные.

База данных и знаний SAR base

Для прогноза в PASS используется SAR base, которая создается на основе анализа обучающей выборки, содержащей структурные формулы и спектры активности органических соединений. SAR base включает в себя словарь названий видов биологической активности, словарь MNA-дескрипторов, описания структур и активностей веществ из обучающей выборки, данные и знания о взаимосвязях «структура—биологическая активность».

При включении обучающей выборки в SAR base для каждого вещества выборки генерируются MNA-дескрипторы. Если структура молекулы не полностью определена, т.е. содержит неопределенный атом или остаток, то вещество не включается в SAR base. Если в SAR base обнаруживается эквивалентная структура, то имеющийся спектр активности вещества дополняется новыми активностями. Данные и знания о взаимосвязях «структура—биологическая активность» генерируются в ходе описанной ниже процедуры обучения.

В PASS версии 1.917 SAR base содержит 57978 описаний структуры молекул и спектра активности лекарств, лекарственно-подобных и биологически активных веществ. Словарь MNA-дескрипторов включает 44041 дескриптор 1-го и 2-го уровней.

В разных источниках информации биологические активности веществ описаны неодинаковыми терминами. Поэтому спектры активности в обучающей выборке были стандартизованы. В SAR base версии 1.917 общее количество разных видов биологической активности равно 4463, но 769 из них представлены только одним соединением, 504 — двумя, а тремя и более — 3190. В список прогнозируемых по умолчанию видов активности включено 2005 названий, из них 224 — это фармакологические эффекты, 1756 — молекулярные механизмы действия и 25 — различные побочные эффекты и виды специфической токсичности. Средняя точность их прогноза составляет 88% по скользящему контролю с исключением по одному.

Алгоритм прогноза биологической активности

В SAR и QSAR используются методы логического вывода на основе классической, индуктивной, вероятностной, нечеткой и других логик, методы анализа сходства и кластеризации, методы линейной, нелинейной и непараметрической регрессии, искусственные нейронные сети, статистические методы анализа и принятия решений [4—14, 20—24] и т.д. Алгоритм прогноза PASS отобран среди исследованных в течение десятков лет сотен различных вариантов [26, 32, 43]. Его описание удобнее всего выполнить на основе ставшего уже классическим байесовского подхода, который можно сформулировать следующим образом.

Для химического соединения C по его структуре, записанной в виде множества из m MNA-дескрипторов $\{D_1, \dots, D_m\}$, оценим вероятность $P(A_k|C)$

того, что соединение C имеет активность A_k . Согласно формуле Байеса:

$$P(A_k|C) = P(C|A_k) \cdot P(A_k) / P(C)$$

где $P(C|A_k)$ — вероятность структуры C при условии, что химическое соединение имеет активность A_k ; $P(A_k)$ — априорная вероятность активности A_k ; $P(C)$ — априорная вероятность структуры C .

Эту формулу более удобно записать в виде отношения правдоподобия — отношения вероятностей «имеет» и «не имеет» активность A_k структура C :

$$P(A_k|C) / P(\bar{A}_k|C) = [P(C|A_k) \cdot P(A_k)] / [P(C|\bar{A}_k) \cdot P(\bar{A}_k)] \quad (1)$$

где $P(\bar{A}_k|C)$, $P(C|\bar{A}_k)$, $P(\bar{A}_k)$ — соответствующие вероятности для \bar{A}_k — отсутствия активности A_k .

Если допустить, что дескрипторы D_1, \dots, D_m независимы в совокупности, то можно записать вероятности $P(C|A_k)$ и $P(C|\bar{A}_k)$ как произведения условных вероятностей для отдельных дескрипторов:

$$P(C|A_k) = P(D_1, \dots, D_m|A_k) = \prod_i P(D_i|A_k)$$

$$P(C|\bar{A}_k) = P(D_1, \dots, D_m|\bar{A}_k) = \prod_i P(D_i|\bar{A}_k)$$

Эти соотношения приближенные, поскольку MNA-дескрипторы заведомо являются зависимыми в силу способа их построения. Но у нас нет приемлемых альтернатив, и нам остается лишь не забывать о приближенности получаемых формул.

Подстановка этого приближения в (1) дает следующее выражение для отношения правдоподобия:

$$P(A_k|C) / P(\bar{A}_k|C) = [P(A_k) / P(\bar{A}_k)] \cdot \prod_i [P(D_i|A_k) / P(D_i|\bar{A}_k)]$$

или для логарифма отношения правдоподобия в виде суммы логарифмов:

$$\ln [P(A_k|C) / P(\bar{A}_k|C)] = \ln [P(A_k) / P(\bar{A}_k)] + \sum_i \ln [P(D_i|A_k) / P(D_i|\bar{A}_k)]$$

Использование формулы Байеса для отношений $P(D_i|A_k) / P(D_i|\bar{A}_k)$ дает выражение:

$$\ln [P(A_k|C) / P(\bar{A}_k|C)] = \ln [P(A_k) / P(\bar{A}_k)] + \sum_i \{ \ln [P(A_k D_i) / P(\bar{A}_k D_i)] - \ln [P(A_k) / P(\bar{A}_k)] \}$$

