

КОМПЬЮТЕРНЫЙ МОЛЕКУЛЯРНЫЙ ДИЗАЙН ЛЕКАРСТВЕННЫХ СРЕДСТВ НА
ОСНОВЕ ИСКУССТВЕННЫХ ИММУННЫХ СИСТЕМ

Г.А. Самигулина, С.В. Чебейко

Институт проблем информатики и управления МОН РК, Казахстан, г. Алматы

Разработан иммуносетевой 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.

Применение современной вычислительной техники и новейших информационных технологий открывает широкие возможности для решения одной из главных проблем современной науки – целенаправленного поиска новых веществ и материалов с заранее заданными свойствами.

Рынок фармацевтической индустрии испытывает недостаток в новых лекарствах. К началу XXI века ежегодно проходили регистрацию и внедрялись в клиническую практику 50-60 новых лекарственных препаратов. За последнее десятилетие количество новых внедренных лекарств существенно сократилось. Традиционными источниками для получения лекарств являются растительные продукты, микробные метаболиты, эндогенные нейромедиаторы, гормоны и классический химический синтез. Экспериментальное тестирование большинства новых соединений исключает их как неперспективные прототипы лекарств из-за низкой целевой активности, сложности синтеза, канцерогенности, высокой токсичности и т.д. Только одно из 100 тыс. исследованных соединений может стать препаратом с выходом на фармацевтический рынок. Общие затраты времени и денег на создание нового препарата составляет 12-15 лет и более 800 млн. долларов [1, 2]. Снижение темпов создания новых лекарств обусловлено многими причинами, одна из которых, возрастание требований к эффективности и безопасности лекарств. Особо следует отметить ситуацию по дефициту новых структур-лидеров, которые могут быть

оптимизированы в терапевтически пригодные лекарства [3].

Создание методов прогнозирования свойств новых химических соединений и направленный молекулярный дизайн соединений с заданным набором свойств являются важнейшими актуальными задачами современной медицинской химии и биоинформатики.

История дизайнера с помощью компьютеров началась более 25 лет назад, когда стало возможным изображение и вращение молекул на экране компьютера [4-9]. Данный метод позволяет разработчикам лекарств сформулировать и оценить рабочие гипотезы «структура – активность». В литературе используются следующие термины: QSAR (Quantities Structure Activity Relationships), который обозначает Количественное Соотношение Структура Активность (КССА) и компьютерное моделирование лекарственных препаратов CAMD (Computer Aided Molecular Design). Работы по молекулярному дизайну лигандов на основе структуры мишени начались с исследований Гудфорда. Немного позже в лечебную практику был внедрен первый лекарственный препарат, полученный методом структурно-ориентированного дизайна - каптоприл. Модели физиологически активных веществ позволяют проводить молекулярный дизайн новых лекарств, основываясь на знании пространственной структуры биомишеней.

Компьютерный молекулярный дизайн основан на концепции взаимосвязи молекулярной структуры и биологической активности химических соединений. О.А. Раевский

[10] выделяет в современном компьютерном дизайне три основных этапа исследований:

1. формирование обучающей выборки соединений с заданным свойством (активностью);
2. описание молекулярной структуры исследуемых соединений (дескрипторов);
3. установление взаимосвязи «структура – свойство (биологическая активность)» с последующим созданием устойчивых прогностических математических моделей.

Количественное описание молекулярной структуры химических соединений в компьютерном молекулярном дизайне осуществляется с помощью дескрипторов. Дескриптор это параметр, который каким-либо образом характеризует определенную химическую структуру. Например, дескриптором может быть любое число, которое можно рассчитать из структурной формулы химического соединения: молекулярный вес, молекулярный объем и т.д. Существует большое количество дескрипторов, но реально полезных для решения поставленной цели достаточно ограниченное количество. Возникает проблема создания дескрипторов наиболее полно характеризующих рассматриваемое соединение и позволяющих в удобной форме использовать их в вычислительном процессе. В настоящее время разработано много разных подходов к выбору оптимальных дескрипторов. В работах [10, 11] приводится систематическая классификация дескрипторов. О.А. Раевским представлена схема нескольких уровней дескрипторов:

- дескрипторы элементного уровня,
- дескрипторы структурной формулы,
- дескрипторы электронной структуры,
- дескрипторы молекулярной формы;
- дескрипторы межмолекулярных взаимодействий.

При таком представлении каждый последующий уровень включает информационное содержание предыдущего. Таким образом, концепция взаимосвязи «структура

– свойство (активность)» является отражением взаимосвязи молекулярных взаимодействий химических соединений между собой и с биологическими объектами.

