

ТОПОЛОГИЧЕСКИЕ ИНДЕКСЫ МАКСИМАЛЬНЫХ ВНЕШНЕПЛОСКИХ ГРАФОВ С ДВУМЯ СИМПЛИЦИАЛЬНЫМИ ВЕРШИНАМИ

Ю. Л. Носов

*Guardian Software Inc., Калгари, Канада; удалённый офис: Липецк, Россия
yl.nosov@yandex.ru*

Рассматриваются максимальные внешнеплоские графы с двумя симплициальными вершинами, уложенные на равносторонней треугольной решётке (РМВП-графы). Слабо двойственные графы РМВП-графов изоморфны и подобны молекулярным графам изомеров сопряжённых полиеновых углеводородов (СПУ). Поэтому РМВП-графы удобно применять для прогнозирования физико-химических свойств СПУ. Предложен новый подход к поиску многочисленных корреляций структура — свойство между топологическими индексами (ТИ) РМВП-графов из некоторого заданного множества и физико-химическими параметрами изомеров (СПУ). Сущность такого подхода заключается в предварительном отборе ТИ, пригодных для прогнозирования свойств неразветвлённых изомеров СПУ, с последующим поиском корреляций структура — свойство уже между отобранными ТИ и физико-химическими параметрами изомеров СПУ. Пригодность предложенного подхода была доказана с помощью расчётного эксперимента. В рамках этого подхода исследовались ТИ РМВП-графов порядков $n = 8, 10, \dots, 28$, слабые двойственные графы которых изоморфны и геометрически подобны молекулярным графам модели СПУ, принятой в работе S. J. Cyvin (1996). Из множества ТИ Балабана, Винера, Шульца, Рандича, Харари и Загребской группы были выбраны простой J и модифицированный J' индексы Балабана.

Ключевые слова: *плоский граф, максимальный внешнеплоский граф, топологический индекс, 3D-индекс Винера, приложения теории графов в естественных науках.*

Введение

Одним из наиболее важных практических применений теории графов [1, с. 9–11] является использование графов и их топологических инвариантов [1, с. 24] для определения корреляционных зависимостей структура — свойство химических соединений. При этом графы представляют молекулярную структуру химических соединений, а топологические инварианты графов отображают их структурные свойства. Поэтому используемые в таком представлении графы называют *молекулярными графами* [2, с. 8], а их топологические инварианты — *топологическими индексами* [2, с. 101].

Впервые топологический инвариант W (теперь называемый индексом Винера) был использован в 1947 году американским химиком Г. Винером для установления корреляционных зависимостей между значениями W и свойствами ациклических химических соединений, молекулярные графы которых являются деревьями (см., например, [3]). Таким образом, индекс Винера был первым топологическим индексом (ТИ), который был использован в качестве молекулярного дескриптора [4, ch. XVIII, XIX] для построения корреляций структура — свойство.

В настоящее время известно около 3300 молекулярных дескрипторов, большинство из которых не являются ТИ [4]. Однако, несмотря на это, ТИ, как и другие мо-

лекулярные дескрипторы, применяются для построения корреляций структура — свойство и структура — биологическая активность [4].

Широкий интерес вызывают сопряжённые полиеновые углеводороды (СПУ) [5]. Это объясняется тем, что они, с одной стороны, входят в класс полиенов, имеющих большое теоретическое и практическое значение. С другой — они имеют относительно простую структуру и большое количество изомеров [6, с. 2210] и конформеров [6, с. 2204]. По этой причине в последние десятилетия было опубликовано большое количество экспериментальных и теоретических исследований СПУ (см., например, [7; 8]).

Молекулярные графы разветвлённых изомеров СПУ со стёртыми вершинами водорода представляют собой простые деревья [1, с. 48] с максимальной степенью вершин 3. Углы между двумя смежными рёбрами в таких графах изменяются в пределах от 120° до 128° (см., например, [7; 8]). Поэтому сегменты [9, с. 218, 219] молекулярных графов разветвлённых изомеров СПУ, а также и молекулярные графы неразветвлённых изомеров СПУ есть простые зигзагообразные цепи [1, с. 26].

В работе [10] предложена методика перечисления изомеров СПУ, использующая максимальные внешнеплоские графы (МВП-графы) [1, с. 131]. В основу этой методики была положена плоская модель молекул СПУ. По этой модели молекулярные графы СПУ представляют собой деревья, в которых рёбра имеют одинаковую длину, а углы между любыми парами смежных рёбер равны 120° . Там же было показано, что слабо двойственные графы [11, с. 211] решётчатых МВП-графов, т. е. МВП-графов, уложенных на решётке [12] из равносторонних треугольников, изоморфны [1, с. 24] и геометрически подобны [13, с. 225] плоским молекулярным графам модели СПУ. Далее решётчатые МВП-графы мы будем называть РМВП-графами.

В силу вышеуказанных особенностей структуры молекул СПУ молекулярные графы большинства изомеров СПУ изоморфны и отличаются друг от друга только относительным расположением вершин и рёбер. Поэтому использование обычных ТИ для прогнозирования свойств разветвлённых изомеров СПУ затруднительно, а для неразветвлённых изомеров невозможно.

Для преодоления этой трудности в [14] было предложено использовать ТИ РМВП-графов с двумя симплициальными вершинами. Для графов этого класса в [14; 15] была построена структурная классификация, включающая графы четырёх типов: «лестница», ступенчатые и квазиступенчатые графы первого и второго видов. Показано наличие соответствия, т. е. изоморфизма [1, с. 24] и геометрического подобия [13, с. 225], между слабо двойственными графами РМВП-графов вышеперечисленных классов и молекулярными графами изомеров и конформеров СПУ.

Настоящая работа является продолжением работ [14; 15] автора по МВП-графам. В работе предлагается новый подход к поиску многочисленных корреляций структура — свойство между ТИ (РМВП-графов) из некоторого заданного множества I_o индексов I_1, I_2, \dots, I_{k_I} и физико-химическими параметрами неразветвлённых изомеров СПУ. Сущность такого подхода заключается в предварительном отборе ТИ, пригодных для прогнозирования свойств неразветвлённых изомеров СПУ, с последующим поиском корреляций структура — свойство уже между отобранными ТИ и физико-химическими параметрами изомеров СПУ.

В качестве критерия отбора (критерия пригодности) ТИ было принято наличие высокой дискриминирующей способности ТИ [9, с. 243] и высокий коэффициент корреляции [16] между ним и эталонными ТИ. Причём в качестве эталонных

ТИ использовались 3D-индексы Винера ${}^3DW(G_i^*)$ и ${}^3DW_H(U_i)$ слабо двойственных графов G_i^* РМВП-графов G_i и графов U_i , полученных объединением графов G_i и G_i^* соответственно [17; 1, с. 36]. Такая методика позволит отбросить как непригодные ТИ, так и физико-химические параметры, имеющие плохую корреляцию с 3D-индексом Винера.

В статье изложена методика нового подхода и приведены результаты исследования и отбора ТИ РМВП-графов, пригодных для прогнозирования свойств неразветвлённых изомеров СПУ.

1. Структура и основные свойства МВП-графов

Везде далее через $G = (V, E)$ обозначаются конечные связные неориентированные графы без петель и кратных рёбер с множеством вершин $V(G)$ и множеством рёбер $E(G)$. Число $|V|$ вершин и число $|E|$ рёбер графа G обозначаются через $n(G)$ и $m(G)$ соответственно (или через n и m , если ясно, о каком графе идёт речь). Число вершин графа называется *порядком графа*, а число рёбер графа — его *размером*. Далее (при необходимости) будем обозначать через G_n граф порядка n , а через $G_{n,i}$, если не оговорено другое — i -й граф порядка n . Все не определяемые ниже термины можно найти в [1].

Маршрутом в графе называется чередующаяся последовательность вершин и рёбер $v_0, x_1, v_1, \dots, v_{n-1}, x_n, v_n$, которая начинается и кончается вершиной и каждое ребро x_i ($i = 1, 2, \dots, n$) последовательности инцидентно двум вершинам v_{i-1} и v_i . Маршрут называется *цепью*, если все его рёбра различны, и *простой цепью*, если все его вершины различны. Цепь называется *циклом*, если $v_0 = v_n$. Простая цепь называется *простым циклом*, если $v_0 = v_n$ и $n \geq 3$. Простой цикл называется *гамильтоновым циклом*, если в него входят все вершины графа.

Обозначим через P_n простую цепь с n вершинами, через $\Delta(G)$ — максимальную степень вершин графа G , а через $G[A]$ — подграф графа G , индуцированный подмножеством вершин $A \subseteq V$. Подмножество вершин V' графа G называется *кликой*, если $G[V']$ — полный граф. Вершина называется *симплициальной вершиной* графа G , если множество смежных с ней вершин индуцирует в G клику.

Внешнеплоским графом называется плоский граф, все вершины которого принадлежат одной (обычно внешней) грани. Причём *внешней гранью* называется неограниченная грань. *Максимальным внешнеплоским графом* называется такой внешнеплоский граф, который при добавлении хотя бы одного ребра перестаёт быть внешнеплоским [1, с. 131].

Геометрически двойственным графом плоского графа G называется граф G^* , каждой вершине f_i^* которого соответствует некоторая грань f_i графа G , причём две вершины f_i^* и f_j^* графа G^* смежны только тогда, когда соответствующие им грани f_i, f_j графа G имеют одно общее ребро [1, с. 138]. *Слабо двойственным графом* графа G называется граф, полученный из геометрически двойственного графа G^* удалением вершины (и смежных с ней рёбер), соответствующей внешней грани [11, с. 211]. Поскольку в статье рассматриваются только слабо двойственные графы, для обозначения слабо двойственного графа плоского графа G мы будем употреблять аналогичное обозначение G^* , используя символ «звездочка».

Необходимые определения и известные результаты по МВП-графам можно найти в [1; 10; 11; 14; 15]. Известно, что все МВП-графы представляют собой триангуляции выпуклого многоугольника [1, с. 131]. Отсюда следует, что любой n -вершинный МВП-граф имеет: а) $2n - 3$ рёбер; б) внешнюю грань, ограниченную гамильтоновым циклом; в) $n - 2$ внутренних треугольных граней; г) $n - 3$ внутренних рёбер

и не менее двух вершин степени 2, являющихся симплициальными вершинами [1, следствие 11.9 (а)].

Количество симплициальных вершин МВП-графов определяет их структуру и тип соответствующих двойственных графов. В соответствии с [14] МВП-графы с двумя симплициальными вершинами называются *МВП-графами класса «2-цепь»*, а МВП-графы, имеющие не менее трёх симплициальных вершин, — *МВП-графами класса «2-дерево»*.

Слабо двойственный граф любого МВП-графа G_n класса «2-дерево» порядка n является простым деревом T_{n-2} с максимальной степенью вершин $\Delta(G_n) = 3$, (слабо двойственный граф МВП-графа G_n класса «2-цепь» есть простая цепь P_{n-2} , которая является частным случаем дерева) [11].

Любой МВП-граф G_n класса «2-цепь» обладает следующими свойствами [14]:

1) гамильтонов цикл C_n графа G_n разделяется симплициальными вершинами s_1, s_2 на две цепи $P_A = (s_1, x_1, x_2, \dots, x_a, s_2)$ и $P_B = (s_1, y_1, y_2, \dots, y_b, s_2)$, причём $X = \{x_1, x_2, \dots, x_a\}$ и $Y = \{y_1, y_2, \dots, y_b\}$;

2) граф G_n является *LL-плоским* графом [14, с. 97], поскольку его можно уложить на плоскости так, что все вершины цепей P_A, P_B будут размещены на двух параллельных линиях A, B без пересечений рёбер и каждое ребро является отрезком прямой.

МВП-графы класса «2-цепь» подразделяются на три типа, различающихся по своей структуре. Графы этих трёх типов в [14, с. 97–98] получили название МВП-графов типов «веер», «лестница» и «цепь». Образцы МВП-графов типа «лестница», «цепь» и «веер» приведены на рис. 1.