или, поскольку $P(\bar{A}_k) = 1 - P(A_k)$, $P(\bar{A}_k D_i) = 1 - P(A_k D_i)$ и $P(\bar{A}_k|C) = 1 - P(A_k|C)$, находим:

$$\ln [P(A_k|C) / (1 - P(A_k|C))] = \ln [P(A_k) / (1 - P(A_k))] + \sum_i \{ \ln [P(A_k D_i) / (1 - P(A_k D_i))] - \ln [P(A_k) / (1 - P(A_k))] \} \quad (2)$$

Смысл полученного выражения (2) вполне прозрачен: логарифм отношения правдоподобия есть сумма логарифма априорного отношения правдоподобия и суммы вкладов отдельных дескрипторов. При этом, если активность не зависит от данного дескриптора, то $P(A_k D_i) = P(A_k)$ и этот дескриптор не влияет на результат — его вклад в сумму нулевой. Это и есть классический результат вероятностного подхода. Но, помимо уже отмеченной приближенности, этот результат имеет и другой существенный, хорошо известный недостаток: вклад некоторых дескрипторов, для которых условная вероятность активности при их наличии в структуре слишком близка к 0 или 1, стремится к бесконечности и подавляет все остальные члены суммы. Это особенно сильно проявляется, когда для вероятностей $P(A_k|D_i)$ используются частотные

оценки по результатам анализа обучающей выборки и значения 0 и 1 — скорее правило, чем исключение.

Для преодоления этого недостатка можно предложить много разных подходов, и они были апробированы в ходе развития PASS. Наилучший результат дал применение вместо $\ln[p/(1-p)]$ так называемого арксинусного преобразования Фишера $\arcsin(2p-1)$: почти на всем интервале изменения p их форма совпадает, но значения арксинуса ограничены величиной $\pm\pi/2$. Точность прогноза повысилась также после замены суммы вкладов дескрипторов их средним значением, что, видимо, компенсирует допущение о независимости дескрипторов. Логарифм априорного отношения правдоподобия [первый член в правой части уравнения (2)] не несет информации о конкретном прогнозируемом веществе и может быть опущен.

Описанный выше байесовский подход поясняет, почему алгоритм прогноза PASS основан на следующей специальной B -статистике: по структуре молекул химического соединения, записанной в виде множества из m MNA-дескрипторов $\{D_1, \dots, D_m\}$, для каждой активности A_k подсчитываются величины B_k :

$$\begin{aligned} B_k &= (S_k - S_{0k}) / (1 - S_k \cdot S_{0k}); \\ S_k &= \sin[\sum_i \arcsin(2P(A_k|D_i) - 1) / m]; \\ S_{0k} &= 2P(A_k) - 1 \end{aligned} \quad (3)$$

При этом, для каждого вида активности, если для всех дескрипторов $P(A_k|D_i) = 1$, то $B_k = 1$; если для всех дескрипторов $P(A_k|D_i) = 0$, то $B_k = -1$; если связи между дескрипторами и активностью A_k нет и $P(A_k|D_i) \approx P(A_k)$, то $B_k \approx 0$.

До версии 1.703 алгоритм прогноза PASS использовал следующие данные о взаимосвязях «структура—активность»:

N — общее количество веществ в SAR base;

N_i — количество веществ, содержащих дескриптор D_i в описании структуры;

N_k — количество веществ, содержащих активность A_k в спектре активности;

N_{ik} — количество веществ, содержащих и дескриптор D_i в описании структуры, и активность A_k в спектре активности.

По этим данным вычислялись частотные оценки вероятностей $P(A_k)$ и $P(A_k|D_i)$:

$$P(A_k) = N_k / N, P(A_k|D_i) = N_{ik} / N_i$$

В PASS 1.703 и последующих версиях оценки вероятностей $P(A_k)$ и $P(A_k|D_i)$ вычисляются в виде следующих сумм по всем N веществам в SAR base:

$$\begin{aligned} P(A_k|D_i) &= \sum_n f_n(A_k) \cdot g_n(D_i) / \sum_n g_n(D_i); \\ P(A_k) &= \sum_n f_n(A_k) \cdot \sum_i g_n(D_i) / \sum_n \sum_i g_n(D_i) \end{aligned} \quad (4)$$

где $f_n(A_k)$ и $g_n(D_i)$ — характеристические функции принадлежности вещества с номером n к множеству веществ, содержащих активность A_k в спектре активности и дескриптор D_i в описании структуры, соответственно; $f_n(A_k)$ принимает значения 0 и 1, а $g_n(D_i) = 0$ и $1/m_n$, где m_n — число дескрипторов молекулы n и $\sum_i g_n(D_i) \equiv 1$.

