

АВТОМАТИКА. ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА

УДК 004.89:004:8

МРНТИ 50.53.19, 28.23.02

Г. А. Самигулина, д.т.н., **С. В. Чебейко**, к.х.н.

Институт проблем информатики и управления

ИНТЕЛЛЕКТУАЛЬНАЯ ТЕХНОЛОГИЯ ПРОГНОЗИРОВАНИЯ ФАРМАКОЛОГИЧЕСКОЙ АКТИВНОСТИ ОРГАНИЧЕСКИХ СОЕДИНЕНИЙ

Разработан иммунно-сетевой подход к моделированию зависимостей «структура - свойство» лекарственных препаратов. Предложенная интеллектуальная технология позволяет уменьшить погрешности энергетических оценок и повысить достоверность прогноза зависимости «структура - свойство» химических соединений.

Ключевые слова: прогнозирование фармакологической активности, интеллектуальная технология, лекарственные препараты.



Дерілік препараттардың «құрылым - қасиет» тәуелділіктерін моделдеуге иммунды желілік тәсіл әзірленді. Ұсынылған зияткерлік технология энергетикалық бағалау қателіктерін азайтуға және химиялық қосылыстардың «құрылым -қасиет» тәуелділігін болжаудың дұрыстығын арттыруға мүмкіндік береді.

Түйінді сөздер: фармакологиялық белсенділікті болжау, зияткерлік технология, дерілік препараттар.



We have developed immune net method for modeling structure - property dependences of medical drugs. The proposed intellectual technology allows to reduce errors of power estimations and to raise reliability of the forecast of structure - property dependence of chemical compounds.

Key words: forecasting pharmacological activity, intellectual technology, medical drugs.

Исследование зависимости биологической активности от структуры химических соединений и поиск на этой основе но-

вых высокоактивных лекарственных веществ - главные проблемы современной фармакологии.

Синтез новых лекарственных веществ является результатом развития фундаментальных знаний и достижений медицинских, биологических, химических и других наук, проведения напряженных экспериментальных исследований, длительного времени, вложения крупных материальных затрат. Например, выявление одного перспективного лекарственного вещества требует изучения не менее 10 тыс. соединений. Разработка нового лекарства длится 12-15 лет, а затраты составляют от 500 млн. до 1,9 млрд. дол. [1].

Компьютерное моделирование все более уверенно входит в практику технологии создания новых синтетических лекарственных веществ. История дизайна с помощью компьютеров началась более 25 лет назад, когда стало возможным изображение и вращение молекул на экране компьютера. В настоящее время начальным этапом поиска фармакологически активных веществ, как правило, является использование методов *in silico*, предваряющих экспериментальные исследования *in vitro* и *in vivo*. Для этих целей разработаны специализированные информационные технологии и на их основе созданы интегрированные системы и программные среды. Российская академия медицинских наук определила в качестве одного из перспективных направлений развития медицинской науки на период до 2025 г. разработку методов оценки безопасности и эффективности лекарственных препаратов на основе технологий биоинформатики и компьютерного конструирования лекарств [2].

Компьютерный молекулярный дизайн основан на концепции взаимосвязи молекулярной структуры и биологической активности химических соединений. Методы компьютерного моделирования лекарств - виртуальный скрининг, молекулярный докинг, искусственные нейронные сети, конструирование лигандов методами *de novo*, QSAR, CoMFA, CoMSIA - направлены на ускорение и оптимизацию поиска новых биологически активных веществ. Наиболее эффективными при исследовании зависимости «структура - свойство» являются подходы, решаемые с по-

мощью доступных компьютерных программ: SARD, CoMFA, MFTA, PASS, Микрокосм, DRAGON [3].

Кроме того, для прогнозирования зависимости «структура - свойство» в последнее время широко используется подход искусственных нейронных сетей. При моделировании биологической активности наиболее часто применяется многослойная нейронная сеть прямого распространения ошибки. С использованием данного подхода получены многочисленные результаты при прогнозировании токсических и генотоксических свойств соединений, магнитной восприимчивости разнообразных органических соединений и др. К основным недостаткам данного подхода относятся:

- сложность выбора параметров;
- трудности, связанные с выбором структуры нейронной сети;
- ресурсоемкий характер обучения нейронной сети;
- эффект переобучения нейронной сети.