Построение адекватной компьютерной молекулярной модели позволяет в дальнейшем прогнозировать различные терапевтические и физико-химические свойства синтезируемых молекул, что определяет актуальность и перспективность развития данного научного направления [12].

Среди методов прогнозирования зависимости «структура – свойство» следует отметить рост исследований по искусственным нейронным сетям [13]. Нейронные сети представляют собой упрощенную математическую модель обработки информации головным мозгом человека. Однако большинство современных архитектур нейронных сетей не воспроизводит в точности функции мозга, и их следует рассматривать как разновидность параллельных алгоритмов.

Искусственные нейронные сети представляют собой достаточно мощный вычислительный метод решения задач, возникающих при обработке химической информации, а именно, при классификации и моделировании разрозненных экспериментальных данных. В рамках поиска зависимостей между структурами органических соединений и их биологической активностью наиболее популярна многослойная нейронная сеть прямого распространения, обучающаяся по методу обратного распространения ошибки. Количество скрытых нейронов определяет количество настраиваемых параметров нейросети (весовых коэффициентов).

Достоинством этой сети является способность обобщать и аппроксимировать данные с высокой точностью, а также обрабатывать многомерную разрозненную информацию. Совокупность параметров, подаваемых на вход нейросети, представляют в виде вектора. При этом необходимыми условиями являются:

- количество и тип используемых дескрипторов одинаковы для всех описываемых структур;

- каждому дескриптору соответствует определенная компонента в виде вектора параметров;

- различным молекулам соответствуют различные векторы параметров.

По типу описываемой информации дескрипторы делятся на фрагментные; топологические; квантово-химические; физико-химические.

К недостаткам данного подхода можно отнести сложность выбора параметров и структуры нейронной сети; ресурсоёмкий характер обучения нейросети; сложности связанные с переобучением сети.

Моделирование биологической активности органических соединений также возможно и с помощью нового биологического направления *искусственного интеллекта – искусственных иммунных систем (ИИС)*.

Под искусственными иммунными системами понимаются информационные методологии, использующие понятия теоретической иммунологии для решения различных прикладных задач. Это относительно новая область исследования. Первые попытки разработки ИИС были предприняты в 70-х годах, масштабные работы начались в 90 годах прошлого столетия.

В работах А.О. Тараканова были разработаны математические основы биологического подхода ИИС. Базовым элементом системы является пептид (это простейший белок, который содержит до 50 аминокислот) [14]. Биологическим прототипом ИИС является иммунная система человека и обработка информации молекулами белков на основе результатов самосборки. В основе подхода ИИС лежит идея взаимодействия между белками иммунной системы человека и чужеродными антигенами.

Тараканов А.О. ввел математическую модель формального пептида, как математи-

ческой абстракции свободной энергии белковой молекулы от ее пространственной формы, описанной в алгебре кватернионов.

Кватернион - это число, составленное из действительной единицы 1 и трех мнимых единиц: i_1, i_2, i_3 с действительными элементами следующего вида:

$$\Lambda = (\lambda_0, \lambda_1, \lambda_2, \lambda_3) = \lambda_0 1 + \lambda_1 i_1 + \lambda_2 i_2 + \lambda_3 i_3.$$

Комплексное число является частным случаем кватерниона. Кватернионы служат для описания тела в пространстве. Любой кватернион можно представить в пространстве точкой или радиус вектором. Кватернионы чрезвычайно удобны для описания пространственного строения молекул белка ввиду дуализма кватернионных единиц, являющихся с одной стороны, ортами реального трехмерного пространства, а с другой, операторами преобразования. Каждый кватернион описывает положение определенного звена полипептидного остова в пространстве.

Формальным пептидом (ФП) называют упорядоченную пятерку:

$$P = \langle n, U, Q, V, v \rangle,$$

которая включает следующие компоненты:

- количество звеньев $n > 0$;

- множество торсионных углов:

$$U = \{ \varphi_k, \psi_k \}, k=1, \dots, n,$$

где $-\pi \leq \varphi_k \leq \pi, -\pi \leq \psi_k \leq \pi$;

- множество единичных кватернионов:

$$Q = \{ Q_0, Q_k \},$$

где кватернионы $Q_k = Q_k(\varphi_k, \psi_k)$ и результирующий кватернион ФП Q_0 определяется как их произведение: $Q_0 = Q_1 Q_2 \dots Q_n$;

- множество коэффициентов $V = \{ v_{ij} \}, i = 1, 2, 3, 4, j \geq i$;

- функция v (без индекса), определенная на элементах результирующего кватерниона Q_0 следующей квадратичной формой:

$$v = - \sum_{j \geq i} v_{ij} q_i q_j.$$

Рассматриваются взаимодействия между белками иммунной системы человека и чужеродными антигенами, процедура молекулярного узнавания посредством опреде-

ления минимальной энергии связи между формальными пептидами. Достоинствами иммунной системы являются: распределенность; самоорганизация и эволюция; относительная простота и «легковесность», т.е. ИИС не особенно требовательна к вычислительным ресурсам; отсутствие централизованного контроля; самообучаемость; индивидуальный подход к уникальным событиям.

