

# ФАРМАКОЛОГИЯ, ФАРМАЦИЯ, МЕДИЦИНСКАЯ ХИМИЯ

УДК 615:681.3:541.4

## ПРИМЕНЕНИЕ КОМПЬЮТЕРНОЙ ИНФОРМАЦИОННОЙ ТЕХНОЛОГИИ ДЛЯ ПРОГНОЗА ФАРМАКОЛОГИЧЕСКОЙ АКТИВНОСТИ СТРУКТУРНО РАЗНОРОДНЫХ ХИМИЧЕСКИХ СОЕДИНЕНИЙ

П.М. Васильев, А.А. Спасов  
Кафедра фармакологии ВолГМУ

Одной из актуальных задач современной фармакологии является поиск новых соединений с высокой биологической активностью. Экспериментальное изучение всех синтезируемых веществ на весь спектр видов активности чрезвычайно длительно и трудоемко. Академией наук России одним из приоритетных направлений исследования свойств органических соединений объявлено использование для этих целей компьютерных информационных технологий, основанных на исследовании зависимостей "структура – свойство" [9]. Для решения таких задач разработаны специализированные системы, позволяющие выполнять прогноз активности новых еще не синтезированных и не испытанных веществ, используя в качестве исходной информации только структурную формулу соединения – например, система PASS [7] и рассматриваемая в настоящем исследовании информационная технология "Микрокосм" [1].

### ЦЕЛЬ РАБОТЫ

Проверка точности прогноза с использованием информационной технологии "Микрокосм" различных видов фармакологической активности существенно различающихся по своему химическому строению соединений и сопоставление полученных результатов с данными о точности прогноза этих же видов активности в системе PASS.

### МЕТОДИКА ИССЛЕДОВАНИЯ

Основой методики прогноза являются полученные методами теории распознавания образов закономерности, связывающие между собой факт наличия/отсутствия активности и структуру соединений, представленную в виде матрицы структурных дескрипторов.

Исследование выполнялось в несколько этапов. **Формирование обучающих выборок.** В качестве обучающих выборок использовались созданные с помощью стандартных компьютерных химических программ 20 наборов структур соединений по различным видам фармакологической (биологической) активности. Каждый набор содержит структуры как активных (достоверно проявляющих данный вид активности), так и неактивных (достоверно не проявляющих данный вид активности) соединений.

Данные по 13 видам активности взяты из справочника [12] и дополнены данными из более поздних литературных источников (совместно с Г.П. Дудченко); в качестве неактивных добавлены случайно выбранные достоверно неактивные структуры из [12].

Данные по 6 видам активности получены в результате анализа журнальных публикаций (совместно с М.С. Новиковым и В.А. Сажиним).

Данные по канцерогенной активности взяты с сайта IARC <http://www.iarc.fr>, формирование обучающей выборки описано в [5].

Всего в базе данных 8863 структуры. Перечень 20 исследовавшихся активностей и объемы соответствующих выборок приведены в табл. 1.

**Формирование моделей обобщенных образов.** Структурные формулы соединений преобразовывались программой-транслятором в дескрипторы специализированного рабочего языка QL2, из которых формировались модели обобщенных образов классов активных/неактивных соединений, представленных в виде матрицы структурных дескрипторов.

Язык QL (QSAR Language, версия 2) представляет собой мультидескрипторный иерархический многоуровневый язык с подструктурной

нотацией, концепции и развитие которого изложены в работах [2, 11].

Таблица 1

## Характеристики обучающих выборок

Вид активности	Число соединений в выборке		
	Всего	Активных	Неактивных
Нейролептическая <sup>1</sup>	645	226	419
Транквилизирующая <sup>1</sup>	532	176	356
Антидепрессантная <sup>1</sup>	628	238	390
Ноотропная <sup>2</sup>	420	217	203
Анальгетическая наркотическая <sup>1</sup>	320	114	206
Местноанестезирующая <sup>1</sup>	324	118	206
Кардиотоническая <sup>1</sup>	304	98	206
Кардиостимулирующая <sup>1</sup>	233	27	206
Гипогликемическая <sup>1</sup>	230	115	115
Анаболическая <sup>1</sup>	175	91	84
Антисептическая <sup>1</sup>	494	234	260
Туберкулостатическая <sup>1</sup>	386	126	260
Фунгицидная <sup>1</sup>	471	211	260
Анти-ВИЧ <sup>2</sup>	1140	543	597
Антигерпесвирусная <sup>2</sup>	412	192	220
Антипикорнавирусная <sup>2</sup>	512	251	261
Противогриппозная <sup>2</sup>	72	18	54
Противолеиозная <sup>2</sup>	252	175	77
Противоопухолевая <sup>1</sup>	821	562	259
Канцерогенная <sup>3</sup>	492	245	247
Всего	8863	3976	4886

<sup>1</sup> Данные из [12] дополнены литературными данными.

