

## ГИПЕРГРАФОВЫЕ МОДЕЛИ МОЛЕКУЛ УГЛЕВОДОРОДОВ И ИХ ПРИМЕНЕНИЕ В КОМПЬЮТЕРНОЙ ХИМИИ

М.И. Скворцова\*, профессор, И.И. Фасхутдинова, студент,

Н.А. Михайлова, старший преподаватель

кафедра Высшей и прикладной математики МИТХТ им. М.В. Ломоносова,

Москва, 119571 Россия

\*Автор для переписки, e-mail: [skvorivan@mail.ru](mailto:skvorivan@mail.ru)

**П**редложен новый способ представления структур насыщенных углеводородов в виде гиперграфов специальной конструкции. Проведено сравнение традиционной графовой модели и предложенной гиперграфовой модели. Для этой цели введен ряд количественных критериев, характеризующих эффективность применения той или иной модели при решении определенных задач компьютерной химии. Для исследований использована некоторая выборка углеводородов, представленных своими структурными формулами. Показано, что по всем рассмотренным критериям гиперграфовая модель превосходит графовую. Кроме того, на основе инвариантов предложенных гиперграфов для соединений заданной выборки построен ряд корреляций «структура–свойство».

**Ключевые слова:** граф, гиперграф, инварианты графа, инварианты гиперграфа, компьютерная химия, корреляции «структура–свойство», насыщенные углеводороды.

### Введение

Компьютерная химия – область науки, vznikшая на стыке химии, математики и информатики. К основным задачам компьютерной химии обычно относят следующие задачи: 1) создание баз данных, содержащих химическую информацию, и программных средств для работы с этими базами; 2) компьютерная генерация всех возможных реакций между заданными реагентами («компьютерный синтез»); 3) построение моделей связи «структура–свойство» и прогнозирование свойств химических соединений при помощи этих моделей; 4) компьютерное конструирование химических структур с заданными свойствами («молекулярный дизайн»); 5) вычисление характеристик электронного и пространственного строения молекул квантово-химическими методами. Кроме этих основных задач, выделяют также и ряд узкоспециальных задач, связанных с построением математических моделей химических соединений (как правило, в виде графов), исследованием этих моделей, разработкой различных комбинаторных алгоритмов на графах и т.д. [1, 2].

Исследования в области компьютерной химии являются одним из приоритетных направлений развития современной химии. Об этом говорит, например, тот факт, что Нобелевская премия по химии за 2013 г. была присуждена именно за работу в этой области [3].

Важное место в компьютерной химии занимают способы описания структуры молекулы. Один из наиболее распространенных подходов к этой проблеме основан на представлении молекулы в виде взвешенного графа, называемого молекулярным. Вершины такого графа соответствуют атомам молекулы, а ребра – связям. Веса вершин кодируют атомы различной химической природы, а веса ребер – связи

разного типа. Такой граф задается своей матрицей  $A = (x_{ij})$ , где  $x_{ii}$  – вес  $i$ -ой вершины графа,  $x_{ij}$  ( $i \neq j$ ) – вес ребра  $(i, j)$ . Обычно предполагают, что  $x_{ij} = 0$  для несмежных вершин  $i$  и  $j$ . Для углеводородных молекул, как правило, используют простые графы, изображающие лишь углеродный скелет молекулы. В этом случае полагают  $x_{ii} = 0$ ,  $x_{ij} = 1$  для смежных вершин  $i$  и  $j$ ,  $x_{ij} = 0$  для несмежных вершин  $i$  и  $j$ , а матрица  $A$  – это матрица смежности соответствующего графа.

Каждой молекулярной структуре могут быть поставлены в соответствие разнообразные инварианты ее молекулярного графа – числа, определяемые по графу (т.е. по его матрице), не зависящие от способа нумерации его вершин. Инварианты графов широко используются для построения математических моделей связи «структура–свойство» в качестве так называемых молекулярных дескрипторов [4]. Однако инварианты используются не только в корреляциях «структура–свойство». Они находят применение и при кодировании, упорядочивании, поиске химических структур в базах данных. В связи с этим задача поиска инвариантов графов с наименьшей степенью вырождения является актуальной [5].

При решении ряда задач компьютерной химии используются также и локальные вершинные инварианты графа (ЛВИ) (например, при поиске канонической нумерации вершин молекулярного графа [6] или при поиске группы симметрии графа [7]). Для эффективного решения подобных задач важно найти такие ЛВИ, которые могли бы количественно «различить» (иначе говоря, дифференцировать) топологи-

чески неэквивалентные вершины графа, т.е. вершины, принадлежащие разным орбитам группы симметрии графа. В связи с этим задача поиска ЛВИ, обладающих максимальной способностью к дифференциации таких вершин, является актуальной. Другое возможное применение ЛВИ – построение по ним новых инвариантов графов или кодов графов [4].

