

Работа выполнялась при финансовой поддержке Федерального агентства по образованию РФ по программе "Развитие научного потенциала высшей школы" 2005 г. (подпрограмма 3, раздел 3, код проекта 4507).

## ИНФОРМАЦИОННАЯ ТЕХНОЛОГИЯ ПРОГНОЗА БИОЛОГИЧЕСКОЙ АКТИВНОСТИ ХИМИЧЕСКИХ СОЕДИНЕНИЙ "МИКРОКОСМ"

П.М. Васильев

Волгоградский государственный медицинский университет

Использование компьютерных информационных технологий (ИТ) для предварительного планирования экспериментальных испытаний новых биологически активных веществ, в частности потенциальных лекарственных препаратов, является одним из актуальных направлений современной фармакологии.

Информационная технология прогноза свойств органических соединений "Микрокосм" создана в период с 1985 по 2006 гг. как совокупность теоретических концепций, математических методов и правил и основанных на них компьютерных алгоритмов и программ, позволяющих расчетным способом оценивать активность химического соединения по его структурной формуле.

ИТ "Микрокосм" показала высокую точность при прогнозе различных видов фармакологической активности (до 98 %), противовирусных и антибактериальных свойств (до 92 %), канцерогенной опасности (89 %), свойств добавок в полимерные композиты (до 99 %), свойств добавок в резиновые смеси (до 97 %).

Парадигму ИТ "Микрокосм" составляют несколько основных теоретических концепций: химические динамические системы высокой сложности, обобщенный образ класса соединений с заданными свойствами, мультидескрипторное иерархическое многоуровневое описание структуры химических соединений, контекстно-зависимые локальные метрики, мегамерные и k-мерные пространства, взаимодополнение решающих правил при компьютерном прогнозе свойств химических соединений, стратегии компьютерного прогноза свойств химических соединений, комплексный подход к компьютерному прогнозу свойств химических соединений.

На базе этих концепций разработаны основные компоненты ИТ "Микрокосм": специализированный мультидескрипторный иерархический многоуровневый язык QL для описания структуры органических соединений, метод построения моделей обобщенных образов классов соединений с заданными свойствами, математические методы построения решающих правил на основе теории распознавания образов, методы принятия решений при различных стратегиях прогноза, метод проверки результатов прогноза на непротиворечивость, метод формирования фармакофоров-образов заданной активности.

Теоретические концепции ИТ "Микрокосм" и

разработанные на их основе методы реализованы в виде единого программного комплекса, состоящего из 12 взаимосвязанных программ: *TranQL2* – транслятор химических структурных формул в коды языка QL; *MakeData* – формирование и упаковка матрицы данных; *ActUtil* – индексирование вида и уровня биологической активности; *Testing* – расчет решающих правил, оценка точности прогноза, прогноз активности неиспытанных соединений; *TestPrognResult* – проверка спектра прогнозных оценок уровня активности на непротиворечивость; *FarmFor* – выявление значимых признаков активности/неактивности ("фармакофоров"); *QL2Txt* – конвертация матрицы структурных данных в текстовую форму; *NegwerAct* – формирование наборов структур соединений с заданным видом биологической активности по базе данных лекарственных препаратов мирового ассортимента; *Str2Sdf*, *u2d*, *Mol2Hin* – конвертеры различных форматов представления структуры химических соединений; *Elsa* – быстрый расчет зарядов на атомах.

Программный комплекс ИТ "Микрокосм" написан на языке программирования Delphi-6 как приложение под Windows-XP; общий объем исходных текстов версии 4.1 (март 2006 г.) – более 30 тыс. строк. ИТ "Микрокосм" имеет мощное информационное наполнение: базы данных системы содержат в общей сложности около 1 млн структур соединений по почти 3 тыс. видам биологической активности.

От других компьютерных систем аналогичного назначения ИТ "Микрокосм" принципиально отличается комплексным подходом к прогнозу, в рамках которого для расчета итоговой оценки активности используется "веер" решающих правил, полученных различными способами; при этом ошибки нескольких методов взаимно компенсируют друг друга, что обеспечивает высокую точность прогноза. Формируемая в терминах языка QL модель обобщенного образа включает все доступные структурные дескрипторы и представляет собой в некотором смысле "гиперструктурную" формулу активных соединений обучающей выборки, построенную на 11 уровнях QL-описания. Применение четырех методов расчета решающих правил (Байеса, расстояния, ближайшего соседа, локального распределения) отдельно по каждому из этих 11 уровней описания позволяет для каждого соединения получать спектр из 44 прогнозных оценок активности. Это дает возможность использовать для обобщения этих 44 оценок и выработки итоговых заключений о наличии активности три различных стратегии прогноза: консервативную, нормальную и рисковую.

ИТ "Микрокосм" позволяет также по структуре соединений выполнять количественную оценку уровня их биологической активности. В этом случае для каждой градации уровня заданной активности (высокая, умеренная, низкая, отсутствие) производится расчет по трем стратегиям прогноза четырех наборов решающих правил – всего получается 12 зависимостей. При прогнозе уровня активности нового соединения отдельно по каждой стратегии рассчитывают три набора оценок – четыре оценки в наборе,

по одной для каждой градации активности. Затем с целью повышения достоверности прогноза производится проверка результатов расчетов на непротиворечивость путем сравнения наборов полученных расчетных оценок с шаблонами уровней активности, формируемых по уровням активности соединений обучающей выборки.