<sup>2</sup> По результатам анализа журнальных публикаций.

<sup>3</sup> Данные из [5].

Алфавит QL задается тремя типами элементарных дескрипторов.

1. Структурный дескриптор (СД) – фрагмент структуры соединения, обладающий достаточно выраженными электродонорными или электроакцепторными свойствами и представляющий собой неводородный атом или группу атомов с учетом ближайшего окружения. Подалфавит СД QL постулирован и состоит из 4352 видов дескрипторов, в том числе 372 видов гетероатомных СД (например,  $-\text{NH}_2$ ,  $=\text{O}$ ,  $-\text{S}^{+<}$ ,  $\text{Cr}^{+n}$ ), 11 углеродных СД (например,  $-\text{C}^{+<}$ ,  $-\text{CH}_3$ ,  $-\text{CH}=\text{}$ ,  $-\text{C}\#$ ), 3 типов циклических СД (простые, сопряженные с  $k$   $\pi$ -электронами или ароматические циклы размером  $n$ ; всего 3963 вида циклических СД) и 6 видов СД инертных газов.

2. Дескриптор длины (ДД) – длина в связях кратчайшего пути по углеродной цепочке между двумя СД (принимает значения от 1 до 99) или число общих атомов при непосредственном на-

ложении СД друг на друга (индекс вхождения, принимает отрицательные значения).

3. Дескриптор связи (ДС) – типы связей, их количество и наличие сопряжения на ДД. Подалфавит ДС постулирован и состоит из 54 видов дескрипторов, каждый из которых формируется сочетанием 4 бинарных индексов: наличия одной или нескольких кратных ( $p$ ,  $P$ ), ароматических ( $a$ ,  $A$ ) и нековалентных ( $n$ ,  $N$ ) связей и индекса сопряжения ( $0$ ,  $1$ ).

Комбинации из двух, трех или четырех элементарных дескрипторов называются простыми составными дескрипторами соответствующего ранга. Дескрипторов второго ранга насчитывается четыре типа, дескрипторов третьего ранга – три типа, дескрипторов четвертого ранга (базовый дескриптор) – один тип.

Структура соединения описывается совокупностью входящих в нее QL-дескрипторов до четвертого ранга включительно – всего 11 типами дескрипторов разного ранга. Такое описание полностью отражает молекулярный граф соединения, поэтому дескрипторы более высоких рангов не используются. Структура органического соединения средней сложности описывается обычно 50–400 видами QL-дескрипторов.

Матрица структурных дескрипторов представляет собой таблицу, в которой по строкам расположены символьные обозначения уникальных QL-дескрипторов 11 типов до четвертого ранга включительно, а по столбцам – номера соединений обучающих выборок. Каждая ячейка такой таблицы содержит число QL-дескрипторов  $j$ -го вида в структуре соединения  $S$ . Дескрипторы в матрице расположены в порядке увеличения ранга и в лексикографическом порядке. Обозначения дескрипторов химически содержательны. Например, дескриптор ( $-\text{OH}$  5  $=\text{O}$  pA.1) представляет собой дескриптор 4-го ранга 11 типа  $\text{СД}_1\text{--ДД--СД}_2\text{--ДС}$ , в котором гидроксил связан с карбонильной группой углеродной цепочкой длиной 5, причем в цепочке есть одна  $\pi$ -связь, две или более ароматических связи, отсутствуют нековалентные связи и присутствует сопряжение. Матрицы для прогноза перечисленных в табл. 1 видов активности содержат от 2406 (противогриппозная) до 22461 (противоопухолевая) видов QL-дескрипторов.

**Формирование решающих правил.** Основной информационной технологии "Микрокосм" является процедура обобщения с помощью трех различных схем принятия решений (стратегий прогноза) спектра промежуточных прогнозных оценок активности, полученных четыремя различными методами прогноза по 11 уровням QL-описания структуры соединений [4].

• **Методы прогноза.** Для получения спектра промежуточных прогнозных оценок используются четыре основанных на теории распознавания образов оригинальных метода, адаптированных

(13)

для использования в пространствах большой размерности. Каждый метод дает 11 прогнозных оценок (по числу типов QL-дескрипторов). Все четыре метода существенно отличаются друг от друга по способу построения решающего правила. Это дает основание полагать, что спектр из 44 прогнозных оценок достаточно адекватно отражает различные особенности обобщенного образа класса активных/неактивных соединений, что позволяет в дальнейшем достоверно оценивать итоговую активность прогнозируемого соединения.