Как известно, обобщением понятия «граф» является понятие «гиперграф» [8]. Гиперграф  $H$  представляет собой некоторое множество вершин  $V = \{v_1, \dots, v_n\}$  и некоторое множество гиперребер  $E = \{e_1, \dots, e_p\}$ , где гиперребро  $e_i$  – это некоторое непустое подмножество множества вершин  $V$ , которое может содержать любое число элементов. Таким образом, обычный граф  $G$  – это частный случай гиперграфа  $H$ , когда любое гиперребро содержит ровно 2 вершины и называется ребром. При изображении гиперграфа на плоскости его гиперребра обычно выделяют путем обведения соответствующих вершин замкнутой линией; если же гиперребро содержит 2 вершины, то их соединяют между собой отрезком прямой. Гиперграф  $H$  задается своей матрицей инцидентий  $B = (b_{ij})$ , где  $b_{ij} = 1$ , если вершина  $v_i = 1$  принадлежит ребру  $e_j$ , и  $b_{ij} = 0$  в противном случае. Для  $H$  может быть также определена матрица смежности  $A = (a_{ij})$ , где  $a_{ij}$  (при  $i \neq j$ ) равно числу гиперребер, которым принадлежит пара вершин  $i$  и  $j$ , и  $a_{ii} = 0$ . Определение инварианта или ЛВИ гиперграфа такое же, как и для простого графа; в частности, инварианты и ЛВИ гиперграфов можно вычислять по каким-либо их матрицам по алгоритмам, разработанным для простых графов.

Гиперграфы также находят применение в компьютерной химии. Это связано с тем, что обыкновенные графы не всегда позволяют адекватно описывать химические соединения, имеющие неклассическое строение (например, молекулы с делокализованными многоцентровыми связями, к которым, в частности, относятся элементоорганические соединения). В ряде работ для описания строения таких соединений было предложено использовать гиперграфы (см., например, [5, 9, 10]). В [10] были рассмотрены графовые и гиперграфовые модели для довольно большого набора соединений неклассического строения, для них были вычислены некоторые известные инварианты и показано, что инварианты гиперграфов существ-

венно менее вырождены, чем аналогичные инварианты графов. Были отмечены и другие преимущества гиперграфового представления таких молекул. Однако, несмотря на перспективность использования инвариантов гиперграфов в моделях связи «структура–свойство» для таких соединений, в литературе подобные корреляции отсутствуют. Следует отметить, что число работ, посвященных применению гиперграфов в химии, очень невелико (по сравнению с числом работ, связанных с обычными графами в химии). При этом для описания структур классического строения гиперграфы не использовались. Это связано, в основном, с тем, что структурная формула молекулы, являющаяся основой для построения каких-либо других молекулярных моделей, уже сама является графом. Отметим, что во всех графовых моделях молекулы две вершины соединяют ребром тогда и только тогда, когда между соответствующими атомами молекулы имеется химическая связь, указанная в структурной формуле. Однако некоторое взаимодействие имеется и между атомами, не связанными непосредственно химической связью, причем чем ближе они друг к другу, тем сильнее это взаимодействие. Эти соображения приводят к идее считать «связанными» не два, а несколько близко расположенных друг к другу атомов (по какому-либо критерию) и использовать для описания молекулярной структуры классического строения гиперграф.

Цели настоящей работы: 1) предложить способ описания структур органических соединений классического строения, представляющих собой насыщенные углеводороды, в терминах гиперграфов; 2) сравнить эффективности использования традиционной графовой модели и предложенной гиперграфовой модели таких соединений при решении ряда задач компьютерной химии; 3) на основе инвариантов предложенных гиперграфов для выборки соединений рассматриваемого класса построить ряд корреляций «структура–свойство».

#### Гиперграфовые модели молекул насыщенных углеводородов

Для исследований была использована выборка, состоящая из 41 соединения, содержащая все насыщенные углеводороды (как ациклические, так и циклосодержащие) с числом атомов углерода 6 и 7, из работы [11]. На рис.1 изображены молекулярные графы углеродных остовов этих соединений (вершины такого графа соответствуют атомам углерода в молекуле, а ребра – химическим связям между этими атомами).