Такая модификация алгоритма PASS не только позволила существенно повысить точность алгоритма прогноза, но и открыла очень интересные возможности. Например, функцию $f_n(A_k)$ можно рассматривать как меру принадлежности к нечеткому множеству веществ, проявляющих активность A_k . Точно также можно рассматривать и веса дескрипторов $g_n(D_i)$ — и тогда дескрипторы могут быть любой однородной природы. На этой основе мы разрабатываем метод

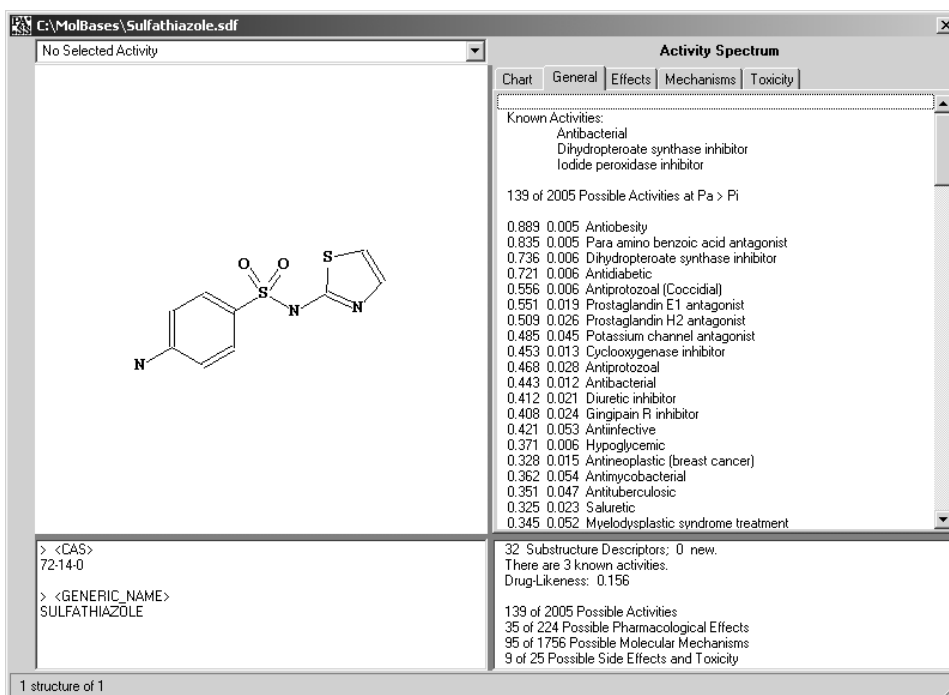


Рис. 2. Окно программы PASS со структурной формулой сульфатазиола и результатами прогноза спектра биологической активности

```

> <PASS_RESULT_COUNT>
35 of 224 Possible Pharmacological Effects at Pa > Pi
95 of 1756 Possible Molecular Mechanisms at Pa > Pi
9 of 25 Possible Side Effects and Toxicity at Pa > Pi

> <PASS_EFFECTS>
0.889 0.005 Antiobesity
0.721 0.006 Antidiabetic
0.556 0.006 Antiprotozoal (Coccidial)
0.468 0.028 Antiprotozoal
0.443 0.012 Antibacterial
0.412 0.021 Diuretic inhibitor
0.421 0.053 Antiinfective
0.371 0.006 Hypoglycemic
. . .
0.149 0.122 Liver fibrosis treatment
0.187 0.184 Antiparasitic

> <PASS.MECHANISMS>
0.835 0.005 Para amino benzoic acid antagonist
0.736 0.006 Dihydropteroate synthase inhibitor
0.551 0.019 Prostaglandin E1 antagonist
0.509 0.026 Prostaglandin H2 antagonist
0.485 0.045 Potassium channel antagonist
. . .
0.035 0.015 TNF convertase inhibitor
0.155 0.136 Iodide peroxidase inhibitor
0.020 0.005 Histone acetylation inducer
. . .
0.134 0.133 Astacin inhibitor

> <PASS.TOXICITY>
0.339 0.067 Carcinogenic, female rats
0.336 0.080 Carcinogenic, female mice
0.246 0.106 Carcinogenic
0.256 0.117 Carcinogenic, male mice
0.166 0.070 Mutagenic, Salmonella
0.213 0.132 Carcinogenic, group 3
0.156 0.083 Mutagenic
0.179 0.124 Carcinogenic, group 2B
0.180 0.135 Vasopressor
    
```

Рис. 3. Пример части спрогнозированного спектра активности сульфатазола.

Структуру см. на рис. 2. Известные (содержащиеся в обучающей выборке PASS) виды активности выделены в прогнозе жирным шрифтом

количественного прогноза биологической активности, и предварительные результаты показывают превосходство такого подхода на основе MNA-дескрипторов по сравнению с 3D COMFA и COMSIA (трехмерный сравнительный анализ молекулярных полей и сравнительный анализ молекулярного сходства).

Главное назначение системы PASS состоит в прогнозе спектров активности именно новых, еще не изученных, веществ. Поэтому общим принципом алгоритма прогноза PASS является исключение из SAR base вещества со структурой молекулы, эквивалентной структуре молекулы вещества, спектр биологической активности которого прогнозируется. Если в SAR base обнаружена эквивалентная структура с номером n , то это вещество исключается из суммирования в (4). Усреднение вкладов дескрипторов выполняется по MNA-дескрипторам прогнозируемого вещества, найденным в SAR base, т.е. тем, для которых $\sum_n g_n(D_i) > 0$.