При иммунносетевом моделировании зависимости «структура – свойство» выполняется следующая последовательность действий:

- описываются структуры исследуемых соединений числовыми параметрами (дескрипторами);

- осуществляется предварительная обработка данных [15]: нормирование, центрирование; заполнение пропущенных данных (методом заполнения средним, методом заполнения главным компонентом, методом случайного заполнения) и т.д.

- выбирается оптимальный набор дескрипторов, строится оптимальная структура иммунной сети [16];

- весь массив данных разбивается на обучающую и контролирующую выборки;

- 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 и нахождения минимальной энергии связывания между формальными пептидами (антителами и антигенами);

- осуществляется оценка решения задачи распознавания образов на основе гомологов [17] и расчет коэффициентов риска прогнозирования на основе ИИС;

- полученную иммунносетевую модель используют для прогноза свойств и активностей неизвестных соединений.

Укрупненная структурная схема иммунносетевого моделирования «структура-свойство» химических соединений приведена на рисунке 1. Здесь D_1, \dots, D_m исходный набор дескрипторов. На основе пакета прикладных программ SPSS 16.0 (факторного анализа и метода главных компонент) и разработанного пакета прикладных программ «PEPTID» [18, 19] осуществляется редукция малоинформативных признаков и нахождение оптимального набора дескрипторов D_1, \dots, D_n .

При решении задачи распознавания образов на основе подхода ИИС экспертами формируются матрицы эталонов (которые получаются в результате свертывания соответствующих временных рядов с информативными дескрипторами) и матрицы образов (которые также получаются в результате сворачивания временных рядов в матрицы). После того, как мы обучили ИИС, то есть ввели матрицы эталонов в компьютер для соответствующих классов, мы вводим матрицу образов и находим меру близости нашего образа с эталонами. Минимальное значение энергии связи указывает, к какому классу принадлежит образ.

Достоинством предложенной интеллектуальной технологии на основе иммунносетевого моделирования является:

- способность системы глубоко анализировать скрытые (латентные) взаимодействия между дескрипторами и основополагающие факторы, влияющие на них;

- распознавать пептиды, находящиеся на границе классов (имеющие схожие структуры);

- сокращение времени на обучение иммунной сети за счет построения оптимальной структуры и редукции дескрипторов, несущих существенные погрешности;



Рис. 1. Структурная схема иммуносетевого моделирования «структура – свойство» химических соединений.

- уменьшение погрешностей энергетических оценок, так называемых ошибок обобщения; повышение достоверности прогноза зависимостей «структура – свойство» химических соединений.

На разработанное программное обеспечение получены два авторских свидетельства о государственной регистрации объекта

интеллектуальной собственности, зарегистрированные в Комитете по правам интеллектуальной собственности Министерства юстиции Республики Казахстан [20, 21].

Таким образом, иммунносетевой подход может быть с успехом применен к моделированию зависимостей «структура – свойство» лекарственных препаратов.