Использование в ИТ "Микрокосм" в качестве основного способа представления химической структуры многомерной (до нескольких сотен тысяч переменных) модели обобщенного образа класса соединений с заданными свойствами позволяет формировать на основе этой модели новый класс объектов – фармакофоры-образы: высокодостоверные интегративные структурные признаки наличия или отсутствия у соединений данного вида активности, в том числе ее высокого уровня. Фармакофор-образ объединяет в одной строке QL-описания (общей формуле) множество структурно-сходных "традиционных" фармакофоров – фиксированных фрагментов химической структуры, отвечающих за наличие у соединений этого вида биологической активности.

При построении фармакофоров-образов на первом этапе из множества всех дескрипторов QL-описания соединений обучающей выборки формируется с использованием гипергеометрического критерия список первичных статистически значимых QL-признаков заданного вида активности. Затем путем попарного сравнения всех признаков в списке выявляются пары QL-дескрипторов, имеющие совпадающие или структурно близкие особенности; эти пары объединяются в промежуточные "общие" признаки. Процедура повторяется до полного исчерпания возможностей попарного объединения признаков – как первичных, так и промежуточных.

Структуру прогнозируемого соединения сравнивают с фармакофорами-образами и определяют, есть ли в ней реализация данного образа в виде конкретного "классического" фармакофора – фиксированного фрагмента химической структуры; если эта реализация присутствует, соединение рекомендуют для последующих экспериментальных испытаний.

Результаты более чем 10-летнего практического использования и многократного независимого тестирования позволяют утверждать, что "Микрокосм" является мощной современной универсальной информационной технологией, позволяющей по структуре химических соединений производить прогноз их биологической активности и выполнять направленный поиск и конструирование веществ с максимальным уровнем полезного эффекта и минимальным побочным действием, что существенно сокращает временные материальные и финансовые затраты при поиске новых лекарственных препаратов.

**КОМПЬЮТЕРНЫЙ СКРИНИНГ НОВЫХ  
ПРОИЗВОДНЫХ БЕНЗИМИДАЗОЛА  
С ВЫСОКОЙ 5-НТ<sub>3</sub>-АНТИСЕРОТОНИНОВОЙ  
АКТИВНОСТЬЮ НА БАЗЕ  
ИНФОРМАЦИОННОЙ ТЕХНОЛОГИИ  
"МИКРОКОСМ"**

**М.В. Черников, П.М. Васильев**

*Волгоградский государственный медицинский университет*

Серотонин является одним из важнейших биологических аминов, играющих ведущую роль как в регуляции многих физиологических функций, так и при ряде серьезных патологических состояний. Поэтому поиск избирательных антагонистов различных подтипов серотониновых рецепторов (в частности 5-НТ<sub>3</sub>-антагонистов) является одной из актуальных задач современной фармакологии.

Ранее на кафедре фармакологии ВолГМУ экспериментально было показано наличие у некоторых новых конденсированных производных бензимидазола достаточно выраженной 5-НТ<sub>3</sub>-антисеротониновой активности. Однако проведение систематических массовых испытаний более чем 1000 синтезированных к настоящему времени соединений этого класса представляет собой практически невыполнимую задачу из-за высокой стоимости и трудоемкости таких исследований.

В связи с этим была поставлена цель: создать на базе информационной технологии "Микрокосм" систему компьютерного скрининга новых высокоактивных 5-НТ<sub>3</sub>-антагонистов среди новых производных бензимидазола с тем, чтобы отобрать из множества синтезированных соединений наиболее перспективные для последующего экспериментального изучения вещества.

При выполнении настоящего исследования было проделано следующее.

1. Сформирована обучающая выборка, в которую были включены структуры новых конденсированных производных бензимидазола, испытанных на 5-НТ<sub>3</sub>-антисеротониновую активность на изолированных атропинизированных предсердиях морских свинок. Всего было испытано 94 соединения, разделенных на 4 группы (высокоактивные, умеренноактивные, низкоактивные, неактивные) в соответствии с изменением положительного хронотропного эффекта, вызванного серотонином.

2. Для целей прогноза определены 6 градаций 5-НТ<sub>3</sub>-антисеротониновой активности: "активные", "высокоактивные", "умеренноактивные", "низкоактивные", "высоко- или умеренноактивные", "умеренно- или низкоактивные".

3. С использованием информационной технологии "Микрокосм" рассчитаны 18 решающих правил (три стратегии прогноза, шесть градаций активности) для прогноза уровня 5-НТ<sub>3</sub>-антисеротониновой активности производных бензимидазола.

4. Произведена оценка прогностической точности каждого из полученных решающих правил с применением процедуры скользящего контроля. Максимальное значение точности прогноза составило 80 % при прогнозе по градации "высокоактивные" с использованием консервативной стратегии.

5. Для каждого из 952 неиспытанных производных бензимидазола по полученным 18 решающим правилам произведен расчет спектра прогнозных оценок каждой из 6 градаций уровня 5-НТ<sub>3</sub>-