К.К. ЭРНАЗАРОВ

(Инновационный Центр Фитотехнологии НАН КР , Бишкек, Кыргызстан)

ТЕОРЕТИЧЕСКОЕ ИССЛЕДОВАНИЕ КОРРЕЛЯЦИОННОЙ ВЗАИМОСВЯЗИ НЕКОТОРЫХ ТОПОЛОГИЧЕСКИХ И МОЛЕКУЛЯРНЫХ ДЕСКРИПТОРОВ СТАНДАРТНЫХ АМИНОКИСЛОТ С ПОМОЩЬЮ МЕТОДА НАИМЕНЬШИХ КВАДРАТОВ

Аннотация

В данной статье получено численное выражение взаимосвязи молекулярных, топологических дескрип-торов и дескриптора связности стандартных аминокислот с помощью методов наименьших квадратов. Полученный метод можно использовать при прогнозировании биологической активности органического соединения с неизвестной биологической активностью.

Ключевые слова: топологические дескрипторы, модификационный индекс Рандича, индекс связности Винера, метод наименьших квадратов.

Кілт сөздер: топологиялық дискрипторлар, Рандичтің модификациялық көрсеткіші, Винердің байланыс көрсеткіші, ең кіші шаршылар әдісі.

Keywords: topological descriptors, modified Randic connectivity index, Wiener connectivity index, the least squares method.

В последние годы теория графов стала важнейшим математическим инструментом, широко используемым во многих областях науки, в том числе и в химии. В современной теоретической химии усиливается тенденция шире использовать математический аппарат для описания молекулярных структур и химических превращений. На первоначальном этапе речь шла о решении частных математических задач: исчисление изомеров, применение комбинаторики для описания химических соединений, определение информационного содержания химических графов. Ныне все более очевидной становится практическая ценность общих топологических подходов для решения химических задач. Можно отметить, что именно теория химического строения молекул в значительной мере способствовала развитию теорию графов. Так, класс нециклических графов - так называемых деревьев – впервые был установлен еще в середине XIX века в работах Кэли на основании молекулярной структуры разветвленных предельных углеводородов. До

сегодняшнего дня использование топологии и теории графов в химии, проходя новые ступени, одновременно создали успехи при исследовании физико-химических свойств различных химических соединений [1]. Оценка фармацевтического свойства биологических активных веществ с помощью топологических индексов модернизируется. В результате, на сегодняшний день, в химии появились новые интеграционные отрасли науки, как математическая и вычислительная химия. Значит, математическое и компьютерное моделирование в исследовании молекулярной строении веществ и их физико- химических свойств при оценке точности экспериментальных методов вносит весомый вклад и одновременно создает большую конкуренцию. Роль молекулярного моделирования в химии достаточно велика, несмотря на явный приоритет экспериментальных исследований в этой области естествознания. Наиболее значимы такие теоретические результаты для тех случаев, когда невозможно, крайне трудно или слишком дорого получить их экспериментальными средствами.

Топологические индексы разрабатывались с целью корреляции и прогнозирования широкого ряда физико-химических параметров химических веществ. Аддитивными параметрами являются те параметры, которые увеличиваются при соединении к молекулам новых инкрементов на некоторое фиксированное значение. Примерами близких к аддитивности параметров являются теплота образования, температура кипения, молекулярная рефракция, плотность и различные теплоты реакции [5].

Таблица 1 – Значения индекса Винера и молярной рефракции стандартных аминокислот, рассчитанные с помощью программы Dragon 5.5

№	Аминокислота	Условное обозначение	Индекс Винера, W	Молярная рефракция, AMR
1	Глицин	Gly R	18	16,15
2	Аланин	Ala R	29	20,57
3	Аргинин	Arg R	29	43,58
4	Аспарагин	Asn R	96	28,43
5	Аспарагиновая кислота	Asp R	96	26,6
6	Цистеин	Cys R	46	28,24
7	Глутамин	Gln R	136	33,2
8	Глутаминовая кислота	Glu R	136	31,36
9	Гистидин	His R	165	37,57
10	Изолейцин	Ile R	92	34,17
11	Лейцин	Leu R	96	34,25