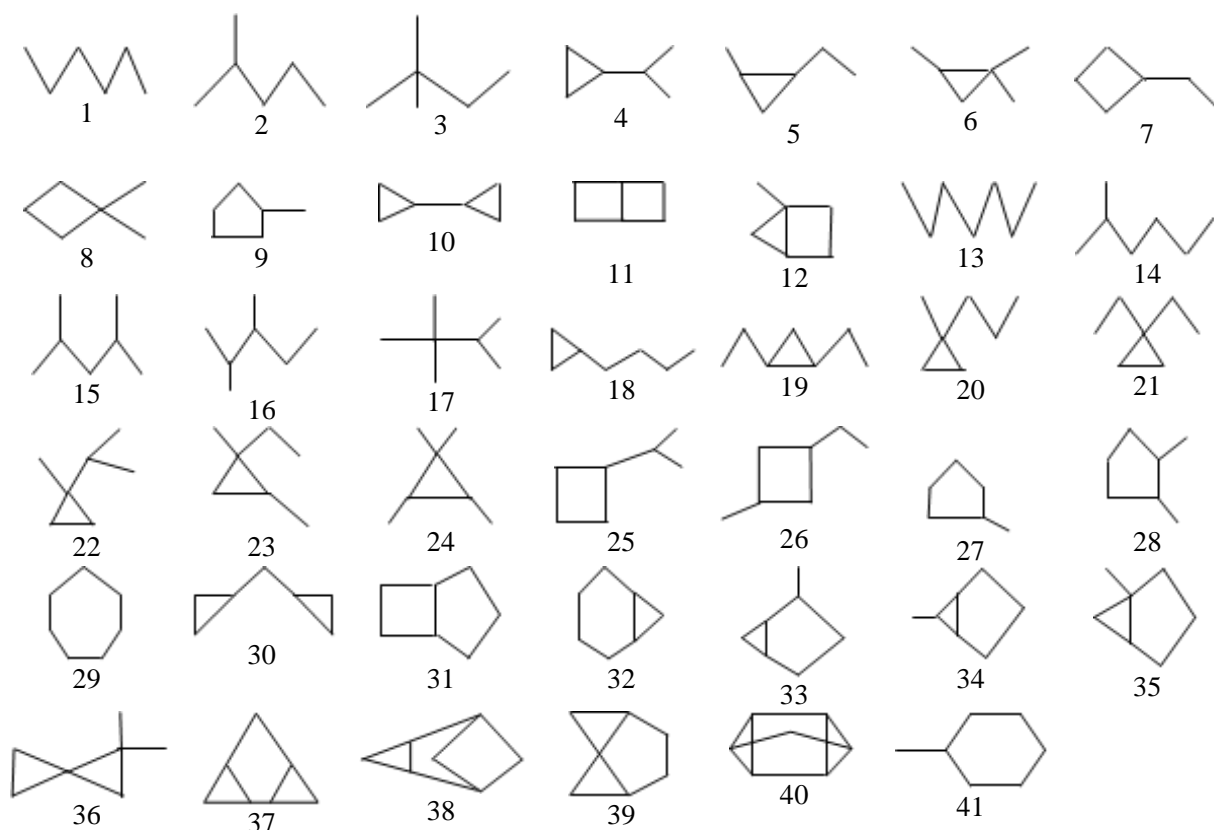


Рис. 1. Соединения, использованные для построения гиперграфовых моделей.

Опишем процедуру построения гиперграфовых моделей вышеуказанных соединений.

Пусть  $G$  – граф с занумерованными вершинами, соответствующий углеродному остову такой молекулы. Пусть  $n$  и  $q$  – числа вершин и ребер графа  $G$ , соответственно.

Как известно, граф можно задать матрицей смежности или матрицей инцидентий, при этом предпочтение отдается матрице смежности. Матрица смежности  $A(G) = (a_{ij}(G))$  графа  $G$  определяется так: элемент  $a_{ij}(G)$  при  $i \neq j$  равен числу ребер, соединяющих вершины  $i$  и  $j$ , т.е. он равен 0 или 1;  $a_{ji}(G) = a_{ij}(G)$ ,  $a_{ii}(G) = 0$  ( $i, j = 1, \dots, n$ ). Для построения матрицы инцидентий  $B(G) = (b_{ij}(G))$  графа  $G$  необходимо занумеровать также и его ребра. Тогда по определению  $b_{ij}(G) = 1$ , если вершина  $i$  и ребро  $j$  инцидентны (иначе говоря, вершина  $i$  принадлежит ребру  $j$ ), и  $b_{ij}(G) = 0$  в противном случае ( $i = 1, \dots, n; j = 1, \dots, q$ ).

Построим гиперграф  $H$  следующим образом: его множество вершин – это множество вершин графа  $G$ , гиперребро  $e_i$  ( $i = 1, \dots, n$ ) состоит из вершины  $i$  и всех ее соседей, с которыми она соединена ребром в  $G$ . Как известно, гиперграф можно задать матрицей смежности или матрицей инцидентий, при этом

предпочтение отдается матрице инцидентий. Матрица смежности  $A(H) = (a_{ij}(H))$  гиперграфа  $H$  определяется так: элемент  $a_{ij}(H)$  при  $i \neq j$  равен числу гиперребер, к которым принадлежит пара вершин  $i$  и  $j$ ,  $a_{ji}(H) = a_{ij}(H)$ ,  $a_{ii}(H) = 0$ . Для построения матрицы инцидентий  $B(H) = (b_{ij}(H))$  гиперграфа необходимо занумеровать также и его гиперребра. Тогда по определению  $b_{ij}(H) = 1$ , если вершина  $i$  и гиперребро  $j$  инцидентны (иначе говоря, вершина  $i$  принадлежит гиперребру  $j$ ), и  $b_{ij}(H) = 0$  в противном случае ( $i = 1, \dots, n; j = 1, \dots, n$ ).

