

УДК 544.165:519.688

ИСПОЛЬЗОВАНИЕ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ПРОГНОЗИРОВАНИЯ СПОСОБНОСТИ НОВЫХ ХИМИЧЕСКИХ СОЕДИНЕНИЙ ИНДУЦИРОВАТЬ ПОЯВЛЕНИЕ ИММУНОСТИМУЛИРУЮЩЕЙ АКТИВНОСТИ В СЫВОРОТКЕ КРОВИ МЫШЕЙ

В.В.Ковалишин, В.В.Холодович, Л.А.Метелица, Л.Л.Чарочкина, С.Е.Могилевич

Институт биоорганической химии и нефтехимии НАН Украины,
02660, г. Киев, ул. Мурманская, 1. E-mail: users@bpci.kiev.ua

Ключевые слова: структура — активность; искусственные нейронные сети; алгоритм пространственного обучения

Изучена возможность применения искусственных нейронных сетей для анализа зависимостей структура — активность в отсутствии обучающей выборки. Показано, что предложенный подход позволяет выявить соединения, обладающие новой, ранее неизвестной активностью, и может использоваться для расчетного скрининга при создании новых лекарственных средств.

THE APPLICATION OF ARTIFICIAL NEURONAL NETWORKS FOR PREDICTION OF THE PROPERTY OF NEW CHEMICAL COMPOUNDS TO INDUCE THE IMMUNOSTIMULATING ACTIVITY IN MICE BLOOD SERUM

V.V.Kovalishin, V.V.Kholodovich, L.O.Metelitsa, L.L.Charochkina, S.Ye.Mogilevich

The possibility of using artificial neuronal networks for the analysis of the structure — activity relationships in the absence of the training set of compounds has been studied. The approach proposed has been shown to allow revealing compounds, which possess new previously unknown activity and it can be used for computational screening while creating new drugs.

ВИКОРИСТАННЯ ШТУЧНИХ НЕЙРОННИХ МЕРЕЖ ДЛЯ ПРОГНОЗУВАННЯ ЗДАТНОСТІ НОВИХ ХІМІЧНИХ СПОЛУК ІНДУЦІЮВАТИ ПОЯВУ ІМУНОСТИМУЛЮЮЧОЇ АКТИВНОСТІ В СИРОВАТЦІ КРОВІ МИШЕЙ

В.В.Ковалішин, В.В.Холодович, Л.О.Метелиця, Л.Л.Чарочкіна, С.Є.Могилевич

Вивчена можливість застосування методу штучних нейронних мереж для аналізу залежностей структура — активність за відсутності навчальної вибірки. Показано, що запропонований підхід дозволяє ефективно виявляти сполуки, які мають нову, раніше невідому активність, і може бути застосований для розрахункового скринінгу при створенні нових лікарських засобів.

Ранее было установлено, что на введение некоторых высокотоксичных химических соединений животный организм отвечает выработкой эндогенных факторов, которые при их перенесении в другой организм оказывают иммуностимулирующее действие, противоположное по направленности влиянию самих соединений [1]. Эти вещества существенно угнетают иммунный ответ на уровне организма. В то же время сыворотка или экстракты определенных тканей животных, обработанных высокими дозами указанных соединений, вызывают при введении в организм других животных резкую стимуляцию иммунитета, кро-вообразования и митотической активности.

Высокая токсичность соединений, обуславливающих появление в сыворотке иммуностимулирующих факторов, существенно затрудняет практическое использование таких сывороток. Поэто-

му нами была изучена возможность использования для их получения малотоксичных соединений. Определенная трудность заключалась в том, что предложенная нами схема (использование в качестве иммуностимулирующего препарата сыворотки животных, обработанных каким-либо веществом) практически не описана. То есть в литературе отсутствуют данные по аналогичной активности других соединений, что делает невозможным составление обучающей выборки и создание экспертной системы для поиска взаимосвязей “структура — активность” на основе стандартных подходов, например, таких как метод частичных наименьших квадратов (ЧНК, Partial Least Squares, PLS) [2], генетических алгоритмов [3] и др. Поэтому на первом этапе исследований для анализа массива химических соединений был разработан оригинальный подход на основе мето-