12	Лизин	Lys R	143	37,89
13	Метионин	Met R	102	37,9
14	Фенилаланин	Phe R	212	45,19
15	Пролин	Pro R-R	62	28,14
16	Серин	Ser S	46	22,12
17	Треонин	Thr R	65	26,53
18	Триптофан	Trp R	369	56,28
19	Тирозин	Tyr R	268	46,87
20	Валин	Val R	65	29,57

В последние годы одним из наиболее важных достижений в области топологических индексов явилось их возрастающее применение при разработке лекарственных препаратов и других биологически активных веществ. Хотя эта область приложений до сих пор находится в зачаточном состоянии, ее потенциальная роль в будущем огромна. Создание лекарственных препаратов требует огромных расходов, а также времени для исследования. Очевидно, что использование топологических индексов, позволяющих миновать некоторые утомительные стадии исследований и значительно сократить время разработки, имеет первостепенное значение для фармацевтической промышленности.

Некоторые ранние работы в этой области основывались на простом обобщении индекса связности Рандича R(G). Для характеристики биологически активных веществ Рандичем, а также Киром предложен обобщенный индекс связности. Эти индексы принимают форму

$$R(G) = \sum_{e_f} \left(\upsilon_i \upsilon_j\right)^{-\frac{1}{2}} \tag{1}$$

где R(G) – индекс Рандича; υ_i и υ_j – соответственно степени вершин i и j в графе G; суммирование проводится по всем ребрам G.

$$H_{R(G)} = \sum_{edg} \left(\upsilon_1 \upsilon_2 \ldots\ldots \upsilon_{h+1}\right)^{-\frac{1}{2}} \tag{2}$$

где υ_1 ,υ_2 ,υ_3 , ……. υ_{h+1} – степени соответствующих вершин и суммирование проводится по всем путям в графе длины $h \succ 1$.

Этот подход применялся во многих областях, включая анестезирующее действие соединений, неспецифическую наркотическую активность, ингибирование ферментов, противогрибковое действие и для предсказания биоконцентрации вредных химических веществ в живых системах.

Наиболее важными свойствами, определяющими активность лекарственных препаратов, являются три физико-химические свойства: липофильность, электронное распределение и форма молекул. Все они связаны с топологической структурой молекул, хотя этот факт, по-видимому, в значительной мере недооценивался. На сегодняшний день электронное распределение в молекулах стандартных аминокислот и формы строения этих молекул изучены на достаточном уровне. Поэтому в данной работе мы в основном подробно остановимся на липофильности стандартных аминокислот и на индексах Рандича, и Винера соответствующие каждой аминокислоте.

В данной статье с помощью метода наименьших квадратов в вычислительной математике приблизительно вычислен функциональная взаимосвязь двух дескрипторов стандартных аминокислот. Точнее говоря, как первичный дескриптор выбран модифицированный индекс связности Рандича, а как вторичный дескриптор выбрана молекулярная масса стандартных аминокислот и их численные значения, вычисленные с помощью программы Dragon 5.5 (табл. 2).

Таблица 2 – Вычисленные значения молекулярных масс стандартных аминокислот и их модифицированные индексы Рандича с помощью программы Dragon 5.5.

№	Название аминокислот	Краткое обозначение	Молекулярная масса,	Модифицированный индекс Рандича, R_M
1	Глицин	Gly	75	15,3
2	Аланин	Ala	89	17,3
3	Валин	Val	117	22,8
4	Лейцин	Leu	131	25,67
5	Изолейцин	Ile	131	26
6	Аспарагиновая Кислота	Asp	133	26,818
7	Глутаминовая Кислота	Glu	147	29,5
8	Серин	Ser	105	21,23
9	Треонин	Thr	119	23,34
10	Цистеин	Cys	121	24,