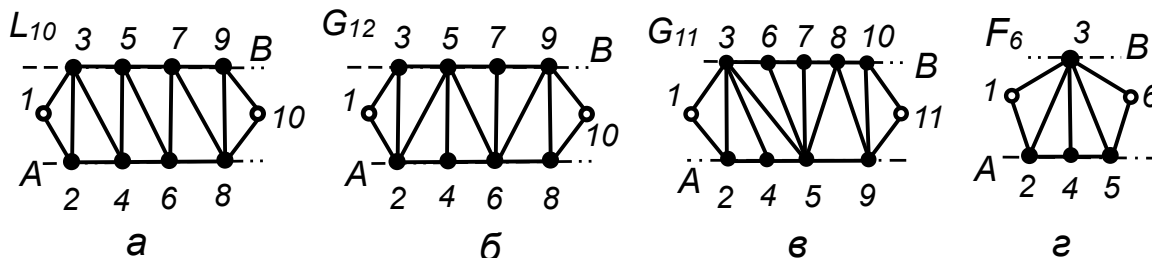


Рис. 1. МВП-графы: а — типа «лестница»; б — типа «цепь» ступенчатый; в — типа «цепь» квазиступенчатый; г — типа «веер»

В работе [15] классификация МВП-графов, включающая графы типа «лестница», «веер» и «цепь», была дополнена ступенчатыми графами [15, с. 291] и квазиступенчатыми графами 1-го и 2-го видов [15, с. 310]. Образцы ступенчатых и квазиступенчатых графов представлены на рис. 1, б, в соответственно.

Как было показано в [15], структура ступенчатых МВП-графов и графов типа «лестница» определяется конфигурацией их цепей $P_A = (s_1, x_1, x_2, \dots, x_a, s_2)$ и $P_B = (s_1, y_1, y_2, \dots, y_b, s_2)$ и рёбер, соединяющих вершины из этих цепей. Эти рёбра можно подразделить на два вида: «поперечные» рёбра $\{x_i, y_i\}$ ($i = 1, 2, \dots, p$), где $p = \lfloor (n-2)/2 \rfloor$, и «диагональные» рёбра одного из двух видов: «диагональные» рёбра первого вида $\{x_i, y_{i+1}\}$ ($i = 1, 2, \dots, g$) или «диагональные» рёбра второго вида $\{y_i, x_{i+1}\}$, причём $g = p - 1$ при чётном n и $g = p$ при нечётном n .

Ясно, что МВП-графы, имеющие «диагональные» рёбра одного вида, являются графами типа «лестница», а графы, имеющие «диагональные» рёбра двух видов, являются ступенчатыми графами. Другое определение ступенчатых и квазиступенчатых МВП-графов, использующее терминологию «поперечных» и «диагональных» рёбер, было дано в [15, с. 291].

Квазиступенчатые МВП-графы имеют одну или несколько пар вершин вида x_i, y_i , не соединённых поперечным ребром $\{x_i, y_i\}$. В качестве примера на рис. 1, в изображен квазиступенчатый МВП-граф G_{11} .

Определение 1. Квазиступенчатый МВП-граф G_n является квазиступенчатым МВП-графом *первого вида* при выполнении следующих условий:

- 1) число n вершин графа G_n чётное;
- 2) МВП-граф G_{n-2} , полученный из графа G_n последовательным удалением симплициальных вершин S_1 и S_2 , является или МВП-графом типа «лестница», или ступенчатым МВП-графом.

В противном случае, т. е. при невыполнении хотя бы одного условия, граф G_n является квазиступенчатым МВП-графом *второго вида*.

Определение 2. *Треугольной решёткой* на плоскости называется множество всех точек пересечения трёх семейств параллельных прямых, разбивающих плоскость на равные треугольники. Точки пересечения называются *узлами решётки*, а прямые — *образующими решётки* [12].

Все графы, изображённые на рис. 1, даны в двухлинейном представлении. Однако для химических приложений более полезно представлять МВП-графы так, чтобы все их внутренние грани были равносторонними треугольниками.

Определение МВП-графов в таком представлении впервые было дано в [15]. Однако, поскольку в этом определении не совсем корректно были использованы термины «дыра» и пересечение вершин, рёбер и граней, дадим новое определение.

Определение 3. [15]. МВП-граф G называется *МВП-графом, уложенным на треугольной решётке*, или *решётчатым МВП-графом (РМВП-графом)* при выполнении следующих условий: (а) вершины и рёбра МВП-графа располагаются на узлах и образующих решётки из равносторонних треугольников без наложения вершин, рёбер и граней; (б) максимальная степень $\Delta(G)$ вершин МВП-графа ограничена, причём $\Delta(G) \leq 6$.

Здесь под наложением вершин, рёбер, граней имеется в виду наличие соответственно вершин, рёбер, граней, занимающих одно и то же место в решётке.

На рис. 2 в качестве примера представлены образцы решётчатых МВП-графов без наложений вершин, рёбер и граней (рис. 2, а), с наложением двух вершин (рис. 2, б) и с наложением двух рёбер (рис. 2, в), с наложением двух рёбер (рис. 2, г).

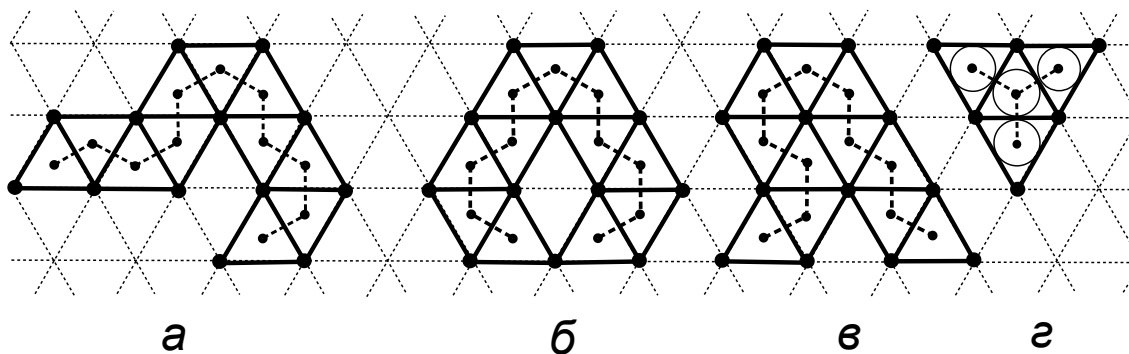


Рис. 2. Решётчатые МВП-графы: решётчатый МВП-граф (а); МВП-граф с наложением двух вершин (б); МВП-граф с наложением двух рёбер (в); шесть апофем, образующих три смежных ребра слабо двойственного графа (г)

Легко видеть, что если все вершины слабо двойственного графа G^* расположены в центрах треугольных граней РМВП-графа G , то в графе G^* все его рёбра будут

иметь одинаковую длину и любые два смежных ребра образуют угол 120° . Этот факт является очевидным следствием из свойств равносторонних треугольников [18, с. 213]. Действительно, в этом случае любое ребро слабо двойственного графа состоит из двух апофем (или двух радиусов вписанной окружности) двух смежных граней, а любые смежные апофемы образуют угол в 120° (см. рис. 2, *з*).

Таким образом, слабо двойственные графы РМВП-графов изоморфны и геометрически подобны молекулярным графам в плоской модели неразветвлённых изомеров СПУ, представленной в [10].

Другой пример РМВП-графов порядка $n = 8$ представлен на рис. 3: граф типа «лестница» (рис. 3, *а*), ступенчатый граф (рис. 3, *б*), квазиступенчатые графы первого вида (рис. 3, *в, з*), квазиступенчатые графы второго вида (рис. 3, *д, е*).

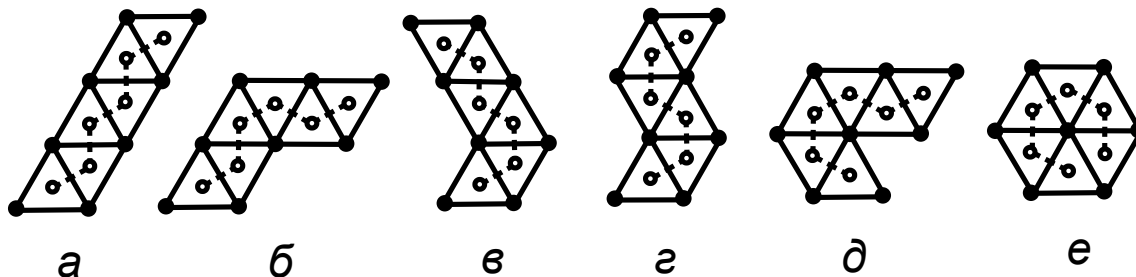


Рис. 3. Решётчатые МВП-графы: типа «лестница» (*а*); ступенчатый (*б*); квазиступенчатые графы первого вида (*в, з*), квазиступенчатые графы второго вида (*д, е*)

Определение 4. *Элементарным расширением МВП-графа* называется операция добавления вершины и двух рёбер, соединяющих её с вершинами внешнего ребра МВП-графа [19]. *Элементарным расширением МВП графа s -типа* называется операция элементарного расширения, при котором новая вершина соединяется рёбрами с вершинами ребра, инцидентного симплициальной вершине [14].

Определение 5. [14]. МВП-граф порядка 4 с двумя симплициальными вершинами называется *графом диамонд* и обозначается через D_4 .

Утверждение 1. [14]. *Любой n -вершинный ($n \geq 5$) МВП-граф класса «2-цепь» может быть получен из графа диамонд D_4 посредством последовательного выполнения $k = n - 4$ операций элементарного расширения s -типа.*

Из утверждения 1 следует, что любой n -вершинный МВП-граф G_n класса «2-цепь» получается из графа D_4 в результате выполнения последовательности (h_1, h_2, \dots, h_k) операций элементарного расширения s -типа. Таким образом, структура и форма МВП-графа G_n класса «2-цепь» определяются последовательностью k операций (h_1, h_2, \dots, h_k) .

Каждая операция элементарного расширения s -типа имеет два варианта выполнения, поскольку новая вершина может быть соединена рёбрами с концевыми вершинами одного из двух рёбер, инцидентных симплициальной вершине. Следовательно, каждому n -вершинному ($n = 4 + k$) МВП-графу G_n класса «2-цепь» можно поставить в соответствие некоторый двоичный кортеж $B(G_n) = (b_1, b_2, \dots, b_k)$, в котором элемент b_i принимает значение 0 или 1 в зависимости от того, на ребре какой цепи (P_B или P_A) выполняется операция h_i [15]. (Далее кортеж $B(G_n)$ мы будем называть *кортежем свойств*.)

Например, МВП-граф L_{10} типа «лестница» имеет кортеж свойств $B(L_{10}) = (1, 0, 1, 0, 1, 0)$, МВП-граф G_{11} типа «цепь» имеет кортеж свойств $B(G_{11}) = (1, 1, 0, 0, 0, 1, 0)$, а МВП-граф F_6 типа «веер» — кортеж свойств $B(F_6) = (0, 0)$ (см. рис. 1, *а, в, з* соответственно).

Замечание 1. Легко видеть, что при построении графа типа «веер» каждая операция элементарного расширения s -типа всегда выполняется на ребре одной цепи, например, цепи P_B (рис. 1, z), а при построении графа типа «лестница» каждая операция элементарного расширения s -типа выполняется поочерёдно — то на ребре цепи P_B , то на ребре P_A (рис. 1, a).

В работе [15] было доказано наличие соответствия (изоморфизма [1, с. 24] и геометрического подобия [13, с. 225]) между слабо двойственными графами РМВП-графов трёх новых видов и молекулярными графами (со стёртыми атомами водорода) изомеров и конформеров СПУ (см. [14, с. 313–314]). А именно: а) слабо двойственные графы РМВП-графов типа «лестница» соответствуют молекулярным графам транс-изомеров СПУ; б) слабо двойственные графы ступенчатых РМВП-графов соответствуют молекулярным графам цис- и смешанных изомеров СПУ; в) слабо двойственные графы квазиступенчатых РМВП-графов 1-го вида соответствуют молекулярным графам конформеров транс-изомеров СПУ; г) слабо двойственные графы квазиступенчатых РМВП-графов 2-го вида соответствуют молекулярным графам конформеров цис- и смешанных изомеров СПУ.

