

УДК 547.1:541.6

В. Е. Кузьмин^{1,2}, А. В. Ляховский², А. Г. Артеменко²¹Одесский Национальный Университет имени И. И. Мечникова, кафедра органической химии, ул. Дворянская 2, Одесса, 65082, Украина²Физико-химический институт им. А. В. Богатского НАН Украины

Люстдорфская дорога 86, г. Одесса, 65080, Украина,

E-mail: 1dqsar@rambler.ru

ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ ЦИРКУЛЯРНЫХ МОДЕЛЕЙ МОЛЕКУЛ ПРИ РЕШЕНИИ ЗАДАЧ QSAR

Предложен новый метод генерации структурных параметров для решения задач QSAR на основе циркулярных моделей молекул. Метод позволяет формировать структурные параметры адекватно описывающие особенности молекулярной структуры. Продемонстрирована эффективность разработанного подхода на примере QSAR анализа ингибиторов ангиотензинпреобразующего энзима.

Ключевые слова: QSAR, молекулярные дескрипторы, гармонический анализ, циркулярные модели молекул.

В настоящее время исследования в области количественных соотношений между структурой и свойствами органических соединений (QSAR¹/QSPR) являются неотъемлемой частью работ, связанных с поиском новых веществ с комплексом полезных свойств. Это объясняется в первую очередь тем, что выявление количественных зависимостей «структура-свойство» экономит время и ресурсы при конструировании новых высокоэффективных препаратов материалов и реагентов.

Современные QSAR методы различаются в основном способом генерации дескрипторов на основе различных уровней детализации молекулярной структуры (1D–4D) и алгоритмов математического описания связи «структура–активность» [1].

Введение

Ранее для описания и оценки структурного подобия циклических структур был предложен метод, в котором структурные параметры, оценивались на основе гармонического анализа их геометрических характеристик [2]. Впервые аналогичные параметры были предложены Кремером и Поплом, затем усовершенствованы Зефириным и Палюлиным [3]. Указанные параметры достаточно хорошо зарекомендовали себя для описания и анализа подобия/различия циклических структур [2].

В настоящей работе мы попытались расширить возможности таких подходов для описания химических структур произвольной топологии. В этой

¹ QSAR/QSPR — quantitative structure activity (property) relationship, количественная связь структура-активность (свойство)

связи было необходимо разработать процедуру для представления различных молекулярных структур в виде циркулярных (циклических) моделей. Для решения этой задачи необходимо найти однозначный способ циклического обхода атомов исследуемых структур произвольного строения. Основная идея предлагаемого подхода заключается в поиске циклического маршрута минимальной длины для всех вершин молекулярного графа. Фактически, речь идет о решении «задачи коммивояжера» на молекулярном графе. Для ее реализации предлагается использовать модифицированный метод ветвей и границ [4]. Таким образом, для построения циркулярной модели молекулы предлагается следующий алгоритм (см. рис.).