Для получения качественного прогноза «да/нет» необходимо определить пороговые значения B -статистики, вычисляемые согласно (3) для каждого вида активности A_k . С помощью методов принятия

статистических решений это можно сделать на основе минимизации функционалов риска. Однако никто не может заранее задать такие функционалы для всех видов активности и для всех возможных практических задач. Поэтому в PASS результат прогноза спектра биологической активности представляется в виде упорядоченного списка названий соответствующих активностей и вероятностей P_a «быть активным» («to be active») и P_i «быть неактивным» («to be inactive»), которые являются функциями значений B -статистики для прогнозируемого соединения. Упорядочение выполняется по убыванию разности $P_a - P_i$, так что более вероятные виды активности находятся в начале спрогнозированного спектра. Спрогнозированный спектр активности может анализироваться любым желаемым образом, но по умолчанию в него включаются активности, для которых $P_a > P_i$.

В качестве примера на рис. 2 и 3 приведены результаты прогноза для сульфатазола. Это вещество найдено в SAR base и исключено из построения прогноза. Известный (в SAR base PASS версии 1.917) спектр активности сульфатазола включает в себя такие виды активности, как «Antibacterial (противобактериальное)», «Dihydropteroate synthase inhibitor (ингибитор дигидроптероатсинтазы)», «Iodide peroxidase inhibitor (ингибитор иодидпероксидазы)». На рис. 3 спрогнозированный спектр активности сульфатазола включает 139 из 2005 прогнозируемых по умолчанию видов активности при принятом по умолчанию условию $P_a > P_i$. Два из имеющихся в SAR base видов биологической активности сульфатазола («Antibacterial», «Dihydropteroate synthase inhibitor») прогнозируются с вероятностью выше 40%, один («Iodide peroxidase inhibitor») — лишь на уровне $P_a > P_i$.

Процедура обучения

Оценки точности прогноза PASS и зависимости, необходимые для получения вероятностей P_a и P_i по значениям B -статистики, являются конечным результатом процедуры обучения, которая состоит в следующем. По данным SAR base, сформированной на основе обучающей выборки, для каждой активности A_k для каждого из N_k активных и для каждого из $N - N_k$ неактивных веществ вычисляются значения B -статистики. Вычисления проводятся в режиме скользящего контроля с исключением по одному, т.е. после «исключения» этого соединения из SAR base, для чего достаточно не включать его в суммы в (4). По полученным выборкам B -статистики строятся гладкие оценки функций их распределения $P_a(B)$ и $P_i(B)$ [37].

Вероятности P_a и P_i являются также, по построению, оценками вероятности ошибок прогноза 1-го и 2-го рода, соответственно. Их можно рассматривать и как меры принадлежности прогнозируемого вещества к нечетким множествам «активных» и «неактивных» веществ. Все эти интерпретации вероятностей P_a и P_i эквивалентны и полезны для анализа результатов прогноза. На их основе можно сконструировать самые разные критерии анализа результатов прогноза, соответствующие решению конкретных практических задач.

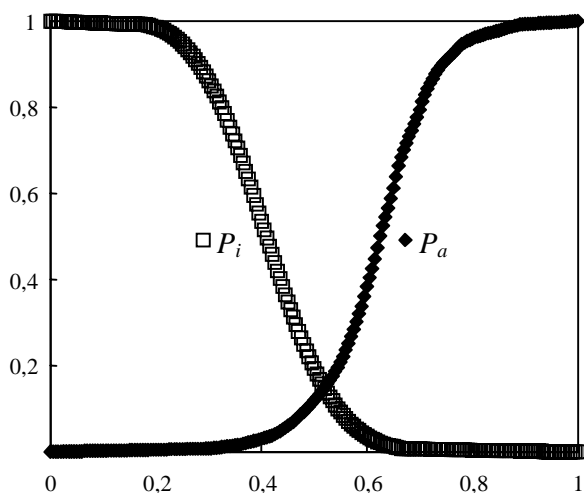


Рис. 4. Оценки вероятностей $P_a(B)$ и $P_i(B)$.

Кривые построены для активности «Alpha adrenoreceptor antagonist» в PASS версии 1.917

На рис. 4 приведен пример оценок вероятностей $P_a(B)$ и $P_i(B)$ как функций значений B -статистики для активности «Alpha adrenoreceptor antagonist (антагонист альфа-адренорецепторов)» в SAR base PASS версии 1.917.

Точка пересечения функций P_a и P_i соответствует равенству вероятностей ошибок 1-го и 2-го рода, и значение в этой точке $MEP = P_a = P_i$ является оценкой максимальной ошибки прогноза (MEP — maximal error of prediction) — характеристикой точности прогноза конкретной биологической активности. Общая точность прогноза оценивается в PASS как среднее по всем прогнозируемым видам активности значение MEP .