На рис. 4 представлены молекулярные графы (по модели Syvin S.J. [10]) изомеров и конформеров гексатриена. Легко видеть наличие геометрического подобия между слабо двойственными графами РМВП-графов порядка $n = 8$, представленными на рис. 3, и моделями молекулярных графов (со стёртыми атомами водорода) изомеров и конформеров гексатриена.

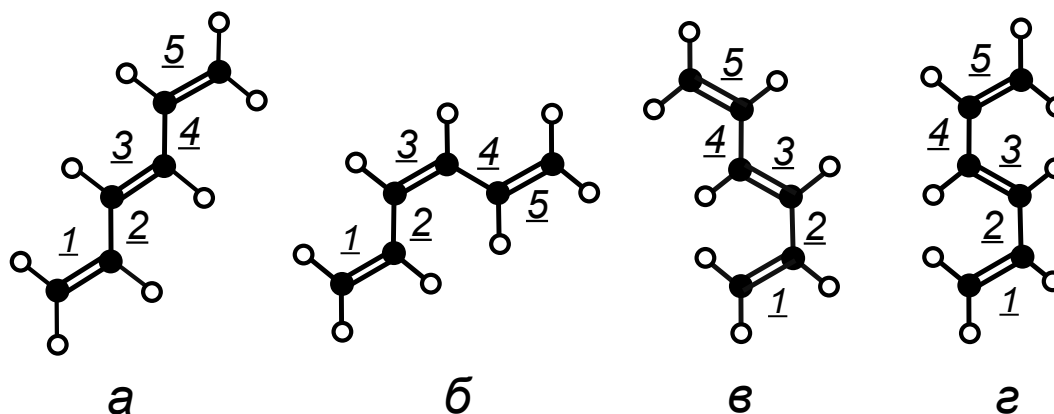


Рис. 4. Модели молекулярных графов изомеров и конформеров гексатриена: транс-изомер (a); цис-изомер (b); конформеры транс-изомера (v, z). Чёрными кружками показаны атомы углерода, пустыми кружками — атомы водорода; двумя параллельными линиями показаны двойные связи

В дальнейшем мы будем объединять РМВП-графы типа «лестница» и ступенчатые РМВП-графы одного порядка в один класс и называть их РМВП-графами класса «изомеры». Аналогично, квазиступенчатые РМВП-графы 1-го вида мы будем называть графами класса «конформеры транс-изомеров» СПУ, а квазиступенчатые РМВП-графы 2-го вида — графами класса «конформеры других изомеров» СПУ. Все квазиступенчатые РМВП-графы одного порядка будем называть РМВП-графами класса «конформеры».

В [15] были разработаны алгоритмы для генерации РМВП-графов по их кортежам свойств [14, с. 293–294], для генерации кортежей свойств ступенчатых РМВП-графов [14, с. 293] и кортежей свойств квазиступенчатых РМВП-графов первого и второго видов [14, с. 310].

2. Исследуемые топологические индексы графов

В этом разделе дадим необходимую терминологию и приведём расчётные формулы топологических индексов, подлежащих исследованию.

Инвариант графа в теории графов — некоторое обычно числовое значение или упорядоченный набор значений, характеризующие структуру графа $G = (V, E)$ и не зависящие от способа нумерации, маркировки, обозначения вершин или графического изображения графа. Инвариант графа G принимает одно и то же значение на любом графе, изоморфном G [1, с. 24; 2, с. 101]. Инвариант играет важную роль при проверке изоморфизма графов, а также в задачах компьютерной химии. Инвариант называется *полным*, если совпадение инвариантов графов необходимо и достаточно для установления изоморфизма.

Инвариант молекулярного графа в задачах компьютерной химии, характеризующий молекулярную структуру химического соединения, называется *топологическим дескриптором* [2; 4]. Числовой топологический дескриптор, имеющий высокую корреляцию с каким-нибудь физико-химическим свойством или биологической активностью химического соединения, называется *топологическим индексом* [2]. *Молекулярный граф* — связный неориентированный граф, находящийся во взаимно-однозначном соответствии со структурной формулой химического соединения таким образом, что вершинам графа соответствуют атомы молекулы, а рёбрам графа — химические связи между этими атомами [2, с. 8; 4, с. 529]. В большинстве случаев молекулярные графы отражают только углеродный скелет молекулы органического соединения (атомы водорода стёрты). Вершины молекулярного графа могут иметь пометки, отражающие тип атома, валентность и его координаты. Рёбра графа могут иметь пометки, отображающие кратность связи.

Возможность применения любого топологического индекса определяется тремя следующими факторами: а) наличием корреляционной связи между топологическим индексом молекулярного графа химического соединения и физико-химическими свойствами соединения; б) сложностью вычисления индекса; в) «дискриминирующей способностью», т. е. способностью индекса принимать разные значения на молекулярных графах химических изомеров.

В теоретико-графовом аспекте дискриминирующая способность — это способность различать n -вершинные неизоморфные графы заданного класса \mathcal{G}_n . Величина $S[I, \mathcal{G}_n]$ дискриминирующей способности индекса I определяется как отношение числа $n(I)$ разных значений индекса I , принимаемых на всех неизоморфных графах порядка n из класса \mathcal{G}_n , к числу $N(\mathcal{G}_n)$ этих графов, т. е. как $S[I, \mathcal{G}_n] = n(I)/N(\mathcal{G}_n)$ [9]. Далее, если ясно, о каком классе графов идёт речь, дискриминирующую способность индекса I будем обозначать через $S[I]$.

Очевидно, что дискриминирующая способность топологического индекса, являющегося полным инвариантом, равна единице. Однако следует отметить, что ни один из известных в настоящее время топологических индексов не является полным инвариантом.

К наиболее известным топологическим индексам относятся индекс Винера W , индекс Рандича χ , индексы Балабана J и J' , индекс Шульца, индекс Харари H , квадратичный индекс Харари H_2 , первый M_1 и второй M_2 Загребские индексы и другие. Наиболее старым (по длительности его использования) из них является *индекс Винера*, названный так в честь американского химика Г. Винера, который впервые в 1947 г. использовал его для установления корреляционных зависимостей между значениями индекса Винера и физико-химическими свойствами ациклических химических соединений [3].

Большинство из вышеперечисленных ТИ используют матрицу топологических расстояний [20]. Индекс Рандича использует степени вершин, а молекулярный топологический индекс Шульца использует и степени вершин и матрицу расстояний.

Для графа $G = (V, E)$ порядка n , вершины которого пронумерованы числами $1, 2, \dots, n$, матрицей расстояний графа G_n называется квадратная матрица $D(G) = [d_{ij}]$ размера $n \times n$, каждый элемент d_{ij} которой есть расстояние между вершинами i, j , определяемое как длина кратчайшей по числу рёбер простой цепи, соединяющей вершины i и j в графе G [20].

В табл. 1 приведены расчётные формулы для ТИ, подлежащих исследованию.

Таблица 1

Топологические индексы РМВП-графов, подлежащие исследованию

Наименование	Расчётная формула	Ссылка
Индекс Винера	$W(G) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n d_{ij} = \sum_{i=1}^n d_i, d_i = \sum_{j=1}^n d_{ij}$	[20]
Молекулярный топологический индекс	$MTI(G) = \sum_{i=1}^n \deg(i) \cdot d_i + \sum_{i=1}^n \deg(i)^2$	[20]
Индекс Балабана	$J(G) = \frac{m}{m-n+2} \sum_{\{i,j\} \in E(G)} (d_i \cdot d_j)^{-1/2}$	[20; 21]
Индекс Балабана модифицированный	$J'(G) = m \sum_{\{i,j\} \in E(G)} (d_i \cdot d_j)^{-1/2}$	[20; 21]
Индекс Рандича	$\chi(G) = \sum_{\{i,j\} \in E(G)} (\deg(i) \cdot \deg(j))^{-1/2}$	[2; 20]
Индекс Харари	$H(G) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (d_{ij})^{-1}$	[2; 20]
Индекс Харари квадратичный	$H_2(G) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (d_{ij})^{-2}$	[2; 20]
Первый Загребский индекс	$M_1(G) = \sum_{i=1}^n \deg(i)^2,$	[2; 20]
Второй Загребский индекс	$M_2(G) = \sum_{\{i,j\} \in E(G)} \deg(i) \cdot \deg(j)$	[2; 20]

Примечание. В табл. 1 величина $d_i = \sum_{j=1}^n d_{ij}$ есть дистанция вершины i в графе G . Остальные обозначения были введены ранее.

Замечание 2. Разные МВП-графы одного порядка имеют одинаковое количество рёбер $m = 2n - 3$. Но при этом для МВП-графов величина $m/(m - n + 2)$ в формуле индекса Балабана J при изменении n изменяется в более узких пределах, чем величина m в формуле модифицированного индекса Балабана J' . Поэтому при построении корреляций структура — свойство для набора изомеров СПУ разных порядков необходимо использовать индекс Балабана J' , а для набора изомеров СПУ одного порядка — индекс J .

Индекс Винера имеет сильные корреляционные связи с физико-химическими показателями некоторых химических соединений. Он характеризуется наглядностью, прямой структурной интерпретацией, простотой вычисления, но имеет низкую дискриминирующую способность [9].

Для преодоления вышеуказанных недостатков в [17] было предложено применять так называемый *3D-индекс Винера* ${}^3D W(G)$, в котором вместо топологических расстояний между парами вершин используются геометрические расстояния, определяемые через декартовы координаты x, y, z вершин графа G

$${}^3D W(G) = \sum_{i,j \in V(G)} L(i, j) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n L(i, j),$$

$$L(i, j) = \sqrt{(X_i - X_j)^2 + (Y_i - Y_j)^2 + (Z_i - Z_j)^2}.$$

Здесь $L(i, j)$ — декартово расстояние между вершинами i и j ; а $X_i, Y_i, Z_i, X_j, Y_j, Z_j$ — декартовы координаты вершин i и j соответственно.

Обычно используют два обозначения для 3D-индекса Винера: ${}^3D W(G)$ для индексов, вычисленных без учёта атомов водорода, и ${}^3D W_H(G)$ для индексов, вычисленных с учётом атомов водорода [17].

В отличие от вышеперечисленных ТИ 3D-индекс Винера отображает геометрическую форму, протяжённость и объём молекулы, поэтому он ещё в большей степени, чем индекс W , пригоден для построения корреляций структура — свойство. При этом 3D-индекс Винера имеет бóльшую дискриминирующую способность по сравнению с индексом Винера W . Отсюда следует, что 3D-индекс Винера возможно будет использовать в качестве эталонного для отбора ТИ, пригодных для построения корреляций структура — свойство неразветвлённых изомеров СПУ.

Единственный очевидный недостаток 3D-индекса Винера заключается в том, что для вычисления этого показателя необходимо знать оптимизированную геометрию каждого изомера СПУ, и это должно быть определено заранее или экспериментально, или квантово-химическими расчётами, прежде чем могут быть определены какие-либо корреляции.

3. Алгоритмы генерации МВП-графов класса «2-цепь»

В [15] был описан алгоритм построения списка рёбер МВП-графа G класса «2-цепь» и его слабо двойственного графа из графа D_4 по заданному кортежу $B(G)$ свойств. В настоящей работе предлагается модифицированный алгоритм, пригодный не только для построения списка рёбер РМВП-графа G класса «2-цепь» по заданному кортежу $B(G)$, но и для получения списков координат концевых вершин всех рёбер РМВП-графа и его слабо двойственного графа. Вычисление вышеперечисленных списков координат вершин осуществляется в подпрограммах `InitCoord` и `ConstructGeometry`.

При модификации алгоритма учитывались следующие основные требования.

1. Вершины создаваемого РМВП-графа G должны располагаться на узлах равносторонней треугольной решётки с горизонтальными рядами [22]. Расположение двух первых граней $\{1, 2, 3\}$, $\{2, 3, 4\}$ графа G должно быть зафиксировано для всех создаваемых РМВП-графов, как это показано на рис. 5 *a*, *б*.