**Метод Байеса (вероятностный параметрический).** Ранее разработанный алгоритм [5] при выполнении настоящего исследования был модифицирован. Логарифм вероятности принадлежности соединения  $C$  к классу активных  $a$  при наличии в нем  $d_i$  дескрипторов  $i$ -го типа  $j$ -го вида  $B_{ij}$

$$\log[P_i(C \in a)] = \log(0,5) + \sum_{j=1}^{d_i} \log[P(B_{ij}|C \in a)],$$

где  $P(B_{ij}|C \in a)$  – априорная вероятность встречаемости дескриптора  $B_{ij}$  в классе активных  $a$  соединений.

Соединение  $C$  считается активным по дескриптору  $i$ -го типа, если

$$\log[P_i(C \in a)] \geq \log[P_i(C \in n)].$$

**Метод расстояния (геометрический параметрический).** Ранее разработанный алгоритм [4] при выполнении настоящего исследования был модифицирован. Вычисляется взвешенное расстояние Пирсона в пространстве дескрипторов  $i$ -го типа от соединения  $C$  до центра класса  $a$

$$D_{ai} = \sum_{j=1}^{d_i} w_{ij} |x_{ij} - c_{aij}|,$$

где  $x_{ij}$  – координаты в пространстве признаков соединения  $C$  по дескриптору  $ij$ ;

$c_{aij}$  – координаты центра класса  $a$  по дескриптору  $ij$ ;

$w_{ij} = (c_{aij} + c_{nij})^{-1}$  – весовой коэффициент для дескриптора  $ij$ .

Соединение  $C$  по дескриптору  $i$ -го типа принадлежит к тому классу, расстояние до центра которого меньше.

**Метод ближайшего соседа (геометрический непараметрический).** Базовый алгоритм описан в [8] и при проведении настоящего исследования адаптирован для пространств большой размерности. В пространстве дескрипторов  $i$ -го типа вычисляются расстояния Евклида от прогнозируемого соединения  $C$  до каждого соединения обучающей выборки

$$D_{il}^2 = \sum_{\substack{j=1 \\ j \in X \cup Y}}^{d_i} (x_{ij} - y_{ijl})^2, \quad l=1, \dots, N,$$

где  $x_{ij}$  – координаты в пространстве признаков соединения  $C$  по дескриптору  $ij$ ;

$y_{ijl}$  – координаты в пространстве признаков соединения  $l$  по дескриптору  $ij$ ;

$N$  – число соединений обучающей выборки.

По дескрипторам  $i$ -го типа соединение  $C$  принадлежит к тому классу, активность которого проявляет ближайшее к соединению  $C$  соединение обучающей выборки (ближайший сосед):

$$D_i^2 = \min_{l=1}^N (D_{il}^2).$$

**Метод локального распределения (смешанный непараметрический геометрический – параметрический вероятностный)** [6]. Вычисляется минимальное число объектов обучающей выборки, необходимых для построения локальной статистики (5 % от общего числа соединений, но не менее 3 и не более 15):

$$L = \begin{cases} 3, & \text{если } \text{trunc}(N/20) < 2 \\ 15, & \text{если } \text{trunc}(N/20) > 14 \\ \text{trunc}(N/20) + 1, & \text{иначе} \end{cases}$$

Для прогнозируемого соединения  $C$  по каждому  $i$ -му типу QL-дескрипторов вычисляются  $L$  ближайших соседей обучающей выборки с использованием супремальной метрики:

$$D_{il} = \max_{\substack{j=1 \\ j \in X \cup Y}}^{d_i} |x_{ij} - y_{ijl}|, \quad l=1, \dots, N.$$

Затем по  $L$  отобранным соединениям методом Байеса производится классификация прогнозируемого соединения.

• **Стратегии прогноза.** Для обобщения спектра прогнозных оценок в информационной технологии "Микрокосм" реализованы три стратегии принятия решения об итоговой активности соединения [3].

**Консервативная стратегия.** Используется процедура простого голосования. Соединение  $C$  считается активным/неактивным, если из 44 промежуточных прогнозных оценок 27 или более одинаковые (95 % доверительный интервал для медианы).

Такой подход учитывает наиболее устойчивые, характерные именно для данного вида свойств, закономерности.