Важной особенностью алгоритма прогноза PASS является его устойчивость к неполноте информации о структурах и спектрах биологической активности химических соединений в обучающей выборке. В специальном исследовании [38] показано, что сокращение наполовину реально известной информации о структурах или активности химических соединений в обучающей выборке лишь незначительно уменьшает точность прогноза в перекрестном контроле. Там же продемонстрировано, что оценка точности по скользящему контролю с исключением по одному даже более жесткая, чем по перекрестному контролю.

Интерпретация результатов прогноза PASS

Необходимо помнить, что вероятность P_a отражает прежде всего сходство структуры молекул данного вещества со структурами молекул наиболее типичных в соответствующем подмножестве «активных» веществ в обучающей выборке. Поэтому никакой прямой корреляции вычисляемых величин P_a с количественными характеристиками активности, как правило, нет. Действительно активное вещество, но имеющее нетипичную для обучающей выборки структуру молекул, может иметь по прогнозу низкое значение P_a , даже возможно $P_a < P_i$. Это очевидно из способа построения функций $P_a(B)$ и $P_i(B)$: значения величин P_a для активных и P_i для неактивных веществ из обучающей выборки распределены строго равномерно. Из этого следует и интерпретация результатов прогноза.

Если, например, величина P_a равна 0,9, то для 90% активных веществ из обучающей выборки значение B -статистики меньше, чем для исследуемого вещества, и только для 10% — больше. Это также означает, что, если мы отклоним предположение о том, что вещество обладает активностью, то, в среднем, мы совершим ошибку с вероятностью 0,9 совершив ошибку.

Если же величина P_a меньше 0,5, но $P_a > P_i$, то, следовательно, более половины активных веществ из обучающей выборки имеют значение B -статистики больше, чем для данного вещества, и если мы отклоним предположение о том, что данное вещество обладает активностью, то, в среднем, совершим ошибку с вероятностью менее 0,5. В этом случае вероятность обнаружить данный вид активности экспериментально невелика, но, если таковая будет найдена, то более 50% шансов за то, что эта структура — оригинальная.

Обширный спрогнозированный спектр активности свидетельствует о том, что структура молекул данного вещества довольно проста, не содержит никаких особенностей, обеспечивающих высокую селективность его биологического действия.

Если при прогнозе оказалось, что в структуре есть несколько новых MNA-дескрипторов, то структура мало похожа на любую из структур в SAR base, и результаты прогноза необходимо рассматривать как очень приблизительные.

Применение прогноза спектров биологической активности органических соединений

Использование PASS позволяет уже на ранних стадиях исследований отобрать из возможных веществ-кандидатов те, которые могут обладать желательными видами биологической активности и с малой вероятностью способны вызывать нежелательные побочные эффекты.

Биологическая активность пептидов в значительной мере обусловлена пространственными особенностями их структуры. Используемые в PASS MNA-дескрипторы 1-го и 2-го уровня не учитывают никаких пространственных особенностей и охватывают лишь небольшие локальные фрагменты молекул, в которых самые дальние атомы разделены не более чем четырьмя связями. Однако и для три- и тетрапептидов и пептидомиметиков PASS обеспечивает высокую точность прогноза [44]. Поскольку прогноз выполняется по структурной формуле вещества, он может быть получен уже на стадии планирования синтеза.

Нередко лекарственный препарат поступает в организм в форме пролекарства. Выполненный нами прогноз спектра биологической активности для ряда пролекарств и их активных метаболитов показал, что в 74% случаев основной эффект препарата предсказывается PASS по структурной формуле его предшественника [45]. Для многих применяемых в медицинской практике препаратов PASS прогнозирует новые вероятные эффекты, которые целесообразно проверить в эксперименте и клинике [46–48]. Так, полученные нами с помощью PASS результаты прогноза спектра биологической активности для 200 наиболее часто используемых в США лекарственных препаратов не просто совпадают с 93% известных фармакологических эффектов и механизмов действия, но и указывают на новые возможные применения некоторых известных препаратов, в частности, ингибирование ан-

гиогенеза (образования кровеносных сосудов) миорелаксантом каризопродол, лечение когнитивных расстройств антигипертензивным препаратом рамиприл, лечение рассеяного склероза антигипертензивным препаратом амлодипин и др. [47, 48].

С помощью PASS нами был выполнен поиск веществ с комбинированным антигипертензивным действием. Для экспериментального решения такой задачи необходимо исследование сотен тысяч веществ с очень низкой вероятностью успеха — задача может быть решена только компьютерными методами [49, 50].

Прогнозируемые с помощью PASS активности были подтверждены в эксперименте для веществ различных химических классов, проявляющих разнообразные эффекты: противоязвенный [51], противоопухолевый [52, 53], противобактериальный [54–56], гепатопротекторный [57, 58], антиоксидантный [59], местноанестезирующий [60], анксиолитический [61, 62], усиления когнитивных функций [63].