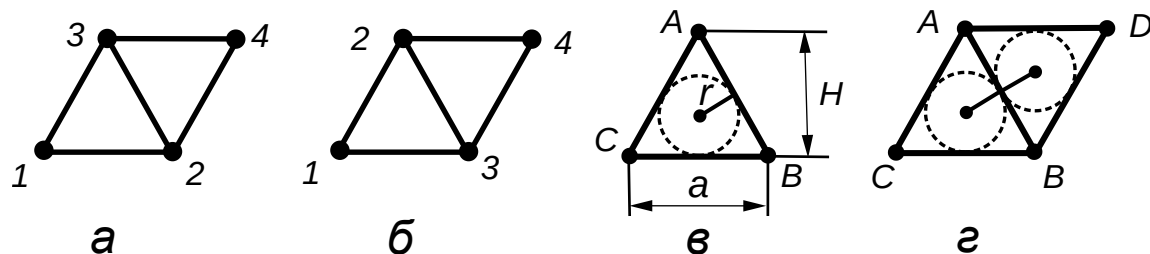


Рис. 5. Фиксированное размещение двух первых граней графа G (*a*, *б*); определение длины L_S ребра слабо двойственного графа по радиусу r вписанной окружности (*в*, *г*)

2. Вершины слабо двойственного графа G^* РМВП-графа G должны располагаться в серединах треугольных граней РМВП-графа (см. рис. 3, *в*, *г*).

3. Основными геометрическими параметрами, используемыми для определения координат вершин РМВП-графа G , являются длина a стороны и высота H треугольника решётки (рис. 5, *в*). В РМВП-графе G величина a есть длина любого его ребра. Величину a и H в графе G целесообразно определять через длину L_S

любого ребра слабо двойственного графа G^* . Далее длину L_S мы будем называть *стандартной длиной рёбер двойственного графа*.

Известно [22, с. 213], что длина L_S равна двум радиусам вписанной окружности (см. рис. 5, ε), т. е. $L_S = 2r$. Но так как $r = a\sqrt{3}/6$, то длина a любого ребра МВП-графа должна быть равна $a = 3L_S/\sqrt{3}$, но поскольку $H = a\sqrt{3}/2$, то высота H должна быть равна $H = 3L_S/2$ (см. рис. 5, ε).

4. В используемой равносторонней треугольной решётке с горизонтальными рядами [22] треугольники направлены вверх и вниз. Поэтому при расчёте координат новой вершины v , смежной с двумя ранее построенными вершинами v_1 и v_2 , следует учитывать, что имеется всего два типа граней (рис. 6). Грани первого типа имеют «остриё», направленное вверх, а грани второго типа — «остриё», направленное вниз. Причём любые две смежные грани являются гранями двух разных типов. Из этого следует, что тип вновь создаваемой грани зависит от типа смежной с ней грани. Если существующая грань есть грань первого типа, то вновь создаваемая грань будет гранью второго типа. И наоборот: если существующая грань есть грань второго типа, то вновь создаваемая грань будет гранью первого типа.

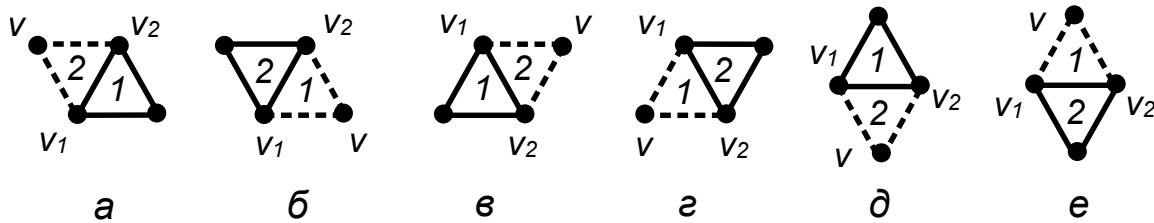


Рис. 6. Варианты расчёта координат новой вершины v ; цифры 1 и 2 показывают тип грани

В алгоритме используются следующие списки и переменные: $B(G)$ — кортеж свойств создаваемого РМВП-графа; ES — список рёбер; $vertFace$ — список номеров вершин граней; TF — список типов граней; v_A и v_B — номера вершин, находящихся на линиях A и B соответственно; w — номер последней (добавляемой) вершины; ng — номер текущей грани; ctf — тип текущей грани. Отметим, что $vertFace(i)$ — список номеров трёх вершин i -й грани, т. е. $vertFace(i) = \{v_{i1}, v_{i2}, v_{i3}\}$, а $TF(i)$ — тип i -й грани, может принимать значение 1 или 2.

В подпрограммах `InitCoord` и `constructGeometry` используются параметры: a — длина ребра РМВП-графа G ; H — высота треугольной грани РМВП-графа; `Coord` — список координат вершин РМВП-графа.

Модифицированный алгоритм построения РМВП-графа G класса «2-цепь» из графа D_4 по кортежу $B(G) = (b_1, b_2, \dots, b_k)$ имеет следующую структуру:

1. Инициализация:

- (a) $ES := ((1, 2), (1, 3), (2, 3), (2, 4), (3, 4)); w := 4;$
- (b) $vertFace = ((1, 2, 3), (2, 3, 4)); TF := (1, 0); ng := 2;$
- (c) если $b_1 = 0$, то $v_A := 4; v_B := 3$, иначе $v_A := 3; v_B := 4;$
- (d) вызвать подпрограмму `InitCoord` для определения начальных координат вершин.

2. Для каждого i -го элемента b_i , ($i = 1, 2, \dots, k$) кортежа свойств B выполнить:

- (a) $w := w + 1; e_1 := (w, v_A)$ и $e_2 := (w, v_B); //$ создать рёбра e_1 и e_2 ;
- (b) $ES := (ES \oplus e_1) \oplus e_2; //$ добавить рёбра e_1 и e_2 в конец списка ES ;
- (c) вызвать подпрограмму `constructGeometry` для пересчёта координат вершин МВП-графа и его слабо двойственного графа;

- (d) $vertFace := vertFace \oplus (w, v_A, v_B)$; // добавить (w, v_A, v_B) в список $vertFace$;
- (e) если $TF[ng] = 1$, то $ctf := 2$, иначе $ctf := 1$; // определить тип следующей грани;
- (f) $TF := TF \oplus ctf$; // добавить тип следующей грани в конец списка TF ;
- (g) $ng := ng + 1$; // задать номер следующей грани;
- (h) если $i \neq k$, то (если $b_i = 1$, то $v_B := w$, иначе $v_A := w$).

В первой части алгоритма (пп. 1 (a)–1 (d)) создаётся исходный список рёбер графа D_4 . Задаются номера v_A, v_B последних вершин, находящихся на линиях A и B , и исходное значение номера $w = 4$ добавляемой вершины, вычисляются исходные значения координат вершин и координат концевых вершин рёбер РМВП-графа и его двойственного графа.

Во второй части (пп. 2 (a)–2 (h)) в цикле на каждом i -м шаге цикла происходит: увеличение номера добавляемой вершины на 1, создание двух рёбер $e_1 := \{w, v_A\}$, $e_2 := \{w, v_B\}$, соединяющих добавляемую вершину w с последними вершинами v_A, v_B , находящимися на линиях A и B (п. 2 (a)); добавление этих рёбер в список SE (п. 2 (b)); добавление вершин новой грани в список $vertFace$ (п. 2 (d)); определение типа новой грани (п. 2 (e)) и добавление его в список TF (п. 2 (f)); задание номера следующей грани (п. 2 (g)); определение номеров вершин v_A, v_B (п. 2 (h)); пересчёт координат вершин и координат концевых вершин рёбер РМВП-графа и его слабо двойственного графа (п. 2 (c)).

Расчёт координат вершин в подпрограмме InitCoord. Согласно п. 1 требований две первые грани любого создаваемого графа G (т. е. две грани исходного графа D_4) зафиксированы (см. рис. 5, а, б). Однако размещение вершин 1, 2, 3, 4 графа зависит от величин v_A, v_B , определяемых по значению b_1 . Если $b_1 = 0$, то $v_A := 4, v_B := 3$ и вершины 2 и 3 располагаются так, как на рис. 5, а. В этом случае координаты вершин 1, 2, 3, 4 определяются следующим образом:

$$x_1 := x_0, \quad x_2 := x_1 + a, \quad x_3 := x_1 + a/2, \quad x_4 := x_2 + a/2, \quad (1)$$

$$y_1 := y_0, \quad y_2 := y_1, \quad y_3 = y_1 + H, \quad y_4 := y_1 + H, \quad (2)$$

если же $b_1 = 1$, то $v_A := 3, v_B := 4$, а вершины 2 и 3 располагаются как на рис. 5, б. В этом случае координаты этих вершин определяются как

$$x_1 := x_0, \quad x_2 := x_1 + a/2, \quad x_3 := x_1 + a, \quad x_4 := x_2 + a, \quad (3)$$

$$y_1 := y_0, \quad y_2 := y_1 + H, \quad y_3 := y_1, \quad y_4 := y_1 + H, \quad (4)$$

где x_i, y_i — координаты x и y i -й вершины.

Расчёт координат вершин в подпрограммах InitCoord и constructGeometry. Согласно п. 2 требований координаты вершин $v_1^*, v_2^*, \dots, v_{n-2}^*$ слабо двойственного графа должны определяться по выражениям

$$x_i^* := (x_{i1} + x_{i2} + x_{i3})/3, \quad y_i^* := (y_{i1} + y_{i2} + y_{i3})/3, \quad (5)$$

где x_i^*, y_i^* — координаты i -й вершины v_i^* ; v_{i1}, v_{i2}, v_{i3} — вершины i -й грани; x_{i1}, x_{i2}, x_{i3} — x -координаты вершин i -й грани; y_{i1}, y_{i2}, y_{i3} — y -координаты вершин i -й грани.

Расчёт координат вершин в подпрограмме constructGeometry. Вычисление координат новой вершины v , добавляемой на каждом шаге алгоритма,

производится с учётом свойств смежных граней, изложенных в п. 4 требований (см. рис. 6). Причём сначала определяются координаты двух вершин v_1, v_2 , с которыми будет соединена новая вершина v . Для этого из списка координат всех вершин `Coord` определяются координаты вершин v_A, v_B и записываются в переменные CA, CB , т. е. $CA = \text{Coord}[v_A]$ и $CB = \text{Coord}[v_B]$. (Отметим, что каждая переменная CB, CA есть список из двух элементов: в первом элементе списка хранится координата x , а во втором элементе списка — координата y .)

Затем определяются координаты x_1, y_1, x_2, y_2 вершин v_1, v_2 согласно правилам:

если $CA[1] < CB[1]$, то $x_1 := CA[1]$; $y_1 := CA[2]$; $x_2 := CB[1]$; $y_2 := CB[2]$;

а если $CA[1] > CB[1]$, то $x_1 := CB[1]$; $y_1 := CB[2]$; $x_2 := CA[1]$; $y_2 := CA[2]$;

Координаты x, y новой вершины v определяются следующими правилами:

если $(y_2 > y_1) \& (x_2 > x_1) \& (TF[ng] = 1)$, то $y := y_2$; $x := x_2 - a$ (рис. 5, а); (6)

если $(y_2 > y_1) \& (x_2 > x_1) \& (TF[ng] = 2)$, то $y := y_1$; $x := x_1 + a$ (рис. 5, б); (7)

если $(y_1 > y_2) \& (x_2 > x_1) \& (TF[ng] = 1)$, то $y := y_1$; $x := x_1 + a$ (рис. 5, в); (8)

если $(y_1 > y_2) \& (x_2 > x_1) \& (TF[ng] = 2)$, то $y := y_2$; $x := x_2 - a$ (рис. 5, г); (9)

если $(y_1 = y_2) \& (x_2 > x_1) \& (TF[ng] = 1)$, то $x := x_1 + \frac{a}{2}$; $y := y_1 - H$ (рис. 5, д); (10)

если $(y_1 = y_2) \& (x_2 > x_1) \& (TF[ng] = 2)$, то $x := x_1 + \frac{a}{2}$; $y := y_1 + H$ (рис. 5, е). (11)

Предложенный модифицированный алгоритм использует базовый алгоритм [15] и подпрограммы для расчёта координат вершин и координат концевых вершин рёбер графов. Таким образом, корректность модифицированного алгоритма обосновывается на корректности базового алгоритма, доказанной в [15], и корректности формул (1)–(5) и (6)–(11), вытекающих из пунктов 1–4 требований и свойств равносторонних треугольников [22, с. 213].