**Нормальная стратегия.** Предусматривает выбор из 4 методов прогноза, по результатам предварительного автотестирования обучающей выборки с обобщением 11 вычисленных в рамках каждого метода прогнозных оценок метода, лучшего по точности. В качестве процедуры голосования используется Байесовский бинарный классификатор [10]

$$L(x) = \sum_{i=1}^{11} w_i x_i + w_0.$$

Весовые коэффициенты

$$w_i = \log \frac{p_{ai}(1-p_{ni})}{p_{ni}(1-p_{ai})}; \quad w_0 = \sum_{i=1}^{11} \log \frac{1-p_{ai}}{1-p_{ni}},$$

где  $p_{ki} = \frac{n_{ki} + 1}{n_{ai} + n_{ni} + 2}$  – априорная вероятность

классификации соединения  $C$  по дескриптору  $i$ -го типа в класс активных  $a$  или неактивных  $n$  соединений;

$n_{ki}$  – число правильно классифицированных по  $i$ -му типу дескриптора активных (неактивных) соединений обучающей выборки.

В рамках каждого метода соединение  $C$  считается активным по всему QL-описанию в совокупности, если  $L(x) \geq 0$ .

При использовании нормальной стратегии прогноза учитываются как стандартные, так и нестандартные, для данного вида свойств, закономерности.

**Рисковая стратегия.** Основана на выборе из 4 методов прогноза по результатам предварительного автотестирования обучающей выборки отдельно по каждому из 11 уровней QL-описания лучшего по точности метода с учетом типа дескриптора. Процедуры итогового голосования нет, каждое из 44 прогнозных множеств рассматривается как самостоятельное информационное пространство.

Рисковая стратегия позволяет косвенно учитывать при прогнозе особенности, связанные с возможным механизмом эффектов, максимально учитывает новизну прогнозируемой структуры.

**Тестирование решающих правил.** Оценка точности прогноза наличия/отсутствия активности проводилась отдельно по каждой из 20 обучающих выборок для каждой из трех стратегий по четырем показателям распознающей и прогностической способности итогового решающего правила: результатам автопрогноза, скользящего, перекрестного и двойного скользящего контроля.

При прогнозе образ тестируемого соединения (в виде QL-представления) сравнивается с моделями обобщенных образов из матрицы и рассчитываются метрики принадлежности прогнозируемой структуры к классу активных/неактивных соединений.

• **Автопрогноз.** Рассчитывается активность каждого из  $N$  соединений обучающей выборки без каких-либо изменений рабочей QL-матрицы.

• **Скользящий контроль.** Каждое из соединений обучающей выборки поочередно исключается из расчетов рабочей QL-матрицы, а затем

используется в качестве независимого тестового объекта – так  $N$  раз.

• **Перекрестный контроль.** По нечетным соединениям обучающей выборки рассчитывается рабочая QL-матрица, четные соединения используются в качестве независимой тестовой выборки – затем наоборот; результаты тестирования усредняются.

• **Двойной скользящий контроль.** Из обучающей выборки исключается одно соединение. На выборке из оставшихся  $N-1$  соединений выполняется процедура скользящего контроля. По результатам этого скользящего контроля рассчитываются параметры итогового решающего правила стратегии. С использованием этого решающего правила производится классификация удаленного из расчетов соединения – процедура повторяется  $N$  раз.

**Программное обеспечение.** Все расчеты выполнены с использованием программного комплекса "Микрокосм" (версия 3.1, июнь 2004 г.), который является практической реализацией теоретических концепций, методов и правил информационной технологии "Микрокосм". Комплекс разработан на языке Delphi-6 в виде приложений для Windows-XP. База данных по структуре и активности соединений формировалась с помощью стандартной программы Chem-Finder-5.

## РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ И ИХ ОБСУЖДЕНИЕ

**Прогноз в информационной технологии "Микрокосм".** В табл. 2–4 приведены результаты тестирования точности прогноза факта наличия/отсутствия у соединений 20 перечисленных в табл. 1 видов фармакологической активности.

В рамках бинарной классификации работоспособные прогнозные модели должны удовлетворять очевидному условию: общая точность прогноза  $F_0$ , точности прогноза активных  $F_a$  и неактивных  $F_n$  соединений должны быть не ниже 70 %.

Наиболее адекватными способами оценки прогностической способности решающих правил являются процедуры, основанные на последовательном изъятии одного или нескольких соединений из состава обучающей выборки с последующим использованием их в качестве тестовых [8]. Это прежде всего скользящий и перекрестный контроль; результаты автопрогноза и двойного скользящего контроля носят справочный характер.

Проведенный с учетом перечисленных ограничений анализ данных табл. 2 показал, что при использовании консервативной стратегии адекватный прогноз возможен для 16 из 20 рассматриваемых видов активности. Для 4 видов активности (фунгицидная, антигерпесвирусная, противогриппозная и противоопухолевая – выделены курсивом) результаты прогноза будут иметь невысокую достоверность. Минимальное значе-

(13)

ние точности прогноза при использовании консервативной стратегии составляет 58 % (Fa в перекрестном контроле для антигерпесвирусной активности), максимальное – 100 %; среднее значение изменяется от 82 % до 99 %.

