

УДК: 539.541 + 539.543

ПРИМЕНЕНИЕ КОМПЬЮТЕРНО-ХРОМАТОГРАФИЧЕСКОЙ СИСТЕМЫ ПРИ РЕШЕНИИ ПРОБЛЕМЫ "СТРУКТУРА – СВОЙСТВО"

© 2001 В.П. Гарькин,¹ С.В. Курбатова,² Н.В. Соловова,³ Е.Е.Финкельштейн⁴

В работе рассматривается применение компьютерно-хроматографической системы в решении проблемы "структура – свойство". Архитектура системы включает банки физико-химических свойств, данные о молекулярной и электронной структуре веществ, банки корреляций.

Создание материалов, обладающих заданными физическими, химическими или биологическими свойствами, является одной из наиболее важных проблем в химической, нефтехимической, фармацевтической и других отраслях науки и промышленности. В связи с этим возрастает особый интерес к проблеме исследования количественных соотношений "структура – свойство".

Достигнутый к настоящему времени уровень знаний о веществе, его строении и композициях, практические потребности ставят перед химиками задачу поиска более эффективного проектирования и создания атомных и молекулярных конструкций, гарантирующих наличие определенных свойств у конечного продукта. Эта проблема включает ряд фундаментальных задач, связанных, прежде всего, с зависимостью свойств веществ от пространственного расположения атомов в их молекулах, а также с исследованием потенциальных (или энергетических) функций систем и пр. Базисом для решения данных проблем должно явиться использование представлений о химической связи для конструирования новых, еще не существующих соединений [1]. Пример различных свойств алмаза и графита свидетельствует о том, что не столько атомы, входящие в состав молекулы, определяют ее свойства, сколько способ связывания их между собой. Именно поэтому возможные подходы к решению проблемы конструирования новых структур с заданными свойствами должны опираться в первую очередь на представления о химической связи и движении электронов внутри молекулы.

Особую актуальность эта проблема приобретает при исследовании веществ с потенциальной биологической активностью и лекарственных препаратов [2].

¹Гарькин Виталий Петрович, кафедра органической химии Самарского государственного университета, 443011, г.Самара, ул. Акад. Павлова, 1; garkin@ssu.samara.ru

²Курбатова Светлана Викторовна, кафедра общей химии и хроматографии Самарского государственного университета, 443011, г.Самара, ул. Акад. Павлова, 1; curbatsv@ssu.samara.ru

³Соловова Наталья Валентиновна, кафедра общей химии и хроматографии Самарского государственного университета, 443011, г.Самара, ул. Акад. Павлова, 1; solovova@ssu.samara.ru

⁴Финкельштейн Елена Евгеньевна, кафедра общей химии и хроматографии Самарского государственного университета, 443011, г.Самара, ул. Акад. Павлова, 1

В этом направлении существует достаточно много сложностей, связанных прежде всего с многообразием веществ, используемых в фармакологии, а также с разнообразием форм и организации живых организмов, на которых осуществляются испытания новых препаратов. Следует отметить, что интерес к этой проблеме неуклонно возрастает. В последнее время значительно увеличилось количество публикаций, посвященных синтезу и исследованию веществ с заранее заданными свойствами и особенностям веществ с потенциальной биологической активностью [3]. Решению проблемы исследования биологически активных веществ посвящена деятельность международного общества по анализу количественных соотношений "структура – активность" (International QSAR and Modeling Society), созданного в 1989 г. на Гордонской конференции в Тилтоне (Нью-Хемпшир, США), и начавшей функционировать в 1996 году Российской секции этого общества [4]. Все изложенное свидетельствует о том, что химическая наука в настоящее время стоит перед необходимостью выработки четких принципов конструирования веществ и, в частности, биологически активных, с заданными свойствами.