Для любого графа и любого гиперграфа справедлива общая формула:  $A = B \cdot B^* - V$ , где  $B$  – матрица инцидентий графа (гиперграфа),  $B^*$  – матрица, получаемая транспонированием матрицы  $B$ , а  $V$  – диагональная матрица, у которой на диагонали стоят степени вершин графа (гиперграфа). Обозначим степень  $i$ -ой вершины графа через  $v_i$ ; тогда степень  $i$ -ой вершины построенного нами гиперграфа равна  $v_i + 1$ .

Пример молекулярного графа  $G$  и соответствующего ему гиперграфа  $H$  дан на рис.2. Граф  $G$  имеет множество вершин  $\{1, 2, 3, 4, 5, 6\}$  и множество ребер  $h_1 = \{1, 2\}$ ,  $h_2 = \{2, 4\}$ ,

$h_3 = \{4,5\}$ ,  $h_4 = \{4,6\}$ ,  $h_5 = \{3,4\}$ ,  $h_6 = \{2,3\}$ .  
 Гиперграф  $H$  имеет то же самое множество вершин и множество гиперребер  $e_1 = \{1,2\}$ ,

$e_2 = \{1,2,3,4\}$ ,  $e_3 = \{2,3,4\}$ ,  $e_4 = \{2,3,4,5,6\}$ ,  
 $e_5 = \{4,5\}$ ,  $e_6 = \{4,6\}$ . Матрицы  $A(G)$ ,  $B(G)$ ,  
 $A(H)$ ,  $B(H)$  для  $G$  и  $H$  даны в табл.



Рис. 2. Пример графа  $G$  и соответствующего ему гиперграфа  $H$ .

Таблица 1. Матрицы  $A(G)$ ,  $B(G)$ ,  $A(H)$ ,  $B(H)$  для графа  $G$  и гиперграфа  $H$ , изображенных на рис. 2

	1	2	3	4	5	6
1	0	1	0	0	0	0
2	1	0	1	1	0	0
3	0	1	0	1	0	0
4	0	1	1	0	1	1
5	0	0	0	1	0	0
6	0	0	0	1	0	0

$A(G)$

	$h_1$	$h_2$	$h_3$	$h_4$	$h_5$	$h_6$
1	1	0	0	0	0	0
2	1	1	0	0	0	1
3	0	0	0	0	1	1
4	0	1	1	1	1	0
5	0	0	1	0	0	0
6	0	0	0	1	0	0

$B(G)$

	1	2	3	4	5	6
1	0	2	1	1	0	0
2	2	0	3	3	1	1
3	1	3	0	3	1	1
4	1	3	3	0	2	2
5	0	1	1	2	0	1
6	0	1	1	2	1	0

$A(H)$

	$e_1$	$e_2$	$e_3$	$e_4$	$e_5$	$e_6$
1	1	1	0	0	0	0
2	1	1	1	1	0	0
3	0	1	1	1	0	0
4	0	1	1	1	1	1
5	0	0	0	1	1	0
6	0	0	0	1	0	1

$B(H)$

**Некоторые локальные вершинные инварианты графа и гиперграфа и исследование их способности различать топологически неэквивалентные вершины графов**

Определим на основе матрицы смежности ряд простейших локальных вершинных инвариантов (ЛВИ) графа и гиперграфа. Пусть  $A^2(G) = (a_{ij}^{(2)}(G))$  – квадрат матрицы смежности графа  $G$  (аналогично определяется матрица  $A^2(H) = (a_{ij}^{(2)}(H))$ ). Рассмотрим следующие ЛВИ:

- 1)  $w_i^{(1)} = \sum_j a_{ij}$  ( $w_i^{(1)} = v_i$  для графа  $G$ );
- 2)  $w_i^{(2)} = \sum_j a_{ij}^2$  ( $w_i^{(2)} = w_i^{(1)}$  для графа  $G$ );
- 3)  $w_i^{(3)} = a_{ii}^{(2)}$  ( $w_i^{(3)} = w_i^{(1)}$  для графа  $G$ );
- 4)  $w_i^{(4)} = \sum_j a_{ij}^{(2)}$ ;
- 5)  $w_i^{(5)} = w_i^{(3)} + w_i^{(4)}$ .