Анализ данных табл. 3 показал, что при использовании нормальной стратегии адекватный прогноз возможен только для 15 видов активности. Для 5 видов активности (ноотропная, фунгицидная, антигерпесвирусная, противогриппозная и противолейкозная) результаты прогноза будут малодостоверны. Минимальное значение точности прогноза при использовании нормальной стратегии составляет 63 % (Fa в перекрестном контроле для антигерпесвирусной активности), максимальное – 100 %; среднее значение изменяется от 82 % до 97 %.

Анализ данных табл. 4 показал, что при использовании рискованной стратегии адекватный прогноз возможен только для 15 видов активности. Для 5 видов активности (ноотропная, фунгицидная, антигерпесвирусная, противолейкозная и противоопухолевая) результаты прогноза будут малодостоверны. Минимальное значение точности прогноза при использовании рискованной стратегии составляет 63 % (Fa в перекрестном контроле для антигерпесвирусной активности), максимальное – 100 %; среднее значение изменяется от 79 % до 96 %.

При совместном применении трех стратегий адекватный прогноз возможен уже для 18 видов

активности. Только для 2 видов активности (фунгицидной и антигерпесвирусной) по всем трем стратегиям результаты прогноза будут малодостоверны.

Сравнение между собой трех стратегий показывает, что по точности они мало отличаются друг от друга: максимальное различие средних значений Fo составляет 4,3 % (между консервативной и рискованной стратегиями в перекрестном контроле), минимальное – 0,6 % (между нормальной и рискованной стратегиями в скользящем контроле). Некоторое преимущество в точности имеет консервативная стратегия. Однако, следует учитывать, что ее применение предусматривает возможность отказа от прогнозирования, т. е. выигрыш в точности может перекрываться потерей части полезной информации. В каждом конкретном случае выбор "лучшей" стратегии прогноза определяется желаемой степенью учета новизны химических структур и характером задачи, которую ставит перед собой исследователь.

**Сравнение систем "Микрокосм" и PASS.**

С целью более жесткой оценки адекватности прогноза фармакологической активности в информационной технологии "Микрокосм" проведено сопоставление полученных результатов с данными о точности прогноза этих же видов активности в системе PASS [7], взятых с сайта <http://www.ibmh.msk.su/pass/> по состоянию на 14 января 2005 г.

Таблица 2

**Точность прогноза различных видов фармакологической активности по результатам тестирования обучающих выборок по консервативной стратегии (%)**

Вид активности	Автопрогноз				Скольльзящий контроль				Перекрестный контроль				Двойной скользящий контроль			
	Fo	Fa	Fn	Fu	Fo	Fa	Fn	Fu	Fo	Fa	Fn	Fu	Fo	Fa	Fn	Fu
Нейролептическая	100	100	99	2	98	99	97	5	97	97	97	4	98	99	97	5
Транквилизирующая	100	100	100	7	90	80	94	16	89	75	95	14	90	80	94	16
Антидепрессантная	100	100	100	5	91	92	90	14	90	94	88	16	91	92	90	14
Ноотропная	99	100	98	14	78	78	78	27	70	71	70	22	78	78	78	27
Анальгетическая нарк.	100	100	99	3	96	97	95	5	95	98	94	5	96	97	95	5
Местноанестезир.	100	100	99	6	90	85	93	8	90	82	94	10	90	85	93	8
Кардиотоническая	100	100	100	8	91	79	96	11	92	80	97	10	91	79	96	11
Кардиостимулир.	100	100	100	4	95	83	96	4	94	74	97	7	95	83	96	4
Гипогликемическая	100	100	99	5	81	81	81	11	82	83	80	12	81	81	81	11
Анаболическая	100	100	100	0	99	99	100	1	99	99	100	1	99	99	100	1
Антисептическая	99	100	99	7	87	82	92	13	86	82	90	14	87	82	92	13
Туберкулостатическая	100	100	100	5	92	86	95	8	91	88	93	7	92	86	95	8
<i>Фунгицидная</i>	100	99	100	7	82	70	92	15	80	68	89	15	82	70	92	15
Анти-ВИЧ	98	95	100	16	81	70	91	25	81	70	90	23	81	70	91	25
<i>Антигерпесвирусная</i>	97	96	99	15	73	61	85	25	72	58	85	20	73	61	85	25
Антипикорнавирусная	98	98	98	10	86	86	86	13	88	90	86	13	86	86	86	13
<i>Противогриппозная</i>	100	100	100	7	91	67	100	4	91	67	100	3	91	67	100	4
Противолейкозная	98	97	99	15	81	85	75	25	79	82	72	21	81	85	75	25