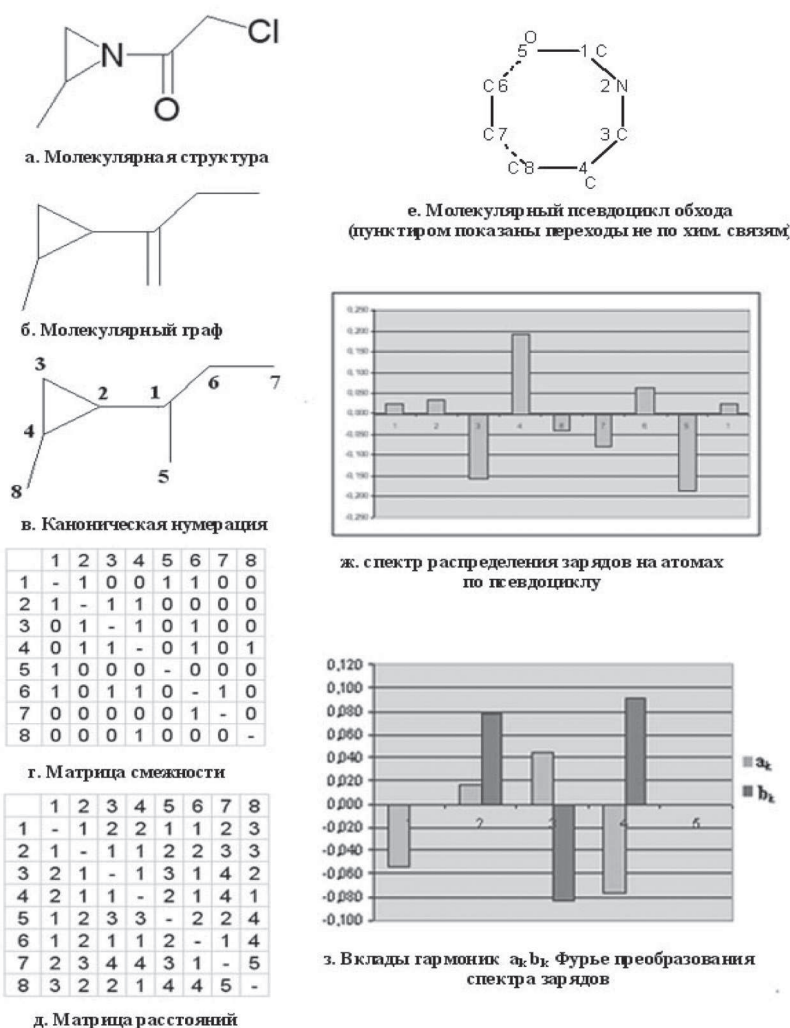


Рис. Схема построения циркулярной модели молекулярной структуры на примере 2-хлор-1-(2-метил-азиридин-1-ил)-этанона

1. Стандартная процедура представления структурной формулы соединения в виде молекулярного графа [5] (см. рис. 1 а и б).

2. Поскольку алгоритм метода ветвей и границ не всегда дает однозначное решение, особенно в случаях наличия вырожденных маршрутов, то для снятия вырождения предлагается использовать каноническую нумерацию атомов, однозначно привязанную к молекулярной структуре (рис. 1 в). Для получения такой нумерации необходим расчет информационных потенциалов на атомах [6].

3. В качестве входной информации метода ветвей и границ используются матрица смежности и матрица расстояний молекулярного графа (см. рис. 1 г, 1 д). Результирующий «псевдоцикл» минимального размера формируется последовательным редуцированием матрицы расстояний [4]. На данном этапе строится циркулярная модель молекулярной структуры (см. рис 1 е), которая определяет пространство дескрипторов для решения задач QSAR.

4. На этом этапе формируются спектры распределения различных свойств атомов по полученному ранее псевдоциклу (см. рис. 1 ж). Используются характеристики атомов, которые могут существенно влиять на проявление изучаемого комплекса свойств веществ (например, заряды на атомах, липофильность и др.)

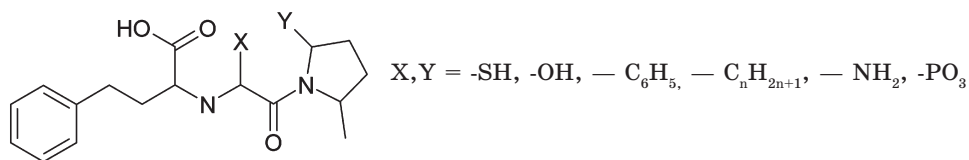
5. Дескрипторами QSAR моделей, являются значения величин гармоник Фурье анализа спектров распределения атомных свойств полученных на этапе 4 (см. рис. 1 з). Гармонический анализ функций распределения структурных параметров по «псевдоциклу» позволяет отразить структурные особенности молекул: от отдельных фрагментов (высоочастотные гармоники) до целой молекулы (низкочастотные гармоники). Для каждой молекулы, представленной рядом молекулярных спектров, разложение выполняется по соотношениям:

$$p(i) = \frac{a_0}{2} + \sum_{k=1}^{M-1} (a_k \cos[2\pi k(i-1)/N] + b_k \sin[2\pi k(i-1)/N]) + \frac{a_{N/2}}{2} \cos[\pi(i-1)],$$

$$\text{где } a_k = \frac{2}{N} \cdot \sum_{i=1}^N P_i \cdot \cos\left(\frac{2\pi \cdot k \cdot (i-1)}{N}\right), \quad b_k = \frac{2}{N} \cdot \sum_{i=1}^N P_i \cdot \sin\left(\frac{2\pi \cdot k \cdot (i-1)}{N}\right)$$

a_k и b_k — коэффициенты, отражающие соответствующие вклады косинусоидальных и синусоидальных гармоник вышеупомянутого спектра распределения свойств; N — общее количество структурных параметров; $M = \text{int}(N-1)/2$ — общее количество гармоник; P_i — значение i -го структурного параметра для молекулы; k — номер гармоники. По существу, a_k и b_k — дескрипторы QSAR моделей.

Для оценки эффективности разработанного подхода и сравнения его с другими методами QSAR нами была решена задача анализа связи между структурой ряда соединений:



и их сродством к ангиотензин-преобразующему энзиму (АПЭ) [7].

Исследовалась способность к ингибированию АПЭ (pIC_{50}) 114 соединений, из которых 76 составляли обучающую выборку, а 38 — тестовую. В качестве инструмента статистического анализа использовался метод частичных наименьших квадратов (PLS). Данная задача является тестовой для анализа эффективности различных QSAR подходов [7]. В работе [7] описаны расчеты, полученные с помощью следующих методов: CoMFA — метод сравнительного анализа молекулярных полей [8], CoMSIA — метод сравнительного анализа индексов подобия [9], EVA анализ собственных значений матриц молекулярного графа [9], HQSAR — гипотетическая решетка активного сайта рецептора [10].

В результате решения данной задачи разработанным методом (Circular) была получена адекватная QSAR модель, статистические характеристики которой приведены в табл. 1.

Таблица 1

Сравнительная характеристика статистических характеристик различных QSAR моделей для АПЭ

Метод	R ²	Q ²	R ² _{test}	A
CoMFA	0.80	0.68	0.49	3
CoMSIA	0.76	0.65	0.52	3
EVA	0.84	0.7	0.36	4
HQSAR	0.84	0.72	0.3	4
Circular	0.86	0.72	0.58	3

где R² — коэффициент детерминации, определяющий «подгонку» модели под обучающую выборку, Q² — коэффициент детерминации в условиях скользящего контроля, определяющий устойчивость полученного решения, R²_{test} — коэффициент детерминации для тестовой выборки, отражающий прогностическую способность QSAR модели, A — количество параметров модели.

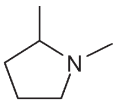
Как видно из табл. 1 все модели достаточно адекватны и имеют близкие статистические характеристики, однако предлагаемый подход (Circular) обладает несколько более высокой прогностической способностью (наибольшее значение R²_{test}).

В отличие от многих других QSAR подходов (например, CoMFA, EVA, HQSAR), в рамках разработанного метода существует возможность оценить вклады различных фрагментов молекулярной структуры в исследуе-

мое свойство. В табл. 2 приведены некоторые фрагменты и соответствующие им вклады.

Таблица 2

Вклады различных фрагментов соединений в исследуемую активность

Фрагмент	Положение	Вклад
-C ₆ H ₅	X	0.802
-OH	Y	0.43
-C=O	X	0.128
	Y	-0.241
=N-	X	-0.830
-NH-	X	-0.841
-SH	Y	-0.904

Как видно из приведенных данных, наиболее увеличивающими сродство к АПЭ являются ароматические заместители и гидроксогруппы, наличие же сульфогидрильных, и аминогрупп уменьшает величину сродства.

Таким образом, на наш взгляд, разработанный метод генерации структурных параметров позволяет вполне эффективно решать задачи QSAR и осуществлять молекулярный дизайн перспективных соединений. Важной особенностью циркулярных моделей является также сравнительно небольшое количество структурных параметров, что значительно снижает вычислительную сложность задачи, не требует дополнительных процедур отсеивания незначимых дескрипторов, и упрощает интерпретацию данных.