Исследуем способность этих ЛВИ графа и гиперграфа количественно различать топологически неэквивалентные вершины в графах  $G$  исходной выборки соединений, и сравним полученные результаты для графовой и гиперграфовой моделей. Введем следующие две количественные характеристики этой способности,  $N_1$  и  $N_2$ , позволяющие провести такое сравнение.

Число  $N_1$  – это число структур в выборке, в которых оказались неразличимые неэквивалентные вершины. Очевидно, что  $0 \leq N_1 \leq 41$ . При этом  $N_1 = 0$  для самой «хорошей» ситуации, когда во всех структурах выборки произошло различение всех неэквивалентных вершин, и  $N_1 = 41$  для самой «плохой» ситуации, когда ни в одной из структур не произошла дифференциация таких вершин.

Для определения  $N_2$  сначала в каждой структуре найдем все неэквивалентные вершины, которые оказались неразличимы при

помощи данного ЛВИ (они будут принадлежать разным классам эквивалентности). Затем определим среди них минимальное количество вершин таких, что при некотором изменении у них значений ЛВИ все неэквивалентные вершины графа станут различимыми. Для примера рассмотрим граф на рис.2. Классы эквивалентности его вершин – {1}, {2}, {3}, {4}, {5,6}, а ЛВИ  $w_i^{(1)}$ , определяемые по графовой модели, равны степеням соответствующих вершин:  $w_1^{(1)}=1, w_2^{(1)}=3, w_3^{(1)}=2, w_4^{(1)}=4, w_5^{(1)}=w_6^{(1)}=1$ . Вершины двух разных классов {1} и {5,6} имеют одно и то же значение ЛВИ; следовательно, при помощи этого ЛВИ не удалось дифференцировать все неэквивалентные вершины. Однако если вершине №1 приписать какое-либо другое число, отличное от уже имеющихся значений ЛВИ, то все неэквивалентные станут различимы. Таким образом, минимальное число вершин, о котором говорилось выше, для этого графа равно 1 (в случае использования графовой модели и ЛВИ  $w_1^{(1)}$ ).

Определим  $N_2$  как общее число таких вершин во всех структурах изучаемой выборки. Так как общее число вершин во всех структурах равно 275, то  $0 \leq N_2 \leq 234$  (в самой «плохой» ситуации, когда в каждом графе значения ЛВИ всех вершин одинаковы, корректировка ЛВИ будет проводиться для  $275-41=234$  вершин). При этом в идеальной ситуации  $N_2 = 0$ , а в самой «плохой», когда для каждого графа ЛВИ всех его вершин одинаковы,  $N_2 = 234$ .

Естественно считать лучшей ту модель, для которой величина  $N_1$  (или  $N_2$ ) меньше.

В табл. 2 и 3 приведены полученные значения параметров  $N_1$  и  $N_2$  для всех 5 вариантов ЛВИ. Из таблиц следует, что по этим показателям гиперграфовая модель превосходит графовую (т.е. значения  $N_1$  и  $N_2$  для гиперграфовой модели меньше, чем для графовой), а для гиперграфовых ЛВИ  $w_i^{(4)}$  и  $w_i^{(5)}$  ситуация близка к идеальной ( $N_1 = 1$ ).

Таблица 2. Значения параметра  $N_1$  для ЛВИ  $w_i^{(1)} - w_i^{(5)}$  для  $G/H$

	$w_i^{(1)}$	$w_i^{(2)}$	$w_i^{(3)}$	$w_i^{(4)}$	$w_i^{(5)}$
$N_1$ для $G/H$	37/13	37/5	37/5	29/1	15/1

Таблица 3. Значения параметра  $N_2$  для ЛВИ  $w_i^{(1)} - w_i^{(5)}$  для  $G/H$

	$w_i^{(1)}$	$w_i^{(2)}$	$w_i^{(3)}$	$w_i^{(4)}$	$w_i^{(5)}$
$N_2$ для $G/H$	81/14	81/5	81/5	37/1	17/1

### Некоторые инварианты и коды графов и гиперграфов и исследование их вырождения

Определим коды  $K_j(G)$  и  $K_j(H)$  ( $j=1, \dots, 5$ ) для графов и гиперграфов как упорядоченные по возрастанию последовательности ЛВИ  $w_i^{(j)}$  ( $j=1, \dots, 5$ ). Например, для графа  $G$  и гиперграфа  $H$ , изображенных на рис. 2, коды  $K_1(G)$  и  $K_1(H)$  таковы:  $K_1(G) = \{1, 1, 1, 2, 3, 4\}$ ,  $K_1(H) = \{4, 5, 5, 9, 10, 11\}$ .