<i>Противоопухолевая</i>	100	100	100	4	88	95	69	18	87	94	65	16	88	95	69	18
Канцерогенная	100	99	100	6	89	85	93	15	87	82	92	14	89	85	93	15
<b>Среднее</b>	<b>99</b>	<b>99</b>	<b>99</b>	<b>7</b>	<b>88</b>	<b>83</b>	<b>90</b>	<b>13</b>	<b>87</b>	<b>82</b>	<b>89</b>	<b>13</b>	<b>88</b>	<b>83</b>	<b>90</b>	<b>13</b>

Примечание. Fo – общая точность прогноза, %; Fa – точность прогноза активных соединений, %; Fn – точность прогноза неактивных соединений, %; Fu – отказ от прогноза, %.

Таблица 3

**Точность прогноза различных видов фармакологической активности по результатам тестирования обучающих выборок по нормальной стратегии (%)**

Вид активности	Метод прогноза	Автопрогноз			Скольльзящий контроль			Перекрестный контроль			Двойной скольльзящий контроль		
		Fo	Fa	Fn	Fo	Fa	Fn	Fo	Fa	Fn	Fo	Fa	Fn
Нейролептическая	Ближ. соседа	100	100	100	96	95	97	95	93	95	96	95	97
Транквилизирующая	Ближ. соседа	100	100	100	84	81	86	83	77	87	84	81	86
Антидепрессантная	Локал. распр.	99	100	99	85	84	86	82	80	83	85	84	86
<i>Ноотропная</i>	Локал. распр.	94	92	96	71	71	72	69	66	71	71	71	72
Анальгетическая нарк.	Ближ. соседа	100	100	100	96	94	97	95	94	96	96	94	97
Местноанестезир.	Ближ. соседа	100	100	100	90	86	92	91	86	94	90	86	92
Кардиотоническая	Ближ. соседа	100	100	100	91	88	92	88	84	90	91	88	92
Кардиостимулир.	Ближ. соседа	100	100	100	94	81	96	94	81	96	94	81	96
Гипогликемическая	Ближ. соседа	100	100	100	80	82	78	78	78	78	80	82	78
Анаболическая	Расстояния	99	100	99	99	99	99	98	98	99	99	99	99
Антисептическая	Локал. распр.	98	98	98	84	81	86	82	78	85	84	81	86
Туберкулоустатическая	Расстояния	92	92	92	89	87	90	88	85	90	89	87	90
<i>Фунгицидная</i>	Байеса	89	83	95	78	66	87	75	66	82	78	66	87
Анти-ВИЧ	Локал. распр.	91	90	92	77	76	77	75	73	76	77	76	77
<i>Антигерпесвирусная</i>	Локал. распр.	91	89	93	71	67	74	72	63	80	71	67	74
Антипикорнавирусная	Ближ. соседа	100	100	99	80	86	75	82	88	76	80	86	75
<i>Противогриппозная</i>	Расстояния	94	78	100	92	67	100	93	72	100	92	67	100
<i>Противолеязычная</i>	Ближ. соседа	100	100	99	81	85	71	75	80	65	81	85	71
Противоопухолевая	Байеса	90	88	92	81	83	77	78	81	73	81	83	77
Канцерогенная	Локал. распр.	99	99	99	85	83	88	83	81	84	85	83	88
<b>Среднее</b>		<b>97</b>	<b>95</b>	<b>98</b>	<b>85</b>	<b>82</b>	<b>86</b>	<b>84</b>	<b>80</b>	<b>85</b>	<b>85</b>	<b>82</b>	<b>86</b>

Таблица 4

**Точность прогноза различных видов фармакологической активности по результатам тестирования обучающих выборок по рискованной стратегии (%)**