11	Метионин	Met	149	32,56
12	Аргинин	Arg	89	35
13	Лизин	Lys	146	29,88
14	Гистидин	His	155	33,6
15	Пролин	Pro	115	24,44
16	Фенилаланин	Phe	165	35,63
17	Тирозин	Tyr	181	38,57
18	Триптофан	Trp	204	45
19	Аспарагин	Asn	132	26
20	Глутамин	Gln	146	29,5

С помощью программы Dragon 5.5. можно вычислить множество дескрипторов, которые имеют корреляции высокого порядка. Но компьютерные эксперименты, проведенные с помощью программы Dragon 5.5, показали, что некоторые дескрипторы, несмотря на то, что они обусловливают друг друга, не имеют корреляции высшего порядка [3]. Главным характеризующим биологической активности органических веществ является их липофильность. При вычислении корреляции между липофильностью и количеством углеродов в молекуле стандартных аминокислот расчет показывал, что они имеют корреляции в меньшей степени ($k < 75$). Но, несмотря на это, как показывали теоретические и экспериментальные исследования, активность биологически активных органических веществ, бесспорно, зависит от количества углеродов в молекуле в некотором характере. Значит, возможно, что исследование корреляции между дескриптором, описывающим свойства биологически активного соединения и дескриптором, описывающим ее структуру, не всегда дает результат высшей точности и удаление этих погрешностей требуют более сложных вычислений статистической математики.

Исследование корреляции между дескрипторами, которые описывают только структуру биологически активного соединения, намного облегчает вышеуказанную проблему. Так как проведенные компьютерные эксперименты показали, что модифицированный индекс Рандича стандартных аминокислот и их молекулярные массы имеют между собой корреляции на высшем порядке. Структурный индекс связности Рандича и его разные производные являются одним из основных дескрипторов, характеризующие активность биологически активных веществ. А молекулярную массу можно рассматривать как количественно- структурный дескриптор, и в стандартных аминокислотах эта величина в какой-то мере зависит от количества углеродов в соединений. В дальнейшем ходе работы с помощью метода наименьших квадратов и используя таблицу 2, приближенным путем установлена функциональная зависимость модификационного индекса Рандича стандартных аминокислот от их молекулярной

массы. Для этого берем функцию зависимости модификационной индекса Рандича $X_{mod.}$ от молекулярной массы M:

$$X_{mod.}(M) = A \cdot M + B$$

где, $X_{mod.}(M)$ – модифицированный индекс Рандича, соответствующий на молекулярную массу M; A и B – постоянные коэффициенты; M – молекулярная масса стандартной аминокислоты.

Здесь неизвестные коэффициенты A и B определяются с помощью метода наименьших квадратов [4]:

$$\left(\sum_{\kappa=1}^{N} M_k^2\right) A + \left(\sum_{k=1}^{N} M_k\right) B = \sum_{k=1}^{N} M_k R_k$$

$$(\textstyle\sum_{k=1}^{N} M_k) A + NB = \textstyle\sum_{k=1}^{N} R_k \qquad (3)$$

Суммы, требуемые для нормальных уравнений (3), легко получить, используя значения табл. 2. Линейная система, включающая A и B, имеет вид

Решением линейной системы будет $\mathrm{A} \approx 0{,}23$ и $\mathrm{B} \approx -2{,}3$. Таким образом, выражение, построенное методом наименьших квадратов, имеет вид:

$$R_M(M) = 0{,}23M - 2.3 \qquad (4)$$

С помощью данных в таблице 1 и методом наименьших квадратов установлены линейная функциональная зависимость между молярной рефракцией AMR и индексом связности Винера W стандартных аминокислот:

$$AMR(W) = 0{,}9412 \cdot W + 22{,}51 \qquad (5)$$

Полученное выражение (4) является линейной функцией зависимости модифицированного индекса связности Рандича от молекулярной массы стандартных аминокислот. А выражение (5) так же является функцией зависимости молярной рефракции стандартных аминокислот AMR от индекса связности Винера W. Устанавливая численное выражение между двумя дескрипторами, имеющие между собой корреляции высокого порядка, с помощью разных интерполяционных методов, можно предсказать биологическую активность неизвестного соединения.