Наиболее ярким примером эффективности применения компьютерного прогноза является поиск новых антиконвульсантов, анксиолитиков и ноотропных препаратов. Химики из шести университетов пяти стран (Россия, Молдова, Бельгия, Греция, Португалия) была создана база данных, содержащая 5494 виртуальные структурные формулы веществ, которые они могли синтезировать. На основе прогноза было отобрано 30 соединений, которые были синтезированы и протестированы экспериментально на соответствующие виды активности в лаборатории профессора Т.А. Ворониной (Институт фармакологии РАМН, Москва). Оказалось, что практически все из исследованных экспериментально веществ проявляли прогнозируемую активность, а некоторые из них превосходили в эксперименте препараты сравнения [61, 63]. Эти вещества имеют высокую степень новизны структуры молекул по отношению к известным анксиолитикам и ноотропам. Если бы не использовался компьютерный прогноз, а исследование осуществлялось путем скрининга, то затраты на синтез и проведение биологических испытаний были бы больше в 182 раза!

Поскольку прогнозируемые спектры биологической активности содержат оценки вероятностей основных и побочных фармакологических эффектов, молекулярных механизмов действия, специфической токсичности, то выбор наиболее перспективных веществ из числа доступных образцов химических соединений можно осуществлять на основе сложных критериев. При этом во внимание принимается как наличие требуемых видов биологической активности и механизмов действия, так и отсутствие нежелательных побочных эффектов и токсичности. И задачи поиска базовой структуры, обладающей требуемыми свойствами, и оптимизации ее с точки зрения снижения побочных и токсических эффектов, обычно решаемые последовательно, будут решаться одновременно. Более того, показано, что алгоритм PASS может быть успешно применен для дискриминации «лекарственно-подобных» (drug-like) и не «лекарственно-подобных» (drug-unlike) веществ [64], что позволяет расширить область использования системы и «отфильтровывать» уже на ранних стадиях исследования химические соединения, вероятность которых стать лекарством сравнительно невелика.

Для расширения возможностей применения PASS нами обеспечена возможность прогноза спектра биологической активности веществ через Интернет [32, 65, 66]. Используя стандартные браузеры Netscape или Internet Explorer, пользователь может послать на сайт <http://www.ibmc.msk.ru/PASS/> структурную формулу вещества, представленную в виде mol-файла, и автоматически получить на дисплее своего компьютера результаты прогноза наиболее вероятных для данного химического соединения видов биологической активности. Более двух тысяч ученых из России, Украины, Латвии, США, Германии, Великобритании, Франции, Индии и ряда других стран получили через Интернет спектры биологической активности для нескольких тысяч веществ, спрогнозированные с использованием Интернет-версии PASS.

Совместно с коллегами из Национального института рака США (NCI NIH) выполнен поиск противовирусных (ВИЧ) препаратов в базе данных NCI. Показано, что использование PASS позволяет от 2 до 17 раз повысить вероятность успеха в первичных испытаниях [67]. Возможности компьютерного прогнозирования биологической активности заинтересовали докт. Марка Никлауса из Национального института здоровья США (Лаборатория медицинской химии, Национальный раковый институт, NCI), и он обратился к нам с предложением о сотрудничестве. Мы провели прогнозирование спектра биологической активности для 250000 органических соединений, зарегистрированных NCI. Оказалось, что наряду с различными механизмами противоопухолевого действия веществ, для многих соединений предсказываются не известные ранее эффекты.

Результаты прогноза представлены на сервере NCI в Интернете <http://cactus.nci.nih.gov/ncidb2/> и используются в настоящее время американскими коллегами для целенаправленного отбора соединений с требуемыми видами биологической активности.

На основе PASS разрабатывается система прогноза биотрансформации молекул в организме человека, что очень важно для интегральной и всесторонней оценки воздействия лекарств и ксенобиотиков на человеческий организм. Возможные виды биотрансформации из списка более 1000 биотрансформаций прогнозируются с точностью лучше 85% [68], продолжается разработка прогноза конкретных реакций и структуры метаболитов [69].

Работа над компьютерной системой PASS осуществляется при поддержке ряда российских и международных грантов, включая РФФИ 03-07-90282, РФФИ 05-07-90123, CRDF RC1-2064, INTAS 00-711, INTAS 03-55-5218, ISTC 574, ISTC 3197, Research Council of Norway 165190/V40.

ЛИТЕРАТУРА

1. Negwer M., Scharnow H-G. Organic Chemical-Drugs and Their Synonyms, 8th edition. Weinheim: Wiley-VCH, 2001.
2. Баренбойм Г.М., Маленков А.Г. Биологически активные вещества. Новые принципы поиска. М.: Наука, 1986, 363 с.
3. Арчаков А.И., Поройков В.В. Биоинформатика. Новые биомедицинские технологии. Обзор. НИИ биомедицинской химии РАМН, 1999, 47 с.
4. Hansch C., Leo A., Hoekman D. Exploring QSAR. V. II. N. Y., ACS Professional Reference Book, Oxford University Press USA, 1995.
5. Livingstone D. Data Analysis for Chemists. Applications to QSAR and Chemical Product Design. Oxford University Press, 1995.