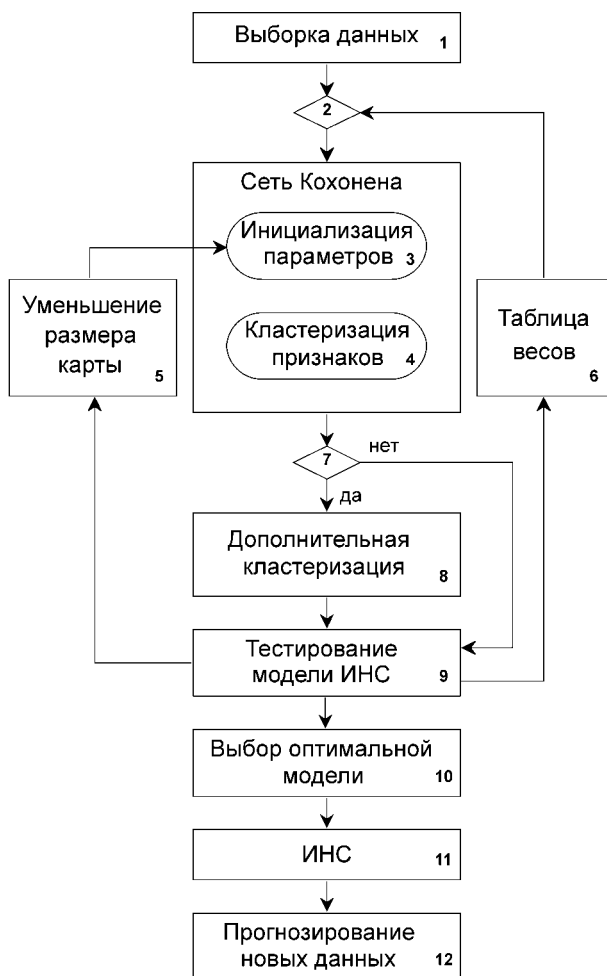


Рис. Общая схема анализа данных с помощью алгоритма пространственного обучения.
 1 - выборка исходных экспериментальных данных;
 3, 4 - сеть Кохонена; 5 - блок контроля и изменения карт Кохонена; 6 - таблица весовых коэффициентов;
 7 - блок контроля кластеров на разрывы в геометрическом пространстве; 8 - блок дополнительной кластеризации;
 9 - блок тестирования модели (кластерного разбиения);
 10 - блок выбора наиболее оптимальной модели;
 11 - блок расчета на основе ИНС; 12 - блок прогноза новых соединений.

да искусственных нейронных сетей (ИНС) [4]. ИНС очень хорошо приспособлены для проведения подобных исследований, так как они позволяют выявить сходство между разнородными группами данных. После того, как были установлены указанные соотношения, с их помощью можно предсказывать свойства объектов, не входивших в исходную группу данных.

Для расчета непрерывных признаков использовался метод CoMFA [2]. Авторы метода исходили из принципа, что биологическая активность зависит от конформации молекулы и что лиганд-рецепторное узнавание и взаимодействие, как правило, происходит без разрыва ковалентных связей. Основное допущение метода состоит в том, что молекулярная активность чувствительна к локализованным в пространстве отличиям в интенсивностях молекулярных полей. Как правило, рассчитываются два типа полей: электростатическое

(Кулоновское) и стерическое (потенциал Леннарда-Джонса) [5]. Интенсивность обеих типов полей измеряется в узлах трехмерной решетки как энергия взаимодействия молекулы с атомом-зондом, обладающим определенным зарядом и стерическими характеристиками.