ЛИТЕРАТУРА

1 Уилсон Р. Введение в теорию графов. М.: «Мир» - 1977., -202 стр.
2 CS Chem3D Ultra Molecular Modeling and Analysis. 2001 Cambridge Soft. www.cambridgesoft.com.
3 Dragon Evaluation Version 5.5. -2007. (Software for Molecular Descriptor Calculations)
4 Джон Г. Мэтьюз, Куртис Д. Финк. Численные методы. Использование MatLab, 3-е издание. Перевод с английского. – М.: «Вильямс», 2001. – 720 стр.
5 Соловьев М. Е., Соловьев М. М. Компьютерная химия. М: СОЛОН-Пресс, 2005, -536 стр.
6 Cherif F. Matta. Quantum Biochemistry. Electronic structure and biological activity, WILEY-VCH Verlag GmbH & KGaA. -2011, 920 pages.

REFERENCES

1 Wilson R. Introduction to graph theory, Moscow: “Mir” – 1977., -202 p., (in Russ)
2 CS Chem3D Ultra Molecular Modeling and Analysis. 2001 Cambridge Soft. www.cambridgesoft.com.
3 Dragon Evaluation Version 5.5. -2007. (Software for Molecular Descriptor Calculations)
4 John H. Mathews and Kurtis K. Fink. Numerical Methods Using Matlab (3th Edition). Translation from English. Moscow: “Viliams”, 2001. -720 p., (in Russ)
5 Soloviev M.E., Soloviev M. M. Computer Chemistry, - Moscow: “SOLON-Press”, 2005, -536 p., (in Russ)
6 Cherif F. Matta. Quantum Biochemistry. Electronic structure and biological activity, WILEY-VCH Verlag GmbH & KGaA. -2011, 920 pages.

Резюме

К.К.Эрназаров

СТАНДАРТТЫ АМИНҚЫШҚЫЛДАРДЫҢ МОЛЕКУЛАЛЫҚ ДЕСКРИПТОРЫ МЕН

ТОПОЛОГИЯЛЫҚ КОРРЕЛЯЦИЯЛЫҚ ӘРЕКЕТТЕСУЛЕРІН ЕҢ КІШІ ШАРШЫЛАР ӘДІСІМЕН

ТЕОРИЯЛЫҚ ЗЕРТТЕУ

(Қырғыз Республикасы ҰҒА Инновациялық фитотехнология орталығы, Бішкек. Қырғызстан)

Бұл мақалада стандартты аминқышқылдардың молекулалық дескрипторының және топологиялық корреляциялық әрекеттесулері арасындағы байланыс ең кіші шаршылар

әдісінің көмегімен сандық түрде сипат-талған. Алынған әдісті пайдаланып белгісіз биологиялық белсенді заттардың биологиялық белсенділігін бол-жауда қолдануға болады.

Кілтті сөздер: топологиялық дискрипторлар, Рандичтің модификациялық индексі, Винердің байланыс көрсеткіші, ең кіші шаршылар әдісі.

Summary

K. K. Ernazarov

(Innovation Phytotechnology Centre, National Academy of Sciences

of Kyrgyz Republic, Bishkek, Kyrgyzstan)

THEORETICAL RESEARCH OF CORRELATION RELATIONSHIP SOME TOPOLOGICAL AND MOLECULAR DESCRIPTORS OF STANDARD AMINO ACIDS USING THE LEAST SQUARES METHOD

In this article a numerical expression of the relationship of molecular, topological descriptors and descriptor connectivity of standard amino acids is gained by the method of least squares. The resulting method can be used in predicting the biological activity of organic compounds with unknown biological activity.

Keywords: topological descriptors, modified Randic connectivity index, Wiener connectivity index, the least squares method.

Поступила 14.05.2013 г.