Для реализации алгоритма был разработан программный комплекс в среде Wolfram Mathematica [23], состоящий из четырёх основных функций:

1. `MakeMOPfromHV02` — главная функция, реализующая вышеуказанный алгоритм построения РМВП-графа из исходного графа D_4 по кортежу свойств.
2. `InitCoord` — функция для инициализации координат вершин исходного графа D_4 и его двойственного графа.
3. `constructGeometry` — функция для расчётов координат вершин и координат концевых вершин рёбер создаваемого РМВП-графа G и его слабо двойственного графа G^* .
4. `showGraph` — функция для вывода на экран изображения РМВП-графа G и его слабо двойственного графа G^* .

Комплекс предназначен для расчёта координат вершин и координат концевых вершин рёбер РМВП-графов и их слабо двойственных графов, для вывода на экран изображений этих графов, а также для расчёта 3D-индексов Винера ${}^3D W(G), {}^3D W(G^*)$ и других топологических инвариантов.

Используя формулы (1)–(11), легко доказать следующее утверждение.

Предложение 1. Пусть G — РМВП-граф, G^* — его слабо двойственный граф, имеющий стандартную длину рёбер $L_S \geq 1$, а граф U образован объединением графов G и G^* , т. е. $U = G \cup G^*$. Тогда расстояние $L(v_j, v_i|U)$ между любой парой вершин v_j и v_i графа U пропорционально стандартной длине L_S рёбер слабо двойственного графа G^* , причём $L(v_j, v_i|U) = L_S \cdot L(v_j^0, v_i^0|U_0)$, где U_0 — РМВП-граф,

изоморфный и геометрически подобный графу U , в котором его подграф G_0^* имеет стандартную длину рёбер $L_S^0 = 1$; v_j^0, v_i^0 — вершины U_0 , соответствующие вершинам v_j, v_i графа U ; а $L(v_j^0, v_i^0|U_0)$ — расстояние между вершинами v_j^0, v_i^0 графа U_0 .

Доказательство. Легко видеть, что вершины МВП-графа G , созданного по описанному выше алгоритму, расположены на узлах равносторонней треугольной решётки R (рис. 7, б). В силу формул (5) вершины слабо двойственного графа G^* РМВП-графа G расположены на узлах решётки R_d , находящихся в центрах треугольников решётки R (рис. 7, б). Множество узлов решётки R_d и решётки R образуют множество узлов решётки R_U , на которых располагаются вершины графа U .

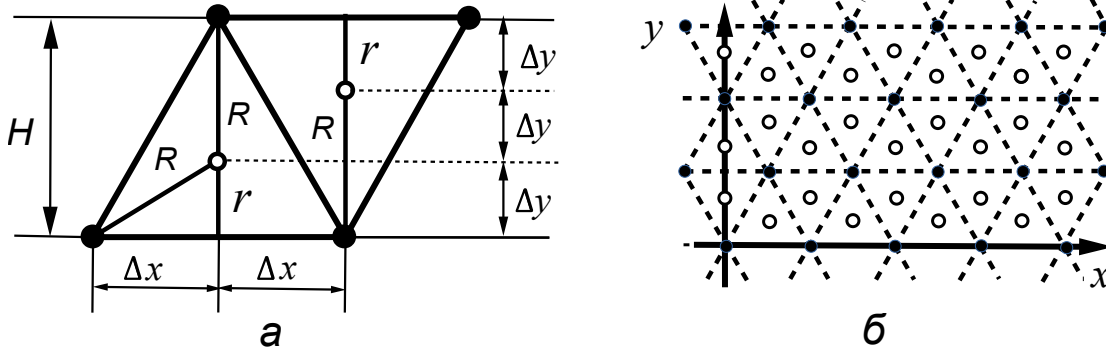


Рис. 7. Равносторонняя треугольная решётка: а — равносторонний треугольник; б — решётка R_U ; чёрными кружками показаны узлы решётки R , пустыми кружками — узлы решётки R_d

Координата x узлов решётки R_U изменяется с шагом Δx , а координата y изменяется с шагом Δy . Причём $\Delta x = a/2$ а $\Delta y = H/3$ (рис. 7, а, б). Из свойств равностороннего треугольника следует, что $H = a\sqrt{3}/2$, а радиус вписанной окружности $r = a\sqrt{3}/6$.

Так как $L_S = 2r = a\sqrt{3}/3$, то $a = 3L_S/\sqrt{3}$ и, следовательно, $H = 3L_S/2$. Откуда для Δx и Δy будем иметь $\Delta x = 3L_S/(2\sqrt{3})$ и $\Delta y = L_S/2$. Или, приняв новые обозначения $C_x = 3/(2\sqrt{3})$ и $C_y = 1/2$, получим $\Delta x = C_x \cdot L_S$ и $\Delta y = C_y \cdot L_S$.

Зададим систему координат на решётке R_U . Для этого в качестве начала системы координат выберем узел решётки R и обозначим его через v_0 . Проведём через узел v_0 ось абсцисс OX по образующей линии, проходящей через узел v_0 , а также — перпендикулярную ей ось ординат OY (рис. 7, б).

Тогда для координат x, y любой вершины v_k графа U получим $x_k = \Delta x \cdot N_k^x$ и $y_k = \Delta y \cdot N_k^y$, где N_k^x и N_k^y есть целые числа. Расстояние $L(v_j, v_i|U)$ между вершинами v_i, v_j графа U будет определяться по выражению

$$L(v_j, v_i|U) = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} = \sqrt{(\Delta x \cdot N_{j,i}^x)^2 + (\Delta y \cdot N_{j,i}^y)^2},$$

где $N_{j,i}^x = (N_j^x - N_i^x)$, $N_{j,i}^y = (N_j^y - N_i^y)$. Откуда, подставив выражения для Δx и Δy , получим $L(v_j, v_i|U) = L_S \sqrt{(C_x \cdot N_{j,i}^x)^2 + (C_y \cdot N_{j,i}^y)^2} = L_S \cdot \Delta L_{j,i}^0$, где величина $\Delta L_{j,i}^0 = \sqrt{(C_x \cdot N_{j,i}^x)^2 + (C_y \cdot N_{j,i}^y)^2}$ не зависит от величины L_S .

Но тогда, если в графе U уменьшить координаты всех вершин в q раз, то получится граф U' , подобный и изоморфный графу U , в котором стандартная длина L'_S рёбер слабо двойственного графа $(G')^*$ будет равна $L'_S = L_S/q$. Причём вершины графа U' будут уже располагаться на решётке R'_U с меньшими шагами решётки $\Delta x', \Delta y'$, поскольку $\Delta x' = C_x \cdot L_S/q$ и $\Delta y' = C_y \cdot L_S/q$. Тогда расстояние $L(v'_j, v'_i|U')$ между вершинами v'_i, v'_j графа U' будет определяться по выражению $L(v'_j, v'_i|U') = L'_S \Delta L_{j,i}^0$.

А если в графе U уменьшить координаты всех вершин в $q = L_S$ раз, то получится граф U_0 , подобный и изоморфный графу U , в котором стандартная длина L_S^0 рёбер слабо двойственного графа G_0^* будет равна 1. Следовательно, расстояние между парой его вершин v_j^0, v_i^0 , соответствующих паре вершин v_j, v_i графа U , будет определяться по формуле $L(v_j^0, v_i^0 | U_0) = \Delta L_{j,i}^0$. \square

Следствие 1. Пусть G — РМВП-граф, G^* — его слабо двойственный граф со стандартной длиной рёбер $L_S \geq 1$ и $U = G \cup G^*$. Тогда 3D-индексы Винера ${}^{3D}W(G^*)$ и ${}^{3D}W_H(U)$ пропорциональны стандартной длине L_S рёбер графа G^* , причём ${}^{3D}W(G^*) = L_S \cdot {}^{3D}W(G_0^*)$ и ${}^{3D}W_H(U) = L_S \cdot {}^{3D}W_H(U_0)$, где G_0 — РМВП-граф, изоморфный и геометрически подобный графу G , а G_0^* — слабо двойственный граф графа G_0 , со стандартной длиной рёбер $L_S^0 = 1$, изоморфный и геометрически подобный графу G^* , и $U_0 = G_0 \cup G_0^*$.

4. Методика отбора ТИ РМВП-графов, пригодных для определения корреляций структура — свойство изомеров СПУ

В основе новой методики лежит использование 3D-индекса Винера и двух видов плоских молекулярных графов СПУ, получаемых из РМВП-графов $G_{n,1}, G_{n,2}, \dots, G_{n,k_n}$ порядка n , где $G_{n,i}$ — i -й РМВП-граф порядка n , а k_n — число графов порядка n . В качестве первого вида молекулярных графов применяются слабо двойственные графы $G_{n,1}^*, G_{n,2}^*, \dots, G_{n,k_n}^*$ РМВП-графов $G_{n,1}, G_{n,2}, \dots, G_{n,k_n}$ порядка n . А в качестве молекулярных графов второго вида — графы $U_{n,1}, U_{n,2}, \dots, U_{n,k_n}$, получаемые объединением графов $G_{n,i}$ и $G_{n,i}^*$, т. е. $U_{n,i} = G_{n,i} \cup G_{n,i}^*$.

Фактически эта методика есть доказательство пригодности ТИ РМВП-графов $G_{n,1}, G_{n,2}, \dots, G_{n,k_n}$ класса «изомеры» и 3D-индекса Винера соответствующих слабо двойственных графов $G_{n,1}^*, G_{n,2}^*, \dots, G_{n,k_n}^*$ для прогнозирования физико-химических свойств изомеров СПУ. Действительно, одним из критериев отбора ТИ РМВП-графов является наличие высокой корреляции между ним и 3D-индексом Винера слабо двойственных графов. При этом сам 3D-индекс Винера слабо двойственных графов проверяется по критерию наличия высокой корреляции между ним и 3D-индексом Винера молекулярных графов $M_{n,1}, M_{n,2}, \dots, M_{n,k_n}$, полученных с помощью квантово-химических расчётов. Но тогда при выполнении двух вышеуказанных критериев ТИ РМВП-графов будет иметь высокую корреляцию и с 3D-индексом Винера молекулярных графов $M_{n,1}, M_{n,2}, \dots, M_{n,k_n}$.

Любой РМВП-граф $G_{n,i}$ является родителем для графов $G_{n,i}^*, U_{n,i}, M_{n,i}$, поэтому для всех графов $G_{n,i}^*, U_{n,i}, M_{n,i}$ пара индексов n, i есть лишь метка, указывающая на i -й родительский РМВП-граф $G_{n,i}$ порядка n .

Список этапов методики отбора ТИ РМВП-графов:

1. Создание РМВП-графов и плоских молекулярных графов изомеров СПУ без атомов водорода и с атомами водорода.
 - 1.a Создать кортежи свойств РМВП-графов порядка n ($n = 8, \dots, 28$) класса «изомеры».
 - 1.b По кортежам свойств создать РМВП-графы порядка n ($n = 8, \dots, 28$) класса «изомеры» $G_n = \{G_{n,1}, G_{n,2}, \dots, G_{n,k_n}\}$.
 - 1.c Для РМВП-графов создать их слабо двойственные графы $G_n^* = \{G_{n,1}^*, G_{n,2}^*, \dots, G_{n,k_n}^*\}$ и графы $U_n = \{U_{n,1}, U_{n,2}, \dots, U_{n,k_n}\}$, где граф $U_{n,j}$

есть объединение РМВП-графа $G_{n,j}$ и его двойственного графа $G_{n,j}^*$, т. е. $U_{n,j} = G_{n,j} \cup G_{n,j}^*$. (Здесь подграф $G_{n,j}^*$ графа $U_{n,j}$ моделирует углерод-углеродную цепочку, а вершины графа $G_{n,j}$ — атомы водорода).