В настоящей работе мы использовали для анализа алгоритм пространственного обучения ИНС [4, 6], который является комбинацией метода ИНС с обратным распространением ошибки [7] и сетей Кохонена [8]. Основная идея метода состоит в том, что вначале определяются кластерные зоны в пространстве вокруг молекул, затем признаки, из которых эти зоны состоят, обобщаются и используются для обучения ИНС и прогнозирования активности новых соединений [6]. Для оценки качества моделей, найденных с помощью данного подхода, была проведена выборочная экспериментальная проверка правильности полученного прогноза активности.

Выборка данных. На первом этапе исследования была создана выборка из 24 соединений для последующего анализа методами компьютерной химии и искусственных нейронных сетей. Молекулярные структуры соединений представлены в табл. 1. 2,2'-Дихлорэтил сульфоксид и 2,2'-дихлорэтил сульфон были предоставлены проф. А.И.Луйком, 2,6-диметилпиридин N-оксид — к.х.н. С.П.Пономаренко, N-(1,1-диоксопиридин-3-ил) дитиокарбамат калия — проф. Л.Н.Шкарапутой (ИБОНХ НАН Украины). Все остальные соединения — Sigma (США).

Молекулярное моделирование. Ко всему набору данных была применена стандартная CoMFA-процедура [2, 5]. Каждый из 24 лигандов был смоделирован с помощью программы Sybyl 6.8 на рабочей станции Silicon Graphics R10000 Octane. Начальные конформации лигандов были оптимизированы на основе алгоритма молекулярной механики с помощью силового поля Tripos [9]. Частичные атомные заряды и необходимые вычисления электростатических взаимодействий рассчитывались на основе формул Гастейгера-Марсили. Модели выравнивались согласно основности структурных элементов, т.е. ориентировались в пространстве таким образом, чтобы достичь максимального совмещения молекулярных единиц: гетероатомов (S, N, O) и структурных элементов (ароматических колец и алифатических цепочек). После такого выравнивания каждая молекула с фиксированной структурой помещалась в пространственный параллелепипед, составленный из кубических ячеек размером 2 \AA . Все вычисления проводили в вакууме. Чтобы уменьшить влияние удаленных полей на QSAR модель, при вычислении электростатических взаимодействий величина энергии уменьшалась согласно квадрату расстояния между атомами. Между каждым атомом молекулы и атомом углерода ("щупом") с зарядом +1 рассчитывалась электростатическая и стерическая энергия взаимодействия в каждом узле решетки. Если энергия стерического или электростатического взаимодействия превышала ± 30 ккал/моль, она была ограничена этим пороговым значением.

Таблица 1

Суммарный статистический анализ выборки данных

№ п/п	Название	Формула	Сеть Кохонена	АПО ИНС*	
			тип активности	вер., %	
1	Этил сульфид	$\text{CH}_3\text{CH}_2\text{SCH}_2\text{CH}_3$	акт.	Акт.	99
2	2,2'-Дихлорэтилсульфид	$\text{ClCH}_2\text{CH}_2\text{SCH}_2\text{CH}_2\text{Cl}$	акт.	Акт.	99,5
3	2,2'-Тиодизэтанол	$\text{HOCH}_2\text{CH}_2\text{SCH}_2\text{CH}_2\text{OH}$	акт.	Акт.	97,5
4	3,3'-Дихлорпропилсульфид	$\text{ClCH}_2\text{CH}_2\text{CH}_2\text{SCH}_2\text{CH}_2\text{CH}_2\text{Cl}$	акт.	Акт.	99
5	2,2'-Дихлорэтилсульфон	$\text{ClCH}_2\text{CH}_2\text{S}(=\text{O})\text{CH}_2\text{CH}_2\text{Cl}$	неакт.	Неакт.	89
6	Этил(2-хлорэтил)сульфон	$\text{CH}_3\text{CH}_2\text{S}(=\text{O})\text{CH}_2\text{CH}_2\text{Cl}$	неакт.	Неакт.	88,5
7	Метил-р-толилсульфон		неакт.	Неакт.	81,5
8	Дифенилсульфон		неакт.	Акт.	79
9	Метил-р-толилсульфоксид		акт.	Акт.	97,5
10	Бензил-р-толилсульфоксид		акт.	Акт.	99
11	Диметилсульфоксид	$\text{CH}_3\text{S}(=\text{O})\text{CH}_3$	акт.	Акт.	97,5
12	Дифенилсульфоксид		акт.	Акт.	97,5
13	2,2'-Дихлорэтилсульфоксид	$\text{ClCH}_2\text{CH}_2\text{S}(=\text{O})\text{CH}_2\text{CH}_2\text{Cl}$	акт.	акт.	99
14	Трис(2-хлорэтил)амин		акт.	акт.	98,5
15	Бис(2-хлорэтил)метиламин		акт.	акт.	100
16	Триметиламин-N-оксид	$(\text{CH}_3)_3\text{N} \rightarrow \text{O}$	акт.	акт.	98
17	Пиридин-N-оксид		акт.	акт.	98
18	2,6-Диметилпиридин-N-оксид		акт.	акт.	100
19	Нитрометан	CH_3NO_2	акт.	акт.	100
20	Нитробензол		акт.	акт.	98
21	Ацетонитрил	$\text{CH}_3\text{C} \equiv \text{N}$	акт.	акт.	97,5
22	Бензонитрил		акт.	акт.	99,5
23	N-(1,1-диоксоетиолан-3-ил) дитиокарбамат калия		акт.	неакт.	87
24	Бутадиенсульфон		неакт.	акт.	100