Рассмотрим следующие 10 инвариантов для  $G/H$ , определяемые по ЛВИ:

$$\min w_i^{(j)}, \max w_i^{(j)} \quad (j=1, 3, 4, 5); \quad S_1 = \sum w_i^{(1)}; \\ S_2 = \sum (w_i^{(1)})^2.$$

Кроме того, рассмотрим следующие 4 инварианта спектрального типа:  $\lambda_{\min}$  – минимальное собственное число матрицы смежности  $A$ ,  $\lambda_{\max}$  – максимальное собственное число матрицы смежности  $A$ ,  $\Delta\lambda = \lambda_{\max} - \lambda_{\min}$ ,  $\det A$  – определитель матрицы  $A$ . Напомним, что инвариант графа (гиперграфа) называется вырожденным на некотором множестве графов (гиперграфов), если в этом множестве имеются графы (гиперграфы) с одинаковыми значениями этого инварианта.

Введем следующую количественную характеристику степени вырождения кода (или инварианта):  $N_3$  – число различных значений кода (инварианта) на заданной выборке структур. Очевидно, что  $1 \leq N_3 \leq 41$ ; самая «хорошая» ситуация – когда  $N_3 = 41$ , а самая «плохая» – когда  $N_3 = 1$ . В табл. 4–6 приведены полученные значения параметра  $N_3$ .

Из таблиц следует, что по этим показателям гиперграфовая модель превосходит графовую (т.е. значение  $N_3$  для гиперграфовой модели больше, чем для графовой), и для гиперграфовой модели в ряде случаев ситуация оказывается идеальной или близкой к ней ( $N_3 = 41$  или  $N_3 = 40$ ).

Таблица 4. Значения параметра  $N_3$  для  $\lambda_{\min}, \lambda_{\max}, \Delta\lambda, \det A$

	$\lambda_{\min}$	$\lambda_{\max}$	$\Delta\lambda$	$\det A$
$N_3$ для $G/H$	32/38	36/40	39/41	7/40

Таблица 5. Значения параметра  $N_3$  для кодов  $K_j$  ( $j=1, \dots, 5$ ) и инвариантов  $S^{(1)}, S^{(2)}$  для  $G/H$

	$K_1$	$K_2$	$K_3$	$K_4$	$K_5$	$S^{(1)}$	$S^{(2)}$
$N_3$ для $G/H$	27/40	27/41	27/41	39/40	39/41	6/20	18/40

Таблица 6. Значения параметра  $N_3$  для инвариантов  $\min w_i^{(j)}$  и  $\max w_i^{(j)}$  ( $j=1,3,4,5$ ) для  $G/H$

	$\min w_i^{(j)}$ ( $j=1$ )	$\min w_i^{(j)}$ ( $j=3$ )	$\min w_i^{(j)}$ ( $j=4$ )	$\min w_i^{(j)}$ ( $j=5$ )	$\max w_i^{(j)}$ ( $j=1$ )	$\max w_i^{(j)}$ ( $j=3$ )	$\max w_i^{(j)}$ ( $j=4$ )	$\max w_i^{(j)}$ ( $j=5$ )
$N_3$ для $G/H$	2/6	2/9	5/27	6/28	3/8	3/14	7/33	8/39

**Инварианты гиперграфов в корреляциях «структура–свойство»**

Для соединений №№ 1–41, представленных на рис. 1, были рассмотрены значения их температуры кипения  $t_{\text{экс.}}(^{\circ}\text{C})$ , взятые из работы [11]

(см. табл. 7), а также значения их энергий  $E$  (в эВ), рассчитанные нами при помощи стандартной компьютерной программы HyperChem (методом молекулярной механики; силовое поле – MM<sup>+</sup>) [2]. В табл. 7 приведены значения  $E' = -10E$ , округленные до второго знака после запятой.

Таблица 7. Значения температур кипения и энергий углеводородов, представленных на рис. 1