2. Оценка дискриминирующей способности эталонных ТИ.
 - 2.a Для графов $G_{n,j}^*$ и $U_{n,j}$ определить значения 3D-индексов Винера ${}^3W(G_{n,j}^*), {}^3W(U_{n,j})$ ($n = 8, \dots, 28, j = 1, \dots, k_n$).
 - 2.b Для каждого n определить значения дискриминирующей способности $S[{}^3DW, G_n^*]$ индекса 3DW на множестве G_n^* слабо двойственных графов.
 - 2.c Для каждого n определить значения дискриминирующей способности $S[{}^3W, U_n]$ индекса 3DW на множестве U_n графов.
3. Проверка возможности использования эталонного топологического индекса.
 - 3.a Для графов $G_{n,j}^*$ и $U_{n,j}$ рассчитать 3D-индексы Винера ${}^3DW(G_{n,j}^*)$ и ${}^3DW_H(U_{n,j})$ ($n = 8, \dots, 28, j = 1, \dots, k_n$).
 - 3.b Для каждого n ($n = 8, \dots, 28$) создать переменную ${}^3DW(G_n^*)$, принимающую значения ${}^3DW(G_{n,j}^*)$, и переменную ${}^3DW_H(U_n)$, принимающую значения ${}^3DW_H(U_{n,j})$ ($j = 1, \dots, k_n$).
 - 3.c По графам $G_{n,j}^*$ ($n = 8, \dots, 28, j = 1, \dots, k_n$) получить соответствующие им реальные 3D-молекулярные графы $M_n = \{M_{n,1}, M_{n,2}, \dots, M_{n,k_n}\}$ ($n = 8, \dots, 28$) с атомами водорода.
 - 3.d Для графов $M_{n,j}$ рассчитать 3D-индексы Винера ${}^3DW(M_{n,j})$ и ${}^3DW_H(M_{n,j})$ соответственно без учёта и с учётом атомов водорода.
 - 3.e Для каждого n ($n = 8, \dots, 28$) создать переменную ${}^3DW(M_n)$, принимающую значения ${}^3DW(M_{n,i})$, и переменную ${}^3DW_H(M_n)$, принимающую значения ${}^3DW_H(M_{n,i})$ ($i = 1, \dots, k_n$).
 - 3.f Для каждого n ($n = 8, \dots, 28$) рассчитать коэффициенты корреляции между переменной ${}^3DW(G_n^*)$ с одной стороны и переменными ${}^3DW(M_n)$ и ${}^3DW_H(M_n)$ с другой. В случае наличия высокой корреляции между ${}^3DW(G_n^*)$ и ${}^3DW(M_n)$ или между ${}^3DW(G_n^*)$ и ${}^3DW_H(M_n)$ переменные ${}^3DW(G_n^*)$ можно использовать в качестве эталонных ТИ.
 - 3.g Для каждого n ($n = 8, \dots, 28$) рассчитать коэффициент корреляции между переменными ${}^3DW_H(U_n)$ и ${}^3DW_H(M_n)$. В случае наличия высокой корреляции между этими переменными ${}^3DW_H(U_n)$ можно будет использовать в качестве эталонных ТИ и для прогнозирования физико-химических свойств изомеров СПУ.
4. Исследование и отбор топологических индексов РМВП-графов.
 - 4.a Для РМВП-графов $G_{n,j}$ ($n = 8, \dots, 28, j = 1, \dots, k_n$) рассчитать значения ТИ $I_1(G_{n,j}), I_2(G_{n,j}), \dots, I_{k_I}(G_{n,j})$.
 - 4.b Для каждого $i = 1, \dots, k_I$ определить дискриминирующую способность $S[I_i, G_n]$ индекса I_i на множестве всех n -вершинных РМВП-графов класса «изомеры».
 - 4.c Для каждого $i = 1, \dots, k_I$ определить дискриминирующую способность $S[I_i, G]$ индекса I_i на множестве G всех РМВП-графов класса «изомеры».
 - 4.d Для каждого индекса I_i и для каждого n создать переменную $I_i(G_n)$, принимающую значения $I_i(G_{n,1}), I_i(G_{n,2}), \dots, I_i(G_{n,k_n})$.

- 4.e Для каждого индекса I_i и для каждого n рассчитать коэффициент корреляции между переменными $I_i(G_n)$ и ${}^{3D}W(G_n^*)$.
- 4.f Выполнить отбор ТИ по критерию наибольшей дискриминирующей способности и наибольшему коэффициенту корреляции.

Примечание к пункту 3.c: Здесь под реальными 3D-молекулярными графами подразумеваются молекулярные графы, построенные по координатам атомов и геометрическим параметрам химических связей в молекуле моделируемого химического соединения. К геометрическим параметрам связей относят длины валентных связей, валентные углы и торсионные углы [6]. (Евклидовы координаты x, y, z атомов и длины валентных связей определяются в ангстремах или в борах, а все углы — в градусах.) Напомним, что длина валентной связи — это расстояние между ядрами двух атомов, химически связанных друг с другом. Валентный угол — это угол между двумя линиями связи, имеющими общий атом. Линия связи — это линия, соединяющая ядра двух химически связанных атомов. Торсионный угол, определяемый на цепочке атомов А-В-С-Д, это двугранный угол между двумя плоскостями проходящими через атомы А, В, С и через атомы В, С, Д [6].

Геометрические параметры химических связей молекул могут быть определены экспериментально, например, с помощью методов дифракционного структурного анализа, МВ- или ИК-молекулярной спектроскопии. Для относительно небольших по числу атомов молекул геометрические параметры могут быть определены с помощью квантово-химических расчётов [24] и методами молекулярной механики (см. [24, с. 48–83]).

По каждому пункту методики был разработан соответствующий алгоритм, реализованный в системе Wolfram Mathematica [23].

5. Исследование и отбор топологических индексов РМВП-графов, пригодных для построения корреляций структура — свойство неразветвлённых изомеров СПУ

В разделе приведены результаты расчётных экспериментов, выполненных для проверки возможности применения ТИ РМВП-графов для прогнозирования физико-химических свойств неразветвлённых изомеров СПУ по методике, изложенной в разделе 4.

5.1. Создание РМВП-графов и плоских молекулярных графов изомеров СПУ без атомов водорода и с атомами водорода

Создание РМВП-графов и плоских молекулярных графов изомеров СПУ выполнялось в соответствии с п. 1 методики. Генерировались РМВП-графы класса «изомеры» порядка n ($8 \leq n \leq 28$) и класса «конформеры» для дополнительных исследований, а также все остальные графы, описанные в пункте 1.с методики. Генерация РМВП-графов производилась с помощью программных комплексов, описанных в разделах 1 и 3 соответственно.

5.2. Определение дискриминирующей способности эталонного топологического индекса

Расчётные эксперименты по определению дискриминирующей способности 3D-индексов Винера ${}^{3D}W(G^*)$ и ${}^{3D}W_H(U)$ выполнялись в соответствии с п. 2 методики. По результатам расчётного эксперимента была выявлена высокая дискриминиру-

ющая способность индексов ${}^3D W(G^*)$ и ${}^3D W_H(U)$. Так, например, дискриминирующая способность индекса ${}^3D W_H(U)$ во всём диапазоне изменения порядка РМВП-графов от $n = 8$ до $n = 28$ равна 1. Дискриминирующая способность индекса ${}^3D W(G^*)$ при $n = 8, 10, 12, 14, 16, 20, 24, 28$ равна 1. А при $n = 22, 26$ она равна 0.9076 и 0.9867 соответственно.

Таким образом, индекс ${}^3D W(G^*)$ можно использовать в качестве эталонного и вместе с индексом ${}^3D W_H(U)$ его можно применять для построения корреляций структура — свойство неразветвлённых изомеров СПУ.

5.3. Проверка возможности использования эталонного топологического индекса

Проверка возможности использования эталонного топологического индекса ${}^3D W(G^*)$ плоского молекулярного графа G^* выполнялась по пунктам 3.a–3.g вышеизложенной методики. При этом получение молекулярных графов посредством компьютерных расчётов осуществлялось в три этапа.

На первом этапе с помощью программы, описанной в разделе 3, были сгенерированы ступенчатые и квазиступенчатые РМВП-графы класса «изомеры» и «конформеры» порядка 8, 10, 12 и их слабые двойственные графы порядка 6, 8, 10 соответственно. Эти слабые двойственные графы принимались в качестве плоских молекулярных графов изомеров и конформеров молекул СПУ C_6H_8 , C_8H_{10} , $C_{10}H_{12}$. Первоначально все МВП-графы и их слабые двойственные графы создавались в виде графов и списков координат в Wolfram Mathematica. Затем структура и координаты двойственных графов записывались в файлы формата `cm1` (по одному файлу на граф) и передавались на второй этап.

На втором этапе с помощью программы молекулярного моделирования Avogadro [25] к молекулярным графам, полученным на первом этапе, добавлялись атомы водорода. После чего молекулярные структуры оптимизировались методом молекулярной механики с использованием силовых полей MMFF94 [26]. Все изменения структуры и координат, возникающие в процессе расчёта, сохранялись в файлах формата `cm1` и передавались в качестве исходных данных на третий этап.

На третьем этапе была проведена оптимизация структур молекул с помощью квантово-химических расчётов. Оптимизация структуры молекул C_6H_8 , C_8H_{10} была проведена автором в приближении B3LYP-D3/6-31G с использованием программного пакета Firefly [27], который частично основан на исходном коде GAMESS (US) [28].

Оптимизация структур молекул $C_{10}H_{12}$ с последующим вычислением гармонических частот была проведена в приближении B3LYP-D3BJ/def2-TZVPP [29; 30; 31] с использованием программного пакета Orca 4 [32]. Расчёты проводились к.х.н Д. С. Тихоновым с использованием вычислительного кластера Maxwell научного центра DESY (Гамбург, Германия). Координаты атомов, полученные на третьем этапе, записывались в файлы формата XYZ и использовались для статистической обработки. Извлечение и визуализация результатов квантово-химических расчётов выполнялись с помощью программы Chemcraft [33].

Геометрические параметры изомеров, полученные с помощью квантово-химических расчётов, представлены в табл. 2. С учётом наличия всего двух изомеров C_6H_8 все изомеры и конформеры были выделены в две группы — в одну группу изомеры (конформеры) C_6H_8 и C_8H_{10} , во вторую группу — изомеры (конформеры) $C_{10}H_{12}$.

Таблица 2

Геометрические параметры изомеров C_6H_8 , C_8H_{10} и $C_{10}H_{12}$

Геометрический параметр	Локализация	C_6H_8 и C_8H_{10}			$C_{10}H_{12}$		
		Min	Max	Mean	Min	Max	Mean
$R_O, \text{Å}$	все	1.3434	1.4528	1.3995	1.3530	1.4645	1.4050
$R(C = C), \text{Å}$	терм.	1.3434	1.3443	1.3439	1.3367	1.3679	1.3471
$R(C = C), \text{Å}$	центр.	1.3520	1.3595	1.3563	1.3650	1.3729	1.3685
$R(C - C), \text{Å}$	все	1.4413	1.4528	1.4492	1.4514	1.4645	1.4582
$R(C - H), \text{Å}$	все	1.0859	1.0913	1.0887	1.0945	1.1011	1.0979
$A_{CCC}, \text{град.}$	транс.	123.53	124.89	124.14	123.58	124.50	124.20
$A_{CCC}, \text{град.}$	цис.	126.08	127.37	127.06	126.08	127.07	126.57
торсионный угол, град.	$\langle 0 \rangle$	0.	0.	0.	-0.0004	0.0006	0.00001
торсионный угол, град.	$\langle 180 \rangle$	-180.	-180.	-180.	-179.997	180.	27.27

В табл. 2 приняты следующие обозначения: $R(C - C)$, $R(C - H)$, $R(C = C)$ — длины простых и двойных валентных связей; A_{CCC} — угол между двумя связями трёх атомов углерода. В колонке «локализация» уточняется или вид параметра, или диапазон изменения его величины. Так, например, параметр $R(C = C)$ с локализацией «центр.» («терм.») есть длина центральной (терминальной) двойной связи. Параметр A_{CCC} с локализацией «транс.» («цис.») есть угол между двумя связями, образующими соответственно транс- и цис-конфигурации (конформации). Для торсионных углов обозначение $\langle 0 \rangle$ показывает, что торсионный угол изменяется от 0° до $\pm 90^\circ$, а обозначение $\langle 180 \rangle$ показывает, что торсионный угол изменяется от 180° до $\pm 90^\circ$.