* АПО ИНС - Алгоритм пространственного обучения искусственных нейронных сетей

Статистический анализ соединений. Алгоритм пространственного обучения (АПО) реализует циклическую итерационную процедуру, комбинирующую поочередное применение сети Кохонена и искусственных нейронных сетей. Детально АПО ИНС был описан в работах [4, 6], поэтому мы лишь кратко остановимся на особенностях метода.

Общая блок-схема алгоритма показана на рисунке. Процесс анализа данных условно можно разделить на двенадцать шагов. Входные данные, сформированные на шаге 1, поступают на входы сети Кохонена (блоки 3, 4). В результате обучения сети на выходе формируется карта распределения признаков на кластеры. Кластеры проверяются на наличие разрывов в геометрическом пространстве (блок 7), и при необходимости осуществляется дополнительная кластеризация. После сжатия данных путем замены кластеров их среднеарифметическим значением полученное распределение данных тестируется с помощью ИНС. При этом вес нейронов входного уровня записывается в таблицу, и его показатели используются на следующих циклах обучения вместо входных данных (блок 6). Число рядов и колонок в карте Кохонена уменьшается на единицу (блок 5) и весь процесс сжатия данных повторяется. Из всех моделей кластерного распределения данных выбирается модель с минимальной ошибкой прогноза (блок 10), которая в дальнейшем используется для обучения ИНС и прогнозирования активности новых соединений (блоки 11, 12).

Для тестирования сжатых данных использовались ИНС, состоящие из 3 слоев. Детально метод ИНС описан в работах [4, 7]. Количество нейронов во входном слое соответствовало количеству кластеров, полученных на выходе сети Кохонена. На скрытом слое использовали пять нейронов. Третий слой состоял из двух нейронов и соответствовал количеству прогнозируемых классов активности.

Во избежание проблем “переобучения” ИНС использовался метод “раннего останова” ансамбля нейронных сетей [7], который позволяет существенно улучшить прогнозирующую способность сетей. Каждый ансамбль ИНС формировали из 200 сетей. Усредненное по всем сетям значение активности использовали для вычисления статистических коэффициентов. Качество полученных моделей оценивалось с помощью метода скользящего контроля: каждая молекула последовательно удалялась из обучающей выборки и затем использовалась для оценки качества модели.

Для определения кластеров, наиболее полно отображающих проявление биологической активности исследуемых соединений, были использованы методы отбора признаков, известные в литературе как “pruning methods” [10]. Согласно этим методам чувствительность нейронов рассчитывают, исходя из амплитуды их весов или изменения ошибки нейронных сетей без учета весов. Адекватность данного подхода была доказана на ряде теоретических задач распознавания образов [11], что позволило нам использовать эти методы

для поиска зон, отвечающих за активность исследуемых соединений.