Литература: [1] *Frantz S., Smith A.* Nature Rev. Drug. Discov. -2003. -V.2. -P.95-96.; [2] *Г.Кубиньи.* Российский химический журнал. -2006. - №2. -С.5-17.; [3] *Lohse M. J.* Trends Pharmacol Sci. -1998. -V.19. -P.198.; [4] *Goodford P.J.* J. Med. Chem. -1984. -V.27. -P.557-564.; [5] *Red Shaw S.* In: The Role of Organic Chemistry in Drug Research, 2nd Ed. / Eds. C.R. Ganellin, S.M. Roberts: London: Academic Press. -1993. -P.163-185.; [6] *Holtje H.D., Sippl W., Rognan D., Folkers G.* Molecular Modelling. Basic Principles and Applications, 2nd Ed. Weinheim: Wiley-VCH, 2003.; [7] *Krocmer R.T.* Curt. Protein Pept. Sci.,8, 312, 2007.; [8] *Pegg S. C., Haresco J.J., Kuntz I.D.* J. Comput. Aided. Mol. Des., 15, 911, 2001.; [9] *Bleicher K.H., Bohm H.J., Muller K., Alanine F.I.* Ibid., 2003. -V.2. -P.369-378.; [10] *Раевский О.А.* Российский химический журнал, 2006. -№2. -С.97-108.; [11] *Раевский О.А.* Успехи химии, 1999. -Т.68. -Вып.6. -С.555.; [12] *Testa B., Rier L.B.* Med. Res. Rev., 1991. -V.11. -P.35.; [13] *Гальберштам Н.М., Баскин И.И., Палюлин В.А., Зефирова Н.С.* Нейронные сети как метод поиска зависимостей структура – свойство органических соединений // Успехи химии. -2003. -№72(7). -С.706-727. [14] *Tarakanov A.O.* Formal peptide as a basic of agent of immune networks: from natural prototype to mathematical theory and applications // Proc. of the I Int. workshop of central and eastern Europe on Multi-Agent Systems.-1999.; [15] *Самигулина Г.А.* Разработка интеллектуальных экспертных систем управления на основе технологии искусственных иммунных систем. –Алматы: ИПИУ МОН РК, 2008. -137с.; [16] *Самигулина Г.А.* Моделирование оптимальной структуры иммунной сети в интеллектуальных экспертных системах поддержки принятия решений //Информационные технологии моделирования и управления. –Воронеж, 2008. – Вып.4(47). –С.313-317.; [17] *Самигулина Г.А., Чебейко С.В.* Процедура оценки энергетической погрешности Искусственной Иммунной Системы прогнозирования чумы на основе гомологов //Вестник КазНУ им. Аль-Фараби. -Алматы. -2003. -№3 (38). -С. 61-66.; [18] *Самигулина Г.А., Самигулина З.И.* Разработка программного обеспечения для интеллектуальной системы прогнозирования на основе Искусственных Иммунных Систем //Труды V Всесибирского конгресса женщин-математиков. – Красноярск, 2008, http://www.sfu-kras.ru/files/spis_t.doc; [19] *Самигулина Г.А., Самигулина З.И.* Разработка пакета прикладных программ для обработки, анализа и прогноза данных Искусственными Иммунными Сетями //Вестник Национальной Инженерной Академии Республики Казахстан. –Алматы, 2008. -№ 1(27). –С.39-44.; [20] *Самигулина Г.А.* Предварительная обработка данных AIS (программа для ЭВМ). Свидетельство о государственной регистрации объекта интеллектуальной собственности №286. Зарегистрировано в Комитете по правам интеллектуальной собственности Министерства юстиции Республики Казахстан. –Астана, 2008.; [21] *Самигулина Г.А., Самигулина З.И.* Оценка энергетических погрешностей искусственной иммунной системы по гомологам (программа для ЭВМ). Свидетельство о государственной регистрации объекта интеллектуальной собственности № 396. Зарегистрировано в Комитете по правам интеллектуальной собственности Министерства юстиции Республики Казахстан. –Астана, 2008.

Принято в печать 26.01.10

УДК 004.89:004.4

КОМПЬЮТЕРНЫЙ МОЛЕКУЛЯРНЫЙ ДИЗАЙН ЛЕКАРСТВЕННЫХ СРЕДСТВ НА ОСНОВЕ ИСКУССТВЕННЫХ ИММУННЫХ СИСТЕМ

Г.А.Самигулина., С.В.Чебейко

Институт проблем информатики и управления МОН РК

050010, г.Алматы, ул.Пушкина, 125

chebeiko@rambler.ru, тел.272-46-17, факс 272-37-11

COMPUTER MOLECULAR DESIGN MEDICAL DRUGS ON THE BASIS OF ARTIFICIAL IMMUNE SYSTEMS

Galina.A.Samigulina, Svetlana.V.Chebeiko
Institute for Problems of Informatics and Control MES RK,
050010, Almaty, str.Pushkina, 125
chebeiko@rambler.ru, tel.272-46-17, факс 272-37-11

It is developed immune nets the approach to modeling dependences «structure – property» of medical drags. The offered intellectual technology allows to reduce errors of power estimations and to raise reliability of the forecast of dependence «structure – property» of chemical compounds.

Жасанды иммундық жүйелер негізінде дәрі-дәрмектердің компьютерлік молекулярлық дизайны
Г.А. Самигулина, С.В. Чебейко
Информатика және басқару проблемалары институты,
050010, Алматы қ., Пушкин қ , 125
chebeiko@rambler.ru, тел.272-46-17, факс 272-37-11

Дәрілік препараттардың «құрылым-қасиет» тәуелділіктерін моделдеуге иммунды желілік тәсіл жасалған. Ұсынылған зияткерлік технология энергетикалық бағалардың қателіктерін азайтуға және химиялық қосылыстардың «құрылым-қасиет» тәуелділігін болжаудың анықтығын арттыруға мүмкіндік береді.