Полученные результаты квантово-химических расчётов хорошо согласуются с многочисленными работами по исследованию структуры изомеров СПУ (см., например, [7; 8]). Поэтому их можно использовать для решения поставленных задач.

На заключительном этапе, согласно пунктам 3.d–3.g методики, определялось наличие корреляции между вышеуказанными парами 3D-индексов Винера.

В табл. 3 приведены значения коэффициентов корреляции $R_{(3DW(G^*)/3DW(M))}$, $R_{(3DW(G^*)/3DW_H(M))}$ и $R_{(3DW_H(U)/3DW_H(M))}$ для разных классов РМВП-графов.

Таблица 3

Коэффициенты корреляции

Расчётный коэффициент корреляции	Изомеры C_6H_8 , C_8H_{10} , $C_{10}H_{12}$	Изомеры $C_{10}H_{12}$	Конформеры транс-изомеров $C_{10}H_{12}$	Конформеры других изомеров $C_{10}H_{12}$
$R_{(3DW(G^*)/3DW(M))}$	0.99970	0.99801	0.94198	0.96647
$R_{(3DW(G^*)/3DW_H(M))}$	0.99945	0.99432	0.90045	0.96238
$R_{(3DW_H(U)/3DW_H(M))}$	0.99973	0.99781	0.91467	0.94538

Примечание: В табл. 3 во втором столбце коэффициенты корреляции были определены по объединённым данным всех изомеров C_6H_8 , C_8H_{10} и $C_{10}H_{12}$. Коэффициенты корреляции отдельно для изомеров C_6H_8 и C_8H_{10} не определялись из-за малого количества этих изомеров.

Полученные результаты (табл. 2, 3) позволяют сделать следующие выводы:

1. Молекулярные графы изомеров СПУ практически являются плоскими, поскольку значения их торсионных углов очень мало отличаются от 0° или от 180° .

2. Слабо двойственные графы РМВП-графов не имеют точного геометрического подобия с реальными молекулярными графами СПУ, что объясняется в первую очередь большими значениями валентных углов (все углы больше 120°) и переменными значениями длин связей. Но при этом слабо двойственные графы РМВП-графов схожи по форме с реальными молекулярными графами СПУ и отличаются

от них только по величинам углов и по длинам рёбер.

3. Из табл. 3 видно, что для всех изомеров и конформеров $C_{10}H_{12}$ коэффициенты корреляции имеют очень большие значения. Отсюда следует, что 3D-индексы Винера плоских молекулярных графов, являющихся слабыми двойственными графами решётчатых МВП-графов, пригодны для использования в качестве эталонных.

Отметим, что в табл. 3 расчёт 3D-индексов Винера ${}^3DW(G^*)$, ${}^3DW_H(U)$ графов G^*, U производился для стандартной длины ребра $L_S = 1$. Однако в нашем случае величина стандартной длины ребра двойственных графов не влияет на величины коэффициентов корреляции $R_{({}^3DW(G^*)/{}^3DW(M))}$, $R_{({}^3DW(G^*)/{}^3DW_H(M))}$ и $R_{({}^3DW_H(U)/{}^3DW_H(M))}$.

Действительно, с помощью определения коэффициента корреляции и свойств линейной парной регрессии [6, с. 31–36] легко доказать, что коэффициент корреляции $R'_{X'/Y}$ между переменными X', Y , принимающими значения x'_1, x'_2, \dots, x'_n (где $x'_i = q \cdot x_i$) и y_1, y_2, \dots, y_n , равен коэффициенту корреляции $R_{X/Y}$ между переменными X, Y , принимающими значения x_1, x_2, \dots, x_n и y_1, y_2, \dots, y_n , т. е. $R'_{X'/Y} = R_{X/Y}$. Причём для предикторных значений \hat{Y}', \hat{Y} переменной Y , коэффициентов a, b и a', b' соответствующих уравнений регрессии $\hat{Y} = a + b \cdot X$ и $\hat{Y}' = a' + b' \cdot X'$ легко доказать равенства $a' = a$, $b' = b/q$ и $\hat{Y}' = \hat{Y}$. Отсюда и в силу следствия 1 вытекает справедливость следующего утверждения.

Утверждение 2. Пусть Y — переменная, принимающая значения Y_1, \dots, Y_{k_n} и корреляционно связанная с переменной X , принимающей значения $X_i = {}^3DW(G_{n,i}^*)$ на слабо двойственных графах $G_{n,i}^*$ РМВП-графов $G_{n,i}$ ($i = 1, \dots, k_n$) порядка n , имеющих стандартную длину рёбер $L_S(G_{n,i}^*) = l_S$. И пусть для переменных X, Y выполняется линейное уравнение регрессии $\hat{Y} = a + b \cdot X$ с коэффициентом корреляции $R_{X/Y}$, с величинами скорректированного коэффициента детерминации R_{ck}^2 , F -статистики — F и стандартной ошибки — s .

Тогда для любого другого значения стандартной длины рёбер $L_S(G_{n,i}^*) = l'_S = q \cdot l_S$ ($q > 0$) выполняется линейное уравнение регрессии $\hat{Y}' = a' + b' \cdot X'$, где $\hat{Y}' = \hat{Y}$, $a' = a$, $b' = b/q$ и $X'_i = q \cdot X_i$. Причём для $R'_{X'/Y}$ и параметров $(R_{ck}^2)', F', s'$, выполняются равенства $R'_{X'/Y} = R_{X/Y}$, $(R_{ck}^2)' = R_{ck}^2$, $F' = F$, $s' = s$.

Поэтому для построения корреляций структура — свойство можно использовать плоскую модель молекулярного графа изомеров СПУ, в которой углы между любыми двумя смежными рёбрами равны 120° и все рёбра имеют длину, равную 1.

5.4. Отбор топологических индексов РМВП-графов

Результаты эксперимента по определению дискриминирующей способности топологических индексов Балабана J, J' , Винера W , Рандича χ , Шульца MTI , Харари H, H_2 , Загребских индексов M_1, M_2 РМВП-графов класса «изомеры» представлены в табл. 4.

Результаты эксперимента по определению коэффициента корреляции между 3D-индексом Винера 3DW РМВП-графов класса «изомеры» и топологическими индексами Балабана J, J' , Винера W , Рандича χ , Шульца MTI , Харари H, H_2 , и Загребскими индексами M_1, M_2 представлены в табл. 5. Индекс M_1 не вошёл в табл. 5 из-за самой низкой дискриминирующей способности.

Замечание 3. Поскольку значения индексов J и J' отличаются друг от друга только постоянным коэффициентом, то для коэффициентов корреляции $R_{({}^3DW(G^*)/J)}$ и $R_{({}^3DW(G^*)/J')}$ и для дискриминирующих способностей $S[J]$ и $S[J']$ выполняются ра-

венства $R_{(3DW(G^*)/J)} = R_{(3DW(G^*)/J')}$ и $S[J] = S[J']$. Поэтому в табл. 4 и 5 приведены значения только $R_{(3DW(G^*)/J)}$ и $S[J]$ соответственно.

Таблица 4

Дискриминирующая способность топологических индексов
 $J, J', W, \chi, MTI, H, H_2, M_1, M_2$ РМВП-графов класса «изомеры»

n	k_n	$S[J]$	$S[W]$	$S[\chi]$	$S[MTI]$	$S[H]$	$S[H_2]$	$S[M_1]$	$S[M_2]$
8	2	1.	1	1.	1.	1.	1.	1.	1.
10	3	1.	1.	1.	1.	1.	1.	1.	1.
12	6	1.	0.833	1.	0.833	0.833	0.833	0.667	0.833
14	10	1.	0.7	1.	0.7	0.7	0.7	0.5	0.7
16	20	1.	0.45	0.8	0.55	0.55	0.55	0.3	0.5
18	36	1.	0.361	0.639	0.417	0.417	0.417	0.194	0.361
20	72	1.	0.25	0.444	0.306	0.292	0.306	0.111	0.236
22	136	0.993	0.154	0.309	0.213	0.213	0.221	0.066	0.154
24	272	0.993	0.099	0.199	0.147	0.143	0.154	0.037	0.096
26	528	0.977	0.064	0.127	0.100	0.097	0.106	0.021	0.059
28	1056	0.977	0.037	0.078	0.066	0.065	0.073	0.011	0.035

Примечание: В табл. 4 приняты следующие обозначения: n — порядок РМВП-графа; k_n — число неизоморфных РМВП-графов порядка n . Другие обозначения, были приняты в соответствии с разделом 2.

Таблица 5

Коэффициенты корреляции между $3DW(G^*)$ и топологическими индексами
 $J, J', W, \chi, MTI, H, H_2, M_2$ РМВП-графов класса «изомеры»

n	$R_{(3DW/J)}$	$R_{(3DW/W)}$	$R_{(3DW/\chi)}$	$R_{(3DW/MTI)}$	$R_{(3DW/H)}$	$R_{(3DW/H_2)}$	$R_{(3DW/M_2)}$
8	1.	1.	1.	1.	-1.	-1.	-1.
10	-0.998	0.999	0.961	0.9999	-0.994	-0.991	-0.986
12	-0.993	0.995	0.873	0.997	-0.977	-0.965	-0.943
14	-0.997	0.989	0.857	0.9896	-0.971	-0.956	-0.927
16	-0.979	0.979	0.784	0.981	-0.945	-0.920	-0.874
18	-0.966	0.966	0.762	0.968	-0.931	-0.904	-0.853
20	-0.956	0.953	0.718	0.956	-0.906	-0.873	-0.814
22	-0.941	0.938	0.695	0.940	-0.888	-0.853	-0.792
24	-0.929	0.925	0.667	0.928	-0.867	-0.828	-0.764
26	-0.914	0.9095	0.646	0.913	-0.849	-0.809	-0.743
28	-0.902	0.896	0.625	0.8996	-0.8298	-0.789	-0.721

Примечание: В табл. 5 приняты следующие обозначения: n — порядок РМВП-графа; $R_{(3DW/J)}$, $R_{(3DW/W)}$, $R_{(3DW/\chi)}$, $R_{(3DW/MTI)}$, $R_{(3DW/H)}$, $R_{(3DW/H_2)}$, $R_{(3DW/M_2)}$ — коэффициенты корреляции между 3D-индексом Винера $3DW(G^*)$ с одной стороны и индексами $J, W, \chi, MTI, H, H_2, M_2$ с другой.

Из табл. 4, 5 видно, что по критерию коэффициента корреляции наиболее подходящими топологическими индексами являются индексы Балабана, Шульца, Винера и Харари (в порядке уменьшения коэффициента корреляции). Однако дискриминирующая способность последних трёх индексов резко уменьшается при увеличении порядка РМВП-графов начиная с $n = 12$. Поэтому их применение возможно только для РМВП-графов порядка $8 \leq n \leq 10$.

Единственными топологическими индексами, имеющими высокие значения коэффициента корреляции и дискриминирующей способности во всём диапазоне изменения n , являются индексы Балабана J, J' .

6. Пример построения корреляции структура — свойство для изомеров и конформеров гексатриена

В данном разделе в качестве примера использования предложенного подхода для построения корреляций структура — свойство неразветвлённых изомеров СПУ представлены полученные автором уравнения линейной регрессии нормированной энтальпии образования изомеров и конформеров гексатриена C_6H_8 от индексов Балабана РМВП-графов. На рис. 3 представлены РМВП-графы порядка 8, слабо двойственные графы (даны пунктиром) которых изоморфны и геометрически подобны молекулярным графам изомеров и конформеров гексатриена C_6H_8 .

Значения энтальпии образования и ttt-кодировка изомеров и конформеров гексатриена были взяты из работы [34] и вместе с результатами расчётов 3D-индекса Винера и индексов Балабана $J(G)$, $J'(G)$ представлены в табл. 6. Здесь через G обозначены РМВП-графы порядка 8, а через G^* — слабые двойственные графы графов G , являющиеся простыми цепями P_6 порядка 6.

Данные в четвёртом столбце таблицы были получены из данных в третьем столбце путём вычитания наименьшего значения (энтальпии) и умножения на 1000.