В результате расчетов на основе стандартного метода CoMFA были сформированы два набора данных, состоящие из 560 признаков. Каждый признак соответствовал энергии стерического или электростатического взаимодействия с атомом зондом в данном узле решетки, измеренной в ккал/моль. Далее объединенная выборка данных была проанализирована с помощью сети Кохонена. В процессе анализа использовались карты Кохонена различной величины, что позволило в итоге разделить все соединения только на два класса, используя информацию о пространственном строении соединений. Было принято, что первый класс содержит активные, а второй — неактивные соединения, поскольку в первый класс вошло соединение 2, активность которого была установлена нами ранее экспериментально [1]. В результате расчетов соединения под номером 5-8, 24 были отнесены ко второму классу, тогда как остальные соединения были отнесены к первому классу активных соединений (табл. 1).

На втором этапе исследований проводился общий анализ соединений CoMFA-признаков с помощью АПО ИНС. В процессе анализа исходное пространство признаков было разделено на 12 кластеров по стерическим и на 5 кластеров по электростатическим признакам. Применение методов оптимизации позволило удалить малоинформативные признаки и отобрать всего 5 кластеров из 12 для стерических признаков и 2 кластера из 5 для электростатических признаков.

Активность каждого соединения прогнозировалась с помощью метода скользящего контроля, т.е. каждая молекула последовательно удалялась из обучающей выборки в набор тестирования и использовалась для оценки качества модели. Для получения более достоверной модели использовался статистический прогноз на основе ансамбля из 200 сетей. Результаты анализа для каждого соединения приведены в табл. 1. Эти данные свидетельствуют, что прогнозы, полученные на основе АПО ИНС только для соединений под номером 8, 23 и 24, отличались от прогноза, полученного сетью Кохонена.

Для экспериментальной проверки прогноза было отобрано 10 соединений из обеих групп. Используемые дозы соединений (табл. 2) были максимальными, при которых животные оставались живыми через сутки после введения. Активность соединений оценивалась как способность сыворток крыс, обработанных изучаемыми препаратами, влиять на образование у мышей антител к эритроцитам барана. Как видно из табл. 2, экспериментальные результаты подтвердили данные, предварительно полученные на основе компьютерной модели АПО ИНС. Некоторое отличие результатов прогноза на основе сети Кохонена от прогноза, полученного АПО ИНС, связаны с тем, что они имеют предварительный характер (т.е. позволяют приблизительно разделить соединения

Таблица 2

Влияние сыворотки крыс, обработанных тестируемыми соединениями, на титры антител к ЭБ в крови мышей ($M \pm m$, $n = 3$)

№ п/п	Соединение	Доза, мг на 1 кг массы тела	Титры антител, $\log_2 T$
1	Интактная сыворотка	-	3,3±0,2
2	2,2'-Дихлорэтилсульфон	30	1,4±0,1*
3	Дифенилсульфон	700	4,2±0,4
4	Метил-р-толилсульфоксид	1000	7,2±0,6*
5	Диметилсульфоксид	7000	6,0±0,3*
6	Дифенилсульфоксид	1000	6,0±0,5*
7	2,2'-Дихлорэтилсульфоксид	200	7,0±0,2*
8	2,6-Диметилпиридин-N-оксид	750	5,6±0,3*
9	N-(1,1-диоксотилолан-3-ил) дитиокарбамат калия	500	2,5±0,1*
10	Бутадиенсульфон	2000	6,2±0,3*

* - Показатель достоверно ($p < 0,05$) отличается от такового для интактной сыворотки

на два класса) и всегда требуют дальнейшей проверки АПО ИНС.

Экспериментальная часть

В работе были использованы крысы линии Wistar и белые беспородные мыши. Для получения специфических сывороток крысам подкожно вводили изученные соединения: водорастворимые — в физрастворе, водонерастворимые — в виде тонкой суспензии или эмульсии в подсолнечном масле [12]. Через 24 часа крыс декапитировали и получали сыворотку крови [13].