Моделирование биологической активности органических соединений возможно и с помощью нового биологического направления искусственного интеллекта – искусственных иммунных систем (Artificial Immune Systems). Первые разработки по данной тематике появились в 70-х гг. прошлого столетия. Это адаптивные системы для обработки и анализа многомерных данных, которые представляют собой математическую структуру, имитирующую некоторые функции иммунной системы человека и обладающие следующими свойствами: способностью к обучению, прогнозированию и принятию решений в незнакомой ситуации. Введено понятие формального пептида как математическая абстракция свободной энергии белковой молекулы от ее пространственной формы [4]. В данном биологическом подходе используется идея молекулярного узнавания между антителами человека и чужеродными антигенами посредством определения минимальной энергии связи (решение задачи распознавания образов) между формальными пептидами.

Постановка задачи формулируется следующим образом: необходимо разработать интеллектуальную систему прогнозирования фармакологической активности органических соедине-

ний на основе биологического подхода искусственных иммунных систем.

Входными параметрами интеллектуальной системы являются дескрипторы (числовые характеристики, описывающие химическое соединение). Интеллектуальный анализ данных предусматривает выделение скрытых закономерностей или взаимосвязей между дескрипторами в больших массивах необработанных данных. При реализации поставленной задачи проблема создания оптимальных наборов дескрипторов, позволяющих в удобной форме использовать их в вычислительном процессе, является одной из актуальных.

Предлагаемая технология иммунно-сетевого моделирования зависимости «структура - свойство» включает следующую последовательность действий [5, 6]:

- описываются структуры исследуемых соединений числовыми параметрами (дескрипторами);
- создаются базы данных дескрипторов;
- осуществляется предварительная обработка данных, а именно нормирование, центрирование; заполнение пропущенных данных (методом заполнения средним, методом заполнения главным компонентом, методом случайного заполнения) и т.д.;
- выполняется оценка информативности дескрипторов на основе методов факторного анализа (главных компонент);
- выбирается оптимальный набор дескрипторов;
- производится редукция малоинформативных дескрипторов;
- строится оптимальная структура иммунной сети;
- весь массив данных разбивается на обучающую и контролирующую выборку;
 - выбирается метод обучения иммунной сети: с учителем или без учителя;
 - выбираются параметры для оценки качества обучения иммунной сети;
 - производится обучение иммунной сети, оцениваются результаты этого обучения;
 - решается задача распознавания образов на основе сингулярного разложения матриц и нахождения минимальной энергии связывания между формальными пептидами (антителами и антигенами);

- осуществляется оценка решения задачи распознавания образов на основе гомологов и расчет коэффициентов риска прогнозирования на основе ИИС;

- полученная иммунно-сетевая модель используется для прогноза свойств и активностей неизвестных соединений.

В общем виде можно выделить следующие основные процедуры, выполняемые при функционировании интеллектуальной системы прогнозирования на основе искусственных иммунных систем:

Процедура 1. Предварительная обработка дескрипторов. Необходимым условием является то, что количество используемых дескрипторов одинаково для всех описываемых структур.

Процедура 2. Решение задачи распознавания образов на основе определения минимальной энергии. При выполнении этой процедуры обрабатываются большие массивы дескрипторов, поэтому необходимым условием функционирования интеллектуальной системы является применение алгоритма распараллеливания вычислений для более эффективного и быстрого решения поставленной задачи.