№	$t_{\text{экс.}}(^{\circ}\text{C})$	$t_{\text{расч.}}(^{\circ}\text{C})$	$E'$	$E'_{\text{расч.}}$	№	$t_{\text{экс.}}(^{\circ}\text{C})$	$t_{\text{расч.}}(^{\circ}\text{C})$	$E'$	$E'_{\text{расч.}}$
1	68.7	68.24	818.14	824.91	22	81.5	81.98	880.09	881.07
2	60.3	61.05	818.12	825.80	23	85.2	83.00	880.97	881.07
3	49.7	50.00	817.99	825.47	24	78.0	76.04	881.80	880.82
4	58.3	59.51	751.99	752.38	25	92.7	92.28	881.86	881.03
5	63.0	60.10	752.86	752.38	26	89.5	92.64	882.54	880.68
6	52.6	51.08	753.53	749.09	27	91.3	94.01	891.02	881.07
7	70.7	71.11	754.47	752.38	28	95.6	93.14	891.10	880.82
8	53.6	57.86	754.97	749.09	29	118.4	116.65	891.96	881.03
9	71.8	71.37	763.44	749.09	30	102.0	109.09	813.31	819.06
10	76.0	77.53	686.23	693.71	31	110.5	113.73	826.77	813.92
11	83.0	77.85	687.92	688.07	32	116.0	111.24	825.75	815.66
12	60.5	61.79	684.79	684.79	33	100.0	97.78	823.48	813.92
13	98.5	98.01	945.56	944.89	34	103.0	97.27	824.47	815.66
14	90.0	90.94	945.54	955.21	35	92.0	94.45	824.46	811.88
15	80.5	83.35	945.52	949.55	36	78.0	82.81	809.47	813.92
16	89.8	84.76	945.52	958.56	37	105.0	106.34	756.87	767.85
17	80.9	76.89	945.40	959.79	38	107.0	105.86	748.77	755.59
18	98.0	99.66	879.38	877.83	39	107.5	105.66	749.32	755.59
19	90.0	90.58	880.28	879.08	40	108.5	106.45	679.50	702.39
20	84.9	89.31	880.12	881.03	41	101.0	101.1	891.94	881.07
21	88.6	89.31	880.13	881.03					

Для поиска корреляций «структура–свойство» для этих двух свойств использовались такие структурные характеристики молекул, как  $n$  и  $q$ , а также инварианты гиперграфов  $\lambda_{\min}, \lambda_{\max}$ ,  $\Delta\lambda = \lambda_{\max} - \lambda_{\min}$ ,  $\det A$ ,  $S_1, S_2, \min w_i^{(j)}, \max w_i^{(j)}$  ( $j=1,3,4,5$ ),  $n_2(H)$  – число вершин степени 2 в гиперграфе  $H$ ,  $W(H)$  – индекс Винера гиперграфа  $H$ . Кроме того, наряду с вышеуказанными параметрами рассматрива-

лись и некоторые простейшие функции от них ( $f(x) = x^2, f(x) = x^{0.5}$ ), а также их попарные произведения. При этом инварианты обычных графов (кроме  $q$ ) не рассматривались.

Опишем процедуру нахождения  $W(H)$ . Первоначально строится  $D(H) = (d_{ij}(H))$  – матрица расстояний гиперграфа  $H$ , в которой элемент  $d_{ij}(H)$  равен длине кратчайшей цепи

из гиперребер, соединяющей вершины  $i$  и  $j$ ; для пары вершин, принадлежащих одному гиперребру,  $d_{ij}(H)=1$ . По определению  $W(H)=0.5\sum d_{ij}(H)$ . Инвариант  $W(H)$  является аналогом хорошо известного индекса Винера  $W(G)$  для простых графов, который строится по матрице расстояний  $D(G)$  по такой же формуле [4]. Следует отметить, что при по-

строении  $D(H)$  вышеописанным способом происходит потеря информации о структуре гиперграфа, так как кратности гиперребер при этом не учитываются. Заметим также, что по матрице  $D(G)$  граф  $G$  определяется однозначно, а по матрице  $D(H)$  – нет.

Для температуры кипения  $t_{\text{кип.}}$  была получена следующая корреляция с 4 параметрами:

$$t_{\text{кип.}} = 30.487 + 1.732nq - 3.542\Delta\lambda(H) - 13.792(n_2(H))^{1/2} + 0.165(4n - 2q)^2$$

$$(\delta_{\text{max}} = 8.63, s = 3.62, R = 0.982).$$

Здесь  $\delta_{\text{max}}$  – максимальная ошибка аппроксимации,  $s$  – среднеквадратичное отклонение,  $R$  – коэффициент корреляции. Экспериментальные и расчетные значения температур ки-

пения  $t_{\text{экс.}}(^{\circ}\text{C})$  и  $t_{\text{расч.}}(^{\circ}\text{C})$  приведены в табл. 7.

При добавлении пятого параметра  $(\min w_i^{(1)})^2$  можно получить более точную корреляцию:

$$t_{\text{кип.}} = 36.091 + 1.781nq - 3.295\Delta\lambda(H) - 16.981(n_2(H))^{1/2} + 0.153(4n - 2q)^2 - 0.222(\min w_i^{(1)})^2$$

$$(\delta_{\text{max}} = 7.09, s = 3.09, R = 0.987).$$

Для величины  $E'$  была получена корреляция с тремя параметрами следующего вида:

$$E' = 407.164 + 1.842(4n - 2q)^2 + 1.179qW(H) - 0.151W^2(H)$$

$$(\delta_{\text{max}} = 22.89, s = 8.70, R = 0.994),$$

где статистические характеристики  $\delta_{\text{max}}$ ,  $s$ ,  $R$  имеют тот же смысл, что и выше. Значения энергий молекул  $E'$  и  $E'_{\text{расч.}}$ , рассчитанные при помощи программы HyperChem и при помощи полученного корреляционного уравнения, приведены в табл. 7.