Таблица 6

Энтальпия образования (в кДж моль⁻¹) для изомеров и конформеров C_6H_8 и топологические индексы ${}^3DW(G^*)$, J , J'

Изомер (конформер)	Рисунок	$\Delta_f H$	$\Delta_f H_{nor}$	${}^3DW(G^*)$	$J(G)$	$J'(G)$
ttt	3а	-233.284653	0.000	31.1526	2.0678	14.4752
tct	3б	-233.281451	3.202	29.2197	2.1594	15.1155
ttc	3в	-223.279004	5.649	29.2894	2.1263	14.8844
ctc	3г	-223.272950	11.703	28.1795	2.2132	15.4920
cct	3д	-233.275413	9.240	25.9518	2.2802	15.9616
ccc	3е	-233.265688	18.965	22.3933	2.4621	17.2346

Ниже приведены полученные уравнения линейной регрессии вместе с соответствующими значениями коэффициента корреляции R , скорректированного коэффициента детерминации R_{ck}^2 , F -статистики F и стандартной ошибки s .

1. Для индекса Винера ${}^3DW(G^*)$ и нормированной энтальпии $\Delta_f H_{nor}$ образования:

$$\Delta_f H_{nor} = 63.9872 - 2.0168 \cdot ({}^3DW(G^*)),$$

$$R_{({}^3DW(G^*)/\Delta_f H_{nor})} = -0.9276, R_{ck}^2 = 0.8255, F = 24.6505, s = 2.8206.$$

2. Для индекса Балабана J и нормированной энтальпии $\Delta_f H_{nor}$ образования:

$$\Delta_f H_{nor} = -94.2061 + 46.1337 \cdot J(g),$$

$$R_{(J(G)/\Delta_f H_{nor})} = 0.9558, R_{ck}^2 = 0.8921, F = 42.3177, s = 2.2184.$$

3. Для индекса Балабана J' и нормированной энтальпии $\Delta_f H_{nor}$ образования:

$$\Delta_f H_{nor} = -94.2061 + 6.5905 \cdot J'(g),$$

$$R_{(J'(G)/\Delta_f H_{nor})} = 0.9558, R_{ck}^2 = 0.8921, F = 42.3177, s = 2.2184.$$

Легко видеть, что коэффициенты Балабана J , J' РМВП-графов класса «изомеры» и 3D-индекс Винера их слабо двойственных графов хорошо коррелируют с нормированной энтальпией образования изомеров и конформеров гексатриена.

Заключение

Представлена и апробирована методика отбора топологических индексов (ТИ) РМВП-графов, пригодных для прогнозирования физико-химических свойств изомеров СПУ. Предложенная методика значительно сокращает объём вычислитель-

ных работ по определению корреляционных зависимостей структура — свойство неразветвлённых изомеров СПУ.

По результатам исследования ТИ РМВП-графов класса «изомеры» порядков $n = 8, 10, \dots, 28$ отобраны индекс Балабана J и модифицированный индекс Балабана J' .

Разработан и реализован в системе Wolfram Mathematica [23] модифицированный алгоритм создания РМВП-графов по кортежам их свойств. Кортежи свойств РМВП-графов разных классов генерировались с помощью программ, описанных в [15]. К сожалению, все эти программы генерируют большое количество кортежей свойств с последующей их проверкой на изоморфизм, что увеличивает длительность их работы.

Показано, что 3D-индексы Винера ${}^{3D}W(G^*)$ и ${}^{3D}W_H(U)$ плоских молекулярных графов G^*, U , полученных из РМВП-графов G , обладают высокой дискриминирующей способностью и хорошо коррелируют с 3D-индексами Винера молекулярных графов СПУ, полученных из графов G^*, U расчётными методами молекулярной механики и (или) квантовой химии. Поэтому вполне возможно их применение для прогнозирования физико-химических свойств неразветвлённых изомеров СПУ.

Представляется актуальным продолжение исследований по следующим направлениям: а) отбор информационных топологических индексов, пригодных для прогнозирования физико-химических свойств изомеров СПУ; б) разработка новых топологических индексов, построенных на основе кортежей свойств, пригодных для прогнозирования физико-химических свойств изомеров СПУ; в) разработка алгоритма генерации кортежей свойств РМВП-графов классов изомеров и конформеров без проверки их на изоморфизм.

Автор благодарит к.х.н Д. С. Тихонова (DESY, Германия) за ценные замечания по тексту статьи и за выполненные квантово-химические расчёты. Автор выражает благодарность М. Н. Назарову (МИЭТ) за полезные замечания по разделу 5.3. Автор выражает благодарность к.х.н. Г. А. Журко (Ивановский ГХТУ) за информацию о формулах расчёта торсионных углов в программе молекулярного моделирования Chemcraft [33].

Список литературы

1. Харари Ф. Теория графов. М. : Мир, 1973.
2. Diudea M. V., Gutman I., Jantschi L. Molecular Topology. Huntington, New York : Nova, 2001.
3. Wiener H. Structural determination of paraffin boiling points // Journal of the American Chemical Society. 1947. Vol. 69, no. 1. P. 17–20.
4. Todeschini R., Consonni V. Molecular Descriptors for Chemoinformatics. Vol. I & II. Weinheim : Wiley, 2009.
5. Полиен — Polyene. URL: <https://wikichi.ru/wiki/Polyene> (дата обращения : 05.11.2021).
6. Moss G. P. Basic terminology of stereochemistry (IUPAC Recommendations 1996) // Pure and Applied Chemistry. 1996. Vol. 68, iss. 12. P. 2193–2222.
7. Craig N. C., Demaison J., Groner P., Rudolph H. D. Electron delocalization in polyenes: A semiexperimental equilibrium structure for (3E) 1,3,5-hexatriene and theoretical structures for (3Z,5Z), (3E,5E), and (3E,5Z) 1,3,5,7-octatetraene // The Journal of Physical Chemistry A. 2015. Vol. 119, no. 1. P. 195–204.
8. Panchenko Y. N., De Mare G. R., Aroca R., Bock C. W. All-trans- and t,T,t,C,t,T,t-deca-1,3,5,7,9-pentaenes: Ab initio structures, vibrational analyses,

- and some regularities in the series of related molecules // *Structural Chemistry*. 2000. Vol. 11, no. 2–3. P. 195–204.
9. **Dobrynin A. A., Entringer R., Gutman I.** Wiener index of trees: Theory and applications // *Acta Applicandae Mathematicae*. 2001. Vol. 66, no. 3. P. 211–249.
 10. **Cyvin S. J., Lloyd E. K., Cyvin B. N., Brunvoll J.** Chemical relevance of a pure combinatorial problem: Isomers of conjugated polyenes // *Structural Chemistry*. 1996. Vol. 7, no. 3. P. 183–186.
 11. **Markov M.** On the vertex separation of maximal outerplanar graphs // *Serdica Journal of Computing*. 2008. Vol. 154, no. 5. P. 207–238.
 12. **Uznanski Dan.** "Grid." From MathWorld — A Wolfram Web Resource, created by Eric W. Weisstein. URL: <https://mathworld.wolfram.com/Grid.html> (дата обращения : 05.11.2021).
 13. **Понарин Я. П.** Элементарная геометрия : в 2 т. Т. 1 : Планиметрия, преобразования плоскости. М. : МЦНМО, 2004.
 14. **Носов Ю. Л.** О максимальных внешнеплоских графах с двумя симплицальными вершинами // *Приклад. дискрет. математика*. 2015. Т. 29, № 3. С. 95–109.
 15. **Носов Ю. Л.** Экстремальные по индексу Винера максимальные внешнеплоские графы с двумя симплицальными вершинами // *Челяб. физ.-мат. журн.* 2019. Т. 4, вып. 3. С. 285–322.
 16. **Дрейпер Н., Смит Г.** Прикладной регрессионный анализ : в 2 кн. Кн. 1. М. : Финансы и статистика, 1986.
 17. **Bogdanov B., Nikolić S., Trinajstić N.** On the three-dimensional wiener number // *Journal of Mathematical Chemistry*. 1989. Vol. 3, no. 7. P. 299–309.
 18. CRC Standard Mathematical Tables and Formulas. 33rd ed. / Ed. D. Zwillinger. Boca Raton, London, New York : Chapman and Hall/CRC, 2018.
 19. **Asratian A. S., Oksimets N.** Graphs with Hamiltonian balls // *Australian Journal of Combinatorics*. 1998. Vol. 17, no. 4. P. 185–198.
 20. **Konstantinova E. V., Vidyuk M. V.** Discriminating tests of information and topological indices. Animals and trees // *Journal of Chemical Information and Computer Sciences*. 2003. Vol. 43, no. 6. P. 1860–1871.
 21. **Balaban A. T., Ionescu-Pallas N., Balaban T. S.** Asymptotic values of topological indices J and J' (Average distance sum connectivities) for infinite acyclic graphs // *MATCH Communications in Mathematical and in Computer Chemistry*. 1985. Vol. 17. P. 121–146.
 22. Шестиугольная решётка. URL: <https://google-nfo.org/1117811/1/shestiugolnayareshetka.html> (дата обращения : 25.11.2021).
 23. Wolfram Mathematica. URL: <https://www.wolfram.com/mathematica> (дата обращения : 02.02.2019).
 24. **Lewars E. G.** Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics. Springer : Dordrecht, Heidelberg, London, New York, 2011.
 25. **Hanwell M. D., Curtis D. E., Lonie D. C., Vandermeersch T., Zurek E., Hutchison G. R.** Avogadro: an advanced semantic chemical editor, visualization, and analysis platform // *Journal of Cheminformatics*. 2012. Vol. 4, no. 17. P. 1–17.
 26. **Halgren T. A.** Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94 // *Journal of Computational Chemistry*. 1996. Vol. 17, no. 5–6. P. 450–519.
 27. **Granovsky A. A.** Firefly v.8. URL: <http://classic.chem.msu.su/gran/firefly/index.html> (дата обращения : 08.11.2021).

28. **Schmidt M. W., Baldrige K. K., Boatz J. A., Elbert S. T., Gordon M. S., Jensen J. H., Koseki S., Matsunaga N., Nguyen K. A., Su S., Windus T. L., Dupuis M., Montgomery J. A.** General atomic and molecular electronic structure system // *Journal of Computational Chemistry*. 1993. Vol. 14, no. 11. P. 1347–1363.
29. **Becke A. D.** Density-functional thermochemistry. III. The role of exact exchange // *The Journal of Chemical Physics*. 1993. Vol. 98, no. 7. P. 5648–5652.
30. **Grimme S., Ehrlich S., Goerigk L.** Effect of the damping function in dispersion corrected density functional theory // *Journal of Computational Chemistry*. 2011. Vol. 32, no. 7. P. 1456–1465.
31. **Weigend F., Ahlrichs R.** Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy // *Physical Chemistry Chemical Physics*. 2005. Vol. 7, no. 18. P. 3297–3305.
32. **Neese F.** The ORCA program system // *Wiley Interdisciplinary Reviews: Computational Molecular Science*. 2005. Vol. 2, no. 1. P. 73–78.
33. **Zhurko G. A.** Chemcraft — graphical program for visualization of quantum chemistry computations. Ivanovo, Russia, 2005. URL: <https://chemcraftprog.com> (accessed : 15.04.2021).
34. **Furtula B., Lekishvili G., Gutman I.** A graph theoretical approach to cis/trans isomerism // *Journal of the Serbian Chemical Society*. 2014. Vol. 79, no. 7. P. 805–813.

Поступила в редакцию 09.06.2021.

После переработки 11.01.2022.

Сведения об авторе

Носов Юрий Леонидович, консультант Guardian Software Inc., Калгари, Канада; удалённый офис: Липецк, Россия; e-mail: yl.nosov@yandex.ru.

TOPOLOGICAL INDICES OF MAXIMAL OUTERPLANE GRAPHS WITH TWO SIMPLICIAL VERTICES

Yu.L. Nosov

*Guardian Software Inc., Calgary, Canada; remote office: Lipetsk, Russia
yl.nosov@yandex.ru*

We explore the maximal outerplane graphs (MOP-graphs) with two simplicial vertices, laid out on the lattice