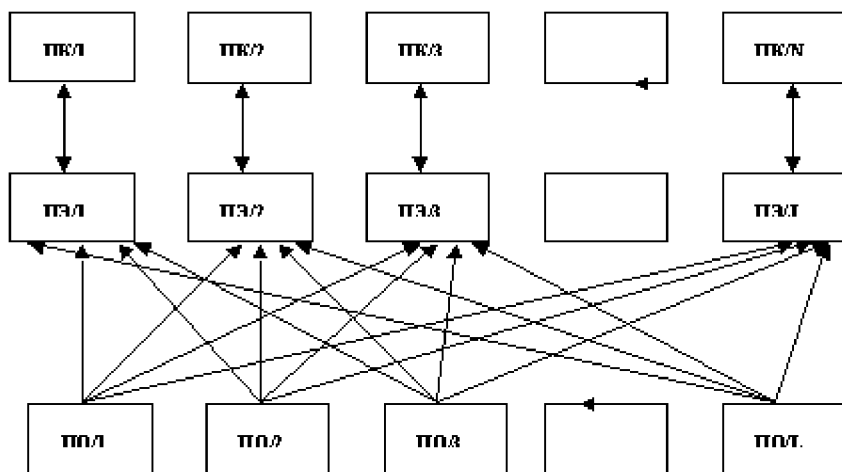


Схема распараллеливания вычислений при иммунно-сетевом моделировании

На рисунке приведена схема распараллеливания вычислений при решении задачи распознавания образов на основе искусственных иммунных систем.

Персональные компьютеры обозначены ПК/1, ПК/2, ПК/3, ... ПК/N, где N – количество компьютеров.

Пептиды «эталонны» обозначены ПЭ/1, ПЭ/2, ПЭ/3, ..., ПЭ/J, где J – количество классов, равное количеству эталонов, при этом N равно J.

Пептиды «образы» обозначены ПЭ/1, ПЭ/2, ПЭ/3, ..., ПЭ/L, где L – количество образов, при этом L больше или равно J.

Алгоритм распараллеливания заключается в следующем. После предварительной обработки с помощью экспертов формируются пептиды «эталонны» (временные ряды, состоящие из информативных дескрипторов) для всех рассматриваемых классов. Каждый «эталон» вводится в отдельный компьютер. Затем из базы данных формируются пептиды «образы». Для определения минимальной энергии связи на основе сингулярного разложения матриц [7] каждый образ сравнивается со всеми «эталонами». Эти вычисления производятся параллельно, что существенно сокращает процесс обработки информации.

Процедура 3. Анализ полученных иммунно-сетевых зависимостей и сравнение используемого алгоритма с другими методами обработки данных (статистическими, нейросетевыми и т.д.).

Достоинством предлагаемой интеллектуальной технологии является возможность (при определенных условиях) использования одновременно нескольких уровней дескрипторов (элементного уровня, структурной формы, электронной структуры, молекулярной формы, межмолекулярного взаимодействия). Применение факторного анализа позволяет выявить латентные взаимодействия между выбранными дескрипторами и проанализировать основополагающие факторы, влияющие на них. Построение оптимальной структуры сети и сокращение времени на обучение осуществляется за счет редукции малоинформативных дескрипторов.

Большая трудность при создании таких систем заключается в способности распознавать образы на границах нелинейно разделенных классов. Например, соединения описываются очень похожими пептидами (временными рядами), но относятся к разным классам с абсолютно непохожими свойствами (биологической активностью). Исполь-

зуемый алгоритм уменьшает погрешности энергетических оценок и повышает достоверность прогноза зависимостей «структура - свойство» химических соединений.

На разработанное программное обеспечение получено авторское свидетельство о государственной регистрации объекта интеллектуальной собственности.

Литература

1 *Кубиньи Г.* В поисках новых соединений-лидеров для создания лекарств // Рос. хим. журн. - 2006. - № 2. - С. 5-17.

2 <http://www.ramrn.ru>

3 *Иванов А. С., Веселовский А.В., Дубанов А. В.* и др. Интегральная платформа «от гена до прототипа лекарства» in silico и in vitro // Рос. хим. журн. - 2006. - № 2. - С. 18-35.

4 *Tarakanov A.O.* Formal peptide as a basic of agent of immune networks: from natural prototype to mathematical theory and applications // Proceedings of the I Int. workshop of central and Eastern Europe on Multi-Agent Systems, 1999. - P. 37.

5 *Самигулина Г.А., Чебейко С.В.* и др. Разработка технологий иммунно-сетевого моделирования для решения различных прикладных задач. - Алматы, 2011. - 217 с.

6 *Самигулина Г. А., Чебейко С. В.* Прогнозирование зависимости «структура - свойство» органических соединений на основе иммунно-сетевого моделирования // Хим. журн. Казахстана. - 2010. - № 3. - С. 164-172.

7 *Самигулина Г. А.* Разработка интеллектуальных экспертных систем прогнозирования и управления на основе искусственных иммунных систем // Проблемы информатики. - 2010. - № 1. - С. 15-22.