6. 3D QSAR in Drug Design. Eds. H. Kubinyi, G. Folkers, Y.C. Martin. V. I—III, Kluwer, Dordrecht, The Netherlands, 1993, 1998, 2002.
7. *Lipnick R.L.* SAR & QSAR Environ. Res., 1999, v. 10 (2-3), p. 239—248.
8. *Hawkins D.M., Basak S.C., Shi X.* J. Chem. Inf. Comp. Sci., 2001, v. 41, № 3, p. 663—670.
9. Handbooks of Cheminformatics: From Data to Knowledge. Ed. J. Gasteiger. 4 V. Weinheim: Wiley-VCH, 2003.
10. *Dean P.M.* Molecular Similarity in Drug Design. London: Blackie Academic, 1995.
11. *Willet P.* Similarity and Clustering in Chemical Information Systems. Research Studies Press Ltd: Letchworth, U.K., 1987.
12. *Barnard J.M., Downs G.M.* J. Chem. Inf. Comp. Sci., 1992, v. 32, p. 644—649.
13. *Wild D. J., Blankey C.J.* J. Ibid., 2000, v. 40(1), p. 155—162.
14. *Brown R.D., Martin Y.C. J.* Ibid., 1996, v. 36, p. 572—584.
15. *Авидон В.В., Лексина Л.А.* Научн.-техн. информ. Сер. 2, 1974, № 3, с. 22—25.
16. *Авидон В.В.* Хим.-фарм. ж., 1974, № 8, с. 22—25.
17. *Голендер В.Е., Розенблит А.Б.* Автоматика и телемеханика, 1974, № 11, с. 99—105.
18. *Авидон В.В., Аролович В.С., Козлова С.П.* Научн.-техн. информ. Сер. 2, 1974, № 12, с. 21—23.
19. *Авидон В.В., Аролович В.С.* Там же, 1975, № 5, с. 26—31.
20. *Прирузян Л.А., Авидон В.В., Розенблит А.Б. и др.* Хим.-фарм. ж., 1977, т. 11, № 5, с. 35—40.
21. *Авидон В.В., Аролович В.С., Козлова С.П., Прирузян Л.А.* Там же, 1978, № 5, с. 88—92.
22. *Авидон В.В., Аролович В.С., Козлова С.П., Прирузян Л.А.* Там же, 1978, № 6, с. 99—106.
23. *Розенблит А.Б., Голендер В.Е.* Логико-комбинаторные методы в конструировании лекарств. Рига: Зинатне, 1984.
24. *Кочетков А.Н., Васильев П.М., Бреслаухов А.Г.* Тез. докл. перв. Всес. конф. по теор. орг. хим. Волгоград, 1991, ч. 2, с. 500.
25. *Буров Ю.В., Корольченко Л.В., Поройков В.В.* Бюлл. ВНИЦ БАВ, 1990, № 1, с. 4—25.
26. *Филимонов Д.А., Поройков В.В. и др.* Экспериментальная и клиническая фармакология, 1995, т. 58, № 2, с. 56—62.
27. *Poroikov V.V., Filimonov D.A.* In: QSAR and Molecular Modeling Concepts, Computational Tools and Biological Applications. Barcelona: Prous Science Publishers, 1996, p. 49—50.
28. *Filimonov D.A., Poroikov V.V.* In: Bioactive Compound Design: Possibilities For Industrial Use; BIOS Scientific Publishers: Oxford, 1996, p. 47—56.
29. *Поройков В.В., Филимонов Д.А. и др.* Хим.-фарм. ж., 1996, т. 30, № 9, с. 20—23.
30. *Глоризова Т.А., Филимонов Д.А., Лагунин А.А., Поройков В.В.* Там же, 1998, т. 32, № 12, с. 32—39.
31. *Filimonov D.A., Poroikov V.V., Borodina Yu.V., Glorizova T.A.* J. Chem. Inf. Comp. Sci., 1999, v. 39, № 4, p. 666—670.
32. *Lagunin A., Stepanchikova A., Filimonov D., Poroikov V.* Bioinformatics, 2000, v. 16, № 8, p. 747—748.
33. *Poroikov V.V., Filimonov D.A.* In: Nitrogen-containing heterocycles and alcaloides. Eds. V.G. Kartsev, G.A. Tolstikov. Moscow: Iridium Press, 2001, v. 1, p. 149—154.
34. *Poroikov V., Filimonov D.* Rational Approaches to Drug Design. Eds. H.-D. Holtje, W. Sippl. Barcelona: Prous Science, 2001, p. 403—407.
35. *Stepanchikova A.V., Lagunin A.A., Filimonov D.A., Poroikov V.V.* Curr. Med. Chem., 2003, v. 10, № 3, p. 225—233.
36. *Dembitsky V.M., Glorizova T.A., Poroikov V.V.* Mini-Reviews in Medicinal Chemistry, 2005, v. 5, p. 319—336.
37. *Poroikov V., Filimonov D.* In: Predictive Toxicology. Ed. by Christoph Helma. Taylor & Francis, 2005, p. 459—478.