Реакцию прямой гемагглютинации (РПГА) с эритроцитами барана (ЭБ) проводили для тести-

рования способности специфических сывороток крови стимулировать антителогенез у экспериментальных животных [14, 15]. С этой целью иммунизировали белых беспородных мышей ЭБ и специфическими сыворотками. Каждой мыши внутрибрюшинно вводили 0,5 мл суспензии ЭБ (концентрация 10%) и через 30 секунд — внутрибрюшинно 0,2 мл сыворотки. Через 7 суток после иммунизации получали сыворотку крови. Для исключения неспецифического иммунного ответа к ЭБ и предотвращения иммунного гемолиза все исследуемые сыворотки прогревали на протяжении 30 мин при 56°C для инактивации комплемента.

РПГА проводили в лунках пластмассовых панелей для микротитрования. Все специфические сыворотки крови разводили последовательно, каждую в ряду панели (6 параллелей) в 2, 4, 8 и т.д. раз в сенсibiliзирующем растворе альбумина сыворотки крови человека (концентрация 0,2%, раствор А). Потом во все лунки планшета вносили равный объем суспензии ЭБ (концентрация 0,2%). Контролем реакции служил тест на полноту оседания эритроцитов в растворе А.

Реакция завершалась через 3-4 часа. Результаты реакции представляли в виде $\log_2 T$ (T-титр последнего разведения сыворотки, при котором еще происходит агглютинация эритроцитов).

Выводы

Предложена экспертная система, являющаяся достаточно эффективным средством для отбора соединений, обладающих новой ранее неизвестной биологической активностью. Хорошее совпадение расчетных и экспериментальных данных позволяет использовать эту систему для проведения расчетного скрининга при создании новых лекарственных средств.

Литература

1. Пат. 51037 (2002). Україна.
2. In 3D QSAR in Drug Design / Ed. H.Kubini. — Leiden: ESCOM, 1993.
3. So S.S., Karplus M. // J. Med. Chem. — 1997. — Vol. 26, №40. — P. 4347-4359.
4. Tetko I.V., Kovalishyn V.V., Livingstone D.J. // J. Med. Chem. — 2001. — Vol. 44, №15. — P. 2411-2420.
5. Cramer R.D., Patterson D.E., Bunce J.D. // J. Am. Chem. Soc. — 1988. — Vol. 110, №18. — P. 5959-5967.
6. Ковалишин В.В., Тетко И.В., Луйк А.И. и др. // Биоорг. химия. — 2001. — Т. 27, №4. — С. 303-313.
7. Tetko I.V., Livingstone D.J., Luik A.I. // J. Chem. Inf. Comput. Sci. — 1995. — Vol. 35, №5. — P. 826-833.
8. Kohonen T. Self-organisation Maps. — Berlin: Springer-Verlag, 2001.
9. Clarc M., Cramer R.D. III and Opendbosch N.V. // J. Comput. Chem. — 1989. — Vol. 10, №8. — P. 982-1012.
10. Tetko I.V., Villa A.E.P., Livingstone D.J. // J. Chem. Inf. Comput. Sci. — 1996. — Vol. 36, №4. — P. 794-803.
11. Kovalishyn V.V., Tetko I.V., Luik A.I. et al. // J. Chem. Inf. Comput. Sci. — 1998. — Vol. 38, №4. — P. 651-659.
12. Фонталин Л.Н., Певницкий Л.А., Соловьев В.В. и др. // Вестник АМН СССР. — 1970. — №7. — С. 75-86.
13. Handbook of Experimental Immunology / Ed. by D.M.Weir. — Blackwell Scientific Publications: Oxford. — Edinburgh, 1967.
14. Зигль Э. Иммунологические методы. — М.: Мир, 1979. — С. 108-110.
15. Banks T., Boursnell J., Francis G. et al. // Biochem. J. — 1946. — Vol. 40, №5-6. — P. 765-771.

Надійшла до редакції 30.07.2007 р.