Вид активности	Метод прогноза и тип дескриптора	Автопрогноз			Скольльзящий контроль			Перекрестный контроль			Двойной скольльзящий контроль		
		Fo	Fa	Fn	Fo	Fa	Fn	Fo	Fa	Fn	Fo	Fa	Fn
Нейролептическая	Ближ. соседа 4	100	100	100	96	95	96	94	93	94	96	95	96
Транквилизирующая	Ближ. соседа 4	100	100	100	82	82	83	81	77	83	82	82	83
Антидепрессантная	Локал. распр. 8	98	98	99	84	81	85	81	77	83	84	81	85
<i>Ноотропная</i>	Локал. распр. 4	92	91	93	74	74	73	67	69	64	74	74	73
Анальгетическая нарк.	Ближ. соседа 4	100	100	100	94	95	94	94	95	94	94	95	94
Местноанестезир.	Ближ. соседа 11	100	100	100	89	86	90	91	87	93	89	86	90
Кардиотоническая	Ближ. соседа 11	100	100	100	89	86	90	84	81	86	89	86	90
Кардиостимулир.	Ближ. соседа 5	100	100	100	94	81	96	93	70	96	94	81	96
Гипогликемическая	Расстояния 11	95	95	96	79	71	87	80	74	86	79	71	87
Анаболическая	Локал. распр. 9	100	100	100	99	99	99	98	100	96	99	99	99
Антисептическая	Байеса 11	92	90	94	82	77	87	81	77	85	82	77	87
Туберкулоустатическая	Расстояния 8	93	91	94	89	85	91	87	83	89	89	85	91
<i>Фунгицидная</i>	Байеса 11	94	90	98	79	67	88	75	64	83	79	67	88
Анти-ВИЧ	Локал. распр. 9	89	90	88	73	73	73	73	75	71	73	73	73
<i>Антигерпесвирусная</i>	Локал. распр. 8	88	88	89	69	69	70	67	64	70	69	69	70

Антипикорнавирусная	Байеса 11	88	94	83	81	86	77	81	83	79	81	86	77
Противогриппозная	Расстояния 1	93	72	100	93	72	100	93	72	100	93	72	100
Противолеиозная	Ближ. соседа 10	100	100	99	81	87	69	75	82	60	81	87	69
Противоопухолевая	Байеса 11	95	94	98	83	86	78	82	88	68	83	86	78
Канцерогенная	Локал. распр. 9	97	97	98	81	80	81	80	79	81	81	80	81
<b>Среднее</b>		<b>96</b>	<b>95</b>	<b>96</b>	<b>85</b>	<b>82</b>	<b>85</b>	<b>83</b>	<b>79</b>	<b>83</b>	<b>85</b>	<b>82</b>	<b>85</b>

В системе PASS используется только одна оценка точности прогноза – максимальная ошибка прогноза МЕР в скользящем контроле; величина 100 – МЕР соответствует в системе "Микрокосм" общей точности прогноза Fo по результатам скользящего контроля. Русским названиям активностей сопоставлены аналогичные английские термины из системы PASS (данные для противолеиозной активности на сайте отсутствуют).

Результаты приведены в табл. 5, анализ данных которой показал следующее.

Информационная технология "Микрокосм" в целом позволяет более точно прогнозировать фармакологическую активность, чем система PASS: средняя точность прогноза по консервативной стратегии в системе "Микрокосм" превышает среднюю точность прогноза в системе PASS на 9,8 %.

По 16 видам активности точность прогноза в системе "Микрокосм" существенно (от 4,8 % до 20,9 %) превышает аналогичный показатель в системе PASS, в том числе:

- более 20 % – для 2 активностей: нейролептическая, противогриппозная;

- от 10 до 20 % – для 7 активностей: антипикорнавирусная, кардиостимулирующая, анаболическая, антисептическая, кардиотоническая, антигерпесвирусная, ноотропная;

- до 10 % – для 7 активностей: противоопухолевая, туберкулостатическая, антидепрессантная, транквилизирующая, анальгетическая наркотическая, местноанестезирующая, канцерогенная.

По одному виду активности (фунгицидная) превышение статистически незначимо и составляет 1,3 %.

По двум видам активности точность прогноза в системе PASS незначительно превышает точность прогноза в системе "Микрокосм": для гипогликемической на 1,9 %, для анти-ВИЧ на 1,4 %; такие отличия статистически незначимы.

Следует также подчеркнуть, что в системе PASS используется только один метод прогноза и не используются интегральные решающие правила (стратегии). Напротив, в информационной технологии "Микрокосм" применяются 4 метода построения первичных решающих правил и реализованные на их основе 3 стратегии прогноза, в дополнение к мощному языку описания структуры соединений QL.

Таблица 5

**Сравнение точности прогноза различных видов фармакологической активности по результатам скользящего контроля в системах "Микрокосм" и PASS**