38. *Poroikov V.V., Filimonov D.A., Borodina Yu.V., Lagunin A.A., Kos A. J.* Chem. Inf. Comp. Sci., 2000, v. 40(6), p. 1349—1355.
39. *Минкин В.И., Симкин Б.Я., Миняев Р.М.* Теория строения молекул (электронные оболочки). Учебное пособие для университетов. М.: Высшая школа, 1979, 407 с.
40. *Марч Н., Кон В., Вашишта П., Лундквист С., Уильямс А., Барт У., Лэнг Н.* Теория неоднородного электронного газа. Под ред. С. Лундквиста и Н. Марча. Пер. с англ. М.: Мир, 1987, 400 с.
41. *Avidon V.V., Pomerantsev I.A., Rozenblit A.B., Golender V.E. J.* Chem. Inf. Comp. Sci., 1982, v. 22, № 4, p. 207—211.
42. *Татевский В.М.* Теория физико-химических свойств молекул и веществ. М.: Изд-во Моск. ун-та, 1987, 239 с.
43. *Филимонов Д.А.* Тез. докл. II Росс. нац. конгресса «Человек и лекарство». М., 1995, с. 62—63.
44. *Мартьянова Н.Б., Филимонов Д.А., Поройков В.В.* Биоорганическая химия, 2000, т. 26, № 5, с. 330—339.
45. *Бородина Ю.В., Поройков В.В., Филимонов Д.А.* Хим.-фарм. ж., 1996, т. 30, № 12, с. 39—42.
46. *Поройков В.В.* Химия в России, 1999, № 2, с. 8—12.
47. *Poroikov V., Filimonov D., Akitov D.* Newsletter of The QSAR and Modelling Society, 2000, № 11, p. 19—22.
48. *Poroikov V., Akitov D., Shabelnikova E., Filimonov D.* SAR and QSAR in Environmental Research, 2001, v. 12, № 4, p. 327—344.
49. *Лагунин А.А., Филимонов Д.А., Поройков В.В.* Хим.-фарм. ж., 2001, т. 35, № 7, с. 28—34.
50. *Lagunin A.A., Gatazkov O.A., Filimonov D.A., e.a.* J. Med. Chem., 2003, v. 46, № 15, p. 3326—3332.
51. *Трапков В.А., Будунова А.П., Бурова О.А., Филимонов Д.А., Поройков В.В.* Вопросы мед. химии, 1997, т. 43, № 1, с. 41—57.
52. *Погребняк А.В., Поройков В.В., Старых В.В., Коновалов Д.А.* Растит. ресурсы, 1998, т. 34, № 1, с. 61—64.
53. *Исляйкин М.К., Данилова Е.А., Кудрик Е.В. и др.* Хим.-фарм. ж., 1997, т. 31, № 8, с. 19—22.
54. *Майборода Д.А., Бабаев Е.В., Гончаренко Л.В.* Там же, 1998, т. 32, № 6, с. 24—28.
55. *Di Giorgio C. e. a.* Antimicrob. Agents Chemoter., 2003, v. 47, № 1, p. 174—180.
56. *Di Giorgio C. e. a.* Experimental Parasitology, 2004, v. 106, p. 67.
57. *Сальникова С.И., Поройков В.В., Мазур И.А.* Тез. докл. респ. научн.-практ. конф. Одесса, 1993, с. 101—102.
58. *Флехтер О.Б., Карачурина Л.Т., Поройков В.В. и др.* Биоорганическая химия, 2000, т. 26, 3, с. 215—223.
59. *Geronikaki A., Poroikov V., Hadjipavlou-Litina D. e.a.* Quant. Struct.-Act. Relat., 1999, v. 18, p. 16—25.
60. *Geronikaki A., Vicini P., Theophilidis G. e.a.* SAR and QSAR in Environmental Research, 2003, v. 14 (5—6), p. 485—495.
61. *Geronikaki A., Babaev E., Dearden J. e.a.* Bioorg. Med. Chem., 2004, v. 12, p. 6559—6568.
62. *Goel R. K., Kumarb V., Mahajan M.P.* Bioorg. Med. Chem. Lett., 2005, v. 15, p. 2145—2148.
63. *Geronikaki A., Dearden J., Filimonov D. e.a.* J. Med. Chem., 2004, v. 47, № 11, p. 2870—2876.
64. *Anzali S., Barnickel G., Cezanne B., Krug M., Filimonov D., Poroikov V.* Ibid., 2001, v. 44, № 15, p. 2432—2437.
65. *Садым А.В., Лагунин А.А., Филимонов Д.А., Поройков В.В.* Хим.-фарм. ж., 2002, т. 36, с. 21—26.
66. *Sadyam A., Lagunin A., Filimonov D., Poroikov V.* SAR and QSAR in Environmental Research, 2003, v. 14 (5—6), p. 339—347.
67. *Poroikov V.V., Filimonov D.A., Ihlenfeldt W-D. e.a.* J. Chem. Inf. Comp. Sci., 2003, v. 43(1), p. 228—236.
68. *Borodina Yu., Sadyam A., Filimonov D., Blinova V., Dmitriev A., Poroikov V.* Ibid., 2003, v. 43, № 5, p. 1636—1646.
69. *Borodina Yu., Rudik A., Filimonov D. e.a.* Ibid., 2004, v. 44, p. 1998.