### Заключение

В настоящей работе: 1) предложен один из возможных вариантов построения гиперграфовых моделей структур насыщенных углеводородов, которые традиционно представляют

графами; 2) проведено сравнение графовых и гиперграфовых моделей одних и тех же соединений по ряду количественных показателей (всего рассмотрено 29 показателей); показано, что во всех случаях гиперграфовая модель превосходит графовую; кроме того, в ряде случаев результат для гиперграфовой модели – «идеальный» или близкий к «идеальному»; 4) на основе инвариантов гиперграфов построен ряд корреляций «структура–свойство»; при этом инварианты графов (кроме числа ребер в графе) для построения корреляций не использовались.

### ЛИТЕРАТУРА:

1. [www.ru.wikipedia.org/wiki/компьютерная\\_химия](http://www.ru.wikipedia.org/wiki/компьютерная_химия)
2. Соловьев М.Е., Соловьев М.М. Компьютерная химия. М.: СОЛОН-Пресс, 2005. 536 с.
3. [www.newsru.com/world/09oct2013/nobel.html](http://www.newsru.com/world/09oct2013/nobel.html)
4. Станкевич М.И., Станкевич И.В., Зефилов Н.С. Топологические индексы в органической химии // Успехи химии. 1988. Т. 57. № 3. С. 337–366.
5. Константинова Е.В. Графы, гиперграфы и их инварианты для характеристики молекулярных структур: дис. ... канд. техн. наук. Новосибирск: РАН. Сиб. отд-ние. Новосибирский ин-т орган. химии, 1992. 198 с.
6. Stankevich I.V., Gal'pern E.G., Chistyakov A.L., Baskin I.I., Skvortsova M.I., Zefirov N.S., Tomilin O.B. Spectral theory of graphs in chemistry. 1. Projection operators and canonical numeration of graph vertices // J. Chem. Inform. & Computer Sci. 1994. V. 34. № 5. P. 1105–1108.
7. Stankevitch M.I., Tratch S.S., Zefirov N.S. Combinatorial models and algorithms in chemistry. Search for isomorphisms and automorphisms of molecular graphs // J. Comput. Chem. 1988. V. 9. № 4. P. 303–320.
8. Емеличев В.А., Мельников О.И., Сарванов В.И., Тышкевич Р.И. Лекции по теории графов: учеб. пособие. М.: ЛИБРОКОМ, 2009. 392 с.

9. Konstantinova E.V., Skorobogatov V.A. Molecular hypergraphs: the new representation of nonclassical molecular structures with polycentric delocalized bonds // *J. Chem. Inform. & Computer Sci.* 1995. V. 35. № 3. P. 472–473.

10. Konstantinova E.V., Skorobogatov V.A. Application of hypergraph theory in chemistry // *Discrete Mathematics*. 2001. V. 235. P. 365–383.

11. Skvortsova M.I., Fedyaev K.S., Palyulin V.A., Zefirov N.S. Molecular design of chemical compounds with prescribed properties from QSAR models containing the Hosoya index // *Internet Electronic Journal of Molecular Design*. 2003. V. 2. № 2. P.70–85.

## **HYPERGRAPH MODELS OF HYDROCARBON MOLECULES AND THEIR APPLICATIONS IN COMPUTER CHEMISTRY**

**M.I. Skvortsova<sup>@</sup>, I.I. Fashutdinova, N.A. Mikhailova**

*M.V. Lomonosov Moscow University of Fine Chemical Technologies, Moscow, 119571 Russia*

<sup>@</sup>*Corresponding author e-mail: skvorivan@mail.ru*

*A new form of representing the structures of saturated hydrocarbons as hypergraphs of special kind is suggested. A vertex of such hypergraph corresponds to a carbon atom in the molecule, and a hyperedge is defined as a set of vertices corresponding to some fixed carbon atom and all carbon atoms bounded with it. The comparison of the traditional graph model and the suggested hypergraph model using definite quantitative criteria related to some tasks of computer chemistry is fulfilled. For these investigations some set of hydrocarbons presented by their structural formulae is used. In particular, different invariants and codes of graphs and hypergraphs were calculated and their degeneration on the given set of structures was studied and compared. Two models were also compared by the ability of some local vertex invariants derived from them to distinguish topologically nonequivalent vertices in the traditional graph model. It is shown that in all cases in accordance with the used criteria the hypergraph model is better than the graph model. Besides, a number of structure–property relationships on the base of invariants of the suggested hypergraphs was obtained.*

**Keywords:** *hypergraph, graph invariants, hypergraph invariants, computer chemistry, structure–property relationships, saturated hydrocarbons.*