Название активности <sup>1</sup>		Точность прогноза, %				
Микрокосм	PASS	Микрокосм, Fo <sup>2</sup>			PASS, 100-MEP	Δ <sup>3</sup>
		Конс.	Норм.	Риск.		
Нейролептическая	Acute neurologic disorders treatment	97,7	96,3	95,7	76,8	20,9
Противогриппозная	Antiviral (influenza)	91,3	91,7	93,1	71,2	20,1
Антипикорнавирусная	Antiviral (picornavirus)	86,1	80,5	81,1	67,2	18,9
Кардиостимулир.	Cardiovascular analeptic	95,1	94,4	94,4	77,0	18,1
Анаболическая	Anabolic	99,4	98,9	98,9	85,5	13,9
Антисептическая	Antiseptic	87,2	83,6	82,0	74,7	12,5
Кардиотоническая	Cardiotonic	90,8	90,8	88,8	79,9	10,9
Антигерпесвирусная	Antiviral (herpes)	73,1	70,9	69,4	62,2	10,9
Ноотропная	Nootropic	78,2	71,4	73,6	68,0	10,2
Противоопухолевая	Antineoplastic	88,3	80,9	83,1	78,5	9,8
Туберкулостатическая	Antituberculosic	92,1	88,6	88,9	83,0	9,1
Антидепрессантная	Antidepressant	90,6	84,9	83,6	81,5	9,1
Транквилизирующая	Anxiolytic	89,5	84,2	82,3	82,3	7,2
Анальгетическая нарк.	Analgesic, opioid	95,7	95,6	94,4	89,1	6,6
Местноанестезир.	Anesthetic local	90,3	89,8	88,6	84,6	5,7
Канцерогенная	Carcinogenic	88,9	85,4	80,7	84,2	4,8

Фунгицидная	Antifungal	82,1	77,7	78,8	80,8	1,3
Анти-ВИЧ	Antiviral (HIV)	81,4	76,6	72,8	82,8	-1,4
Гипогликемическая	Hypoglycemic	81,4	80,0	79,1	83,3	-1,9
<b>Среднее</b>		<b>88,4</b>	<b>85,4</b>	<b>84,7</b>	<b>78,6</b>	<b>9,8</b>

<sup>1</sup> Названия активностей приведены в порядке уменьшения значения  $\Delta$  (см. п. 3).

<sup>2</sup> По консервативной, нормальной и рискованной стратегиям соответственно.

<sup>3</sup> Разница между  $F_0$  по консервативной стратегии в системе "Микрокосм" и 100-МЕР в системе PASS.

Таким образом, полученные результаты позволяют утверждать, что точность прогноза фармакологической активности (на примере 19 различных ее видов) при использовании информационной технологии "Микрокосм" существенно выше, чем при использовании системы PASS.

### ЗАКЛЮЧЕНИЕ

Информационная технология "Микрокосм" позволяет с высокой точностью проводить прогноз различных видов фармакологической активности структурно разнородных химических соединений.

Точность прогноза фармакологической активности с использованием информационной технологии "Микрокосм" значительно превышает точность прогноза активности с использованием системы PASS.

### ЛИТЕРАТУРА

1. Васильев П.М. // Тез. докл. II Российской школы-конференции / СГУ. – Саратов, 2004. – С. 8–9.
2. Васильев П.М. // Тез. докл. XI Российского Национального конгресса "Человек и лекарство". – М., 2004. – С. 586.

Vassiliev P.M., Spasov A.A. Computerized information technology in prognosis of pharmacologic activity of structurally heterogenous chemical compounds // Vestnik of Volgograd State Medical University. – 2005. – № 1. – P. 23–30.

УДК 615.3:547.466.3:615.22

## АНТИШЕМИЧЕСКОЕ ДЕЙСТВИЕ ТОЛИЛЬНЫХ ПРОИЗВОДНЫХ ГАММА-АМИНОМАСЛЯНОЙ КИСЛОТЫ

В.Н. Перфилова, И.Н. Тюренок

Кафедра фармакологии и биофармации ФУВ, НИИ фармакологии ВолГМУ

Экспериментальные данные, полученные на кафедре фармакологии Волгоградского медицинского института, свидетельствуют, что фенильные [1, 3, 5, 6, 9, 11] и толильные [7, 10] производные гамма-аминомасляной кислоты (ГАМК) обладают широким спектром фармакологической активности.  $\beta$ -толил-гамма-аминомасляная кислота (толибут) по некоторым видам активности превосходила  $\beta$ -фенил-гамма-аминомасляную кислоту (фенибут), но последнее, являясь менее токсичным соединением, было внедрено в практику.

Накопленные к настоящему времени экспериментальные данные показывают, что структурные аналоги ГАМК обладают антиангинальной

активностью [2, 4, 8, 11, 12]. Данные по противошумному действию толибутов в литературе отсутствуют.

Учитывая положительный опыт создания высокоэффективных и малотоксичных веществ (пикамилаона и пантогама) на основе ГАМК путем химического связывания с фармакологически активными соединениями – никотиновой и пантотеновой кислотами представляется целесообразным изучить антиангинальную активность новых производных толибутов, включающего в свою структуру лимонную, янтарную, яблочную, щавелевую, глутаминовую и никотиновую кислоты.