До настоящего времени попытки решить проблему "структура – свойство" и соответственно "структура – активность" предпринимались представителями различных областей науки — органической, квантовой, биологической, медицинской химии, фармакологии и т.д. По словам доктора Ханча — основоположника подхода QSAR к этой проблеме, "большинство исследователей смотрят на эту проблему изнутри своей собственной специальности". Объединение усилий исследователей, работающих в данной области, возможно, на наш взгляд, только при использовании системного подхода, сущность которого заключается в создании единой модели описания качественного и количественного аспектов физико-химического поведения и биологической активности соединений, опирающихся на данные об их структуре и физико-химических свойствах. В отношении биологической активности системный подход предполагает использование на разных этапах исследования полной информации об изучаемом соединении: о наличии в молекуле определенных активных с биологической точки зрения структур (или подструктур), вносящих вклад в общую активность молекулы; о способе и формах поступления в организм и приемах регистрации получаемого биологического отклика; о физико-химическом поведении и квантово-химическом описании строения молекул.

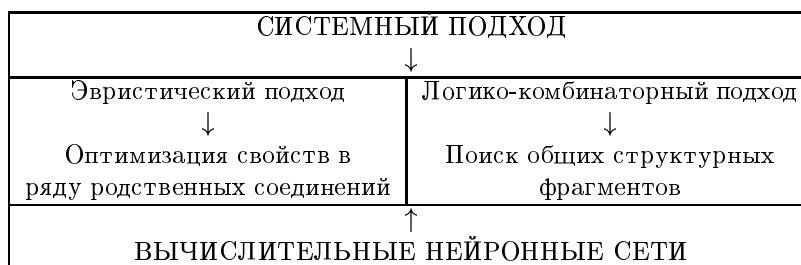
В литературе уже высказывалось мнение о том, что физико-химической основой такого объединения может служить хроматография. Это предположение не лишено здравого смысла, в первую очередь, потому, что хроматографические характеристики, достаточно легко определяемые экспериментально, хорошо коррелируют практически со всеми параметрами, рассматриваемыми в различных подходах и методах. Использование различных вариантов хроматографии, различных детекторов и колонок позволяет выявлять типы ММВ, определяемые различиями в структуре соединений, моделировать биохроматографические процессы, происходящие при введении физиологически активных веществ в живой организм. По мнению отдельных авторов применение хроматографических характеристик для описания биологического действия органических соединений оправдано тем, что в хроматографической колонке так же, как и в биологических системах, реализуется процесс распределения вещества между фазами с различными физико-химическими свойствами. Во многих случаях также предполагается сорбционный механизм действия лекарственного препарата или любого биологически активного вещества, поэтому интересным представляется использование сорбционных характеристик таких веществ, получаемых хроматографическим методом. Поэтому весьма перспективным подходом к реше-

нию проблемы "структура – свойство" является упоминаемый в литературе подход QSRR — Quantitative Structure Retention Relationships.

Очевидно, что для выявления возможных закономерностей необходимо иметь соответствующую базу данных по всем указанным характеристикам соединений с потенциальной биологической активностью, при этом оптимальными программными средствами для обработки информации с целью установления соответствующих корреляций и прогнозирования свойств и активности могут стать широко развивающиеся в настоящее время вычислительные нейронные сети (табл. 1) [5], [6].

Таблица 1

Системный подход к решению проблемы исследования количественных соотношений "структура – свойство" (QSPR)



С точки зрения реализации системного подхода предполагается целесообразным использование математических и программных методов моделирования молекулярных структур биологически активных веществ.

Компьютерная реализация системного подхода, в свою очередь, включает решение ряда задач, связанных с различными возможностями представления структуры в молекулярном виде, дескрипторном или с помощью корреляций.

Таблица 2

Основные методы оптимизации свойств в ряду родственных соединений

ВИД АНАЛИЗА	ПРИНЦИП ОБРАБОТКИ ИНФОРМАЦИИ
Многомерный статистический	Расчет статистических характеристик распределения случайных величин
Многомерное шкалирование	Сокращение размерности многомерного пространства с целью исключения малоинформативных данных
Корреляционный	Выявление связей между переменными
Факторный	Сжатие информации путем введения новых переменных, являющихся линейной комбинацией исходных
Регрессионный и дисперсионный	Построение функциональной зависимости между независимыми переменными и их оптимизация

При этом возможно использование либо оптимизации свойств в ряду родственных соединений (табл. 2), либо поиск способов выявления общих элементов структуры и классификации соединений (табл. 3).

Для описания структуры соединений используют различные методы, позволяющие получать физико-химические и структурные характеристики (табл. 4).

Получение разнообразной информации о физико-химических, электронных, структурных параметрах возможно с использованием компьютерных программ (в част-

Таблица 3

Основные способы выявления общих элементов структуры и классификации соединений

ВИД АНАЛИЗА	ПРИНЦИП ОБРАБОТКИ ИНФОРМАЦИИ
Дискриминантный	Построение разделяющих гиперповерхностей по обучающей выборке
Логико-структурный	Использование законов индуктивной логики в дискриминантном анализе
Кластерный	Классификация объектов по признакам сходства или различия с помощью оптимизации целевой функции
Моделирование 3D структуры и свойств молекул	Расчет и построение 3D структуры молекул методами квантовой химии и молекулярной механики

Таблица 4

Методы описания структуры соединений

МЕТОДЫ	ИЗУЧАЕМЫЕ ХАРАКТЕРИСТИКИ
Система топологических структурных дескрипторов	Индексы Винера, Рэндиша, Балабана, Хосойи и др.
Квантовохимические (полуэмпирические) методы расчета — PMX, CNDO, INDO, AM1, PM3 и ab initio), методы молекулярной механики (MM+)	Энергия образования, длины связей, валентные углы, дипольные моменты, распределение электростатического потенциала, спиновой плотности и пр.
Физико-химические	Стерические, гидрофобные, сорбционные, энергетические характеристики

ности, HYPERCHEM, GAUSSIAN, MOPAC и др.). Нами в работе использован комплекс программ Hyperchem версии 4.5, позволяющий осуществлять расчеты полуэмпирическими методами PMX, CNDO, AM1, PM3 и пр., методами молекулярной механики MM+ и др.

Очевидно, что для выявления возможных закономерностей необходимо иметь соответствующую базу данных по всем указанным характеристикам соединений с потенциальной биологической активностью, при этом оптимальными программными средствами для обработки информации с целью установления соответствующих корреляций и прогнозирования свойств и активности могут стать широко развивающиеся в настоящее время вычислительные нейронные сети (Neural Networks), или искусственные нейронные сети (ИНС), разработка которых явилась результатом коренных изменений в понимании принципов организации вычислительного процесса и подходов к решению прикладных задач.

ВНС могут быть использованы в качестве основного подхода к решению проблемы прогнозирования биологической активности (см. рис. 1) [7]. ВНС основаны на использовании структурных характеристик молекулярных субъединиц и генетического алгоритма, основанного на создании данной молекулярной структуры с использованием заданных макросвойств. Было показано, что использование ВНС для характеристики нелинейных корреляций "структура – свойство" дает значительно лучшие результаты по сравнению с другими подходами.



Рис. 1. Схема использования искусственной нейронной сети для прогнозирования свойств химических соединений

Вычислительные нейронные сети развиваются как одно из направлений в области искусственного интеллекта и представляют собой метод компьютерной обработки экспериментальных данных, основанный на упрощенной математической модели работы клеток мозга. В ВНС заложена способность подходить к обработке информации как к процессу распознавания и классификации образов совершенно произвольной структуры различной сложности при помощи неформализуемых алгоритмов, находимых самими нейронными сетями. Другими словами, ВНС являются наиболее подходящим программным средством для обработки больших массивов данных.

Для разработки алгоритма установления зависимости между структурой молекул и проявляемой ими биологической активностью нами начато формирование автоматизированной базы данных, включающей различные физико-химические, электронные и структурные параметры молекул. Банк данных предназначен для хранения, поиска и выбора информации об органических веществах с потенциальной биологической активностью. Создаваемый банк данных может иметь и самостоятельное значение. Внутри информационной базы возможно построение различных корреляций между величинами хроматографического удерживания веществ с потенциальной биологической активностью, их топологическими характеристиками и некоторыми физико-химическими параметрами. В настоящее время банк содержит систематизированные данные по дескрипторам элементарного уровня (молекулярная масса), топологическим дескрипторам (простые и трехмерные индексы Винера, индексы связанности нескольких порядков), дескрипторам электронной структуры (энергии молекулярных орбиталей, дипольный момент), дескрипторам межмолекулярных взаимодействий (молекулярная рефракция, поляризуемость, липофильность) для двухсот органических соединений различной молекулярной структуры. На основе полученных корреляций возможен прогнозирующий расчет величин хроматографического удерживания, других физико-химических параметров, а также исследование взаимосвязи строения изучаемых органических соединений с их свойствами и биологической активностью.

Таким образом, использование базы данных по физико-химическим свойствам в сочетании с квантово-химическими компьютерными расчетными программами и вычислительными нейронными сетями позволяет создать единую компьютерно-хроматографическую систему и реализовать на практике системный подход к проблеме прогнозирования физико-химических свойств и биологической активности.

Литература

- [1] Раевский А.О., Салегин А.М. Возможности и перспективы конструирования

биологически активных веществ// Успехи химии. 1988. Т.57. № 9. С. 1565–1586.

- [2] Ханч К. Об использовании количественного соотношения "структура – активность" (КССА) при конструировании лекарств// Хим.-фарм. журн. 1980. № 10. С. 15–29.
- [3] Зефиров Н.С., Палюлин В.С., Радченко Е.В. Метод анализа топологии молекулярного поля в исследованиях количественной связи между структурой и активностью органических соединений// ДАН, 1997. Т.352. № 5. 630 с.
- [4] Раевский А.О., Сапегин А.М. Моделирование связи "структура – активность". III. Системный физико-химический подход к конструированию биологически активных веществ// Хим.-фарм. журн. 1990. № 1. С. 43–46.
- [5] Поройков В.В., Раевский А.О. О создании российской секции международного общества по анализу количественных соотношений "структура – активность" и моделированию (QSAR and modeling society)// Хим.-фарм. журн. 1996. № 10. С. 56.
- [6] Баскин И.И., Палюлин В.А., Зефиров Н.С. Вычислительные нейронные сети как альтернатива линейному регрессионному анализу при изучении количественных соотношений "структура – свойство" на примере физико-химических свойств углеводов// ДАН. 1992. Т.333. № 2. С. 189–192.
- [7] Баскин И.И., Палюлин В.А., Зефиров Н.С. Применение искусственных нейронных сетей в химических и биохимических моделях// Вестник Московского Университета. Сер. 2. Химия. 1999. Т.40. № 5. С. 323–326.

COMPUTER-CHROMATOGRAPHIC SYSTEM OF THE QSPR-PROBLEM

© 2001 V.P. Garkin,⁵ S.V. Curbatova,⁶ N.V. Solovova,⁷ E.E. Finkelstein⁸

An application of computer-chromatographic system to the problem solution of quantitative "structure – properties" relationships is investigated. Architecture of this system includes the banks of physico-chemical properties, data set for compounds molecular and electron structure, banks of correlation's.

Поступила в редакцию 14/IV/2001;
в окончательном варианте — 5/VI/2001.

⁵Garkin Vitaliy Petrovich, Dept. of General Chemistry, Samara State University, Samara, 443011, Russia; garkin@ssu.samara.ru

⁶Curbatova Svetlana Victorovna, Dept. of Chemistry and Chromatography, Samara State University, Samara, 443011, Russia; curbatov@ssu.samara.ru

⁷Solovova Nataliya Valentinovna, Dept. of Chemistry and Chromatography, Samara State University, Samara, 443011, Russia; solovova@ssu.samara.ru

⁸Finkelstein Elena Evgenievna, Dept. of Chemistry and Chromatography, Samara State University, Samara, 443011, Russia