Литература

1. R. Todeschini, V. Consonni: Handbook of Molecular Descriptors. Wiley-wich Publishers, Weinheim, 2000. — P. 256.
2. Богатский А. В., Тригуб Л. П., Кузьмин В. Е. Гармонический анализ конформаций макроциклов // Докл. АН СССР. 1984. — Т. 278, № 6. — С. 1401–1405.
3. В. А. Палюлин, Н. С. Зефилов. Концепция молекулярного подобию и ее использование для прогнозирования свойств химических соединений // Успехи химии. — 1975. — С. 88–90.
4. Land A. H, Doig A. G. An autmatic method of solving discrete programming problems// Econometrica. — 1960. — Vol. 28. — P.497–520.
5. King R. B., Rouvray D. H. Graph Theory and Topology in Chemistry. Elsevier, Amsterdam, 1987.
6. Кузьмин В. Е., Огниченко Л. Н., Артеменко А. Г., Ляховский А. В. Прикладные возможности концепции информационного поля молекулы // 3-я Всероссийская конференция «Молекулярное моделирование». — Москва, 2003. — С. 80.
7. J. Sutherland., J O'Brien., L. A Weaver. A comparison of methods for modeling quantitative structure-activity relationships // J. Med. Chem. 2004 — Vol.47. — P.5541-5554.
8. Rappe A. K., Casewit C. J., Colwell K. S., Goddard W. A. Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations // J. Am.Chem. Soc. — 1992. — Vol. 114. — P. 10024–10035.

9. Klebe G., Abraham U., Mietzner T. Molecular Similarity Indices in Comparative Analysis (CoMSIA) of Molecules to Correlate and Predict Their Biological Activity // J. Med. Chem. — 1994. — Vol. 37. — P. 4130–4146.
10. Doweyko A. M. *The Hypothetical Active Site Lattice*. An Approach to Modeling Active Sites from Data on Inhibitor Molecules // J. Math. Chem. — 1988. — Vol 31. — P. 1396–1406.

В. Є. Кузьмін^{1,2}, А. В. Ляховський², А. Г. Артеменко²

¹Одеський Національний Університет ім. І. І. Мечнікова,
кафедра органічної хімії,
вул. Дворянська 2, Одеса, 65082, Україна

²Фізико-хімічний інститут ім. О. В. Богатського НАН України
Люстдорфська дорога 86, г. Одеса, 65082, Україна

ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ ЦИРКУЛЯРНИХ МОДЕЛЕЙ ПРИ ВИРІШЕННІ ЗАВДАНЬ QSAR

Резюме

Запропоновано новий метод генерації структурних параметрів для вирішення завдань QSAR на основі циркулярних моделей молекул. Даний метод дозволяє формувати структурні параметри, що адекватно описують варіювання молекулярної структури. Показано ефективність розробленого підходу на прикладі інгібіторів ангіотензін-перетворюючого ензиму.

Ключові слова: QSAR, молекулярні дескриптори, граф, інформаційне поле, гармонічний аналіз.

Victor E. Kuz'min^{1,2}, Anatoly V Liahovsky², Anatoly G. Artemenko²,

¹ Odessa I. I. Mechnykov National University,
Dworyanskaya 2, Odessa, 65082, Ukraine

²A.V. Bogatsky Physical-Chemical Institute NAS of Ukraine,
Lustdorfskayka Doroga 86, Odessa, 65082, Ukraine

EFFICIENCY STUDY APPLICATION OF CIRCULAR MOLECULAR MODEL TO QSAR ANALYSIS

Summary

New method of structural descriptors generation for the solving of QSAR tasks has been developed by authors. It is shown that mentioned approach allows generation of the set of structural parameters with the quite adequate level of description of molecules and their properties. Efficiency of developed approach was shown base on ACE inhibitors.

Key words: QSAR, molecular descriptors, molecular graph, information field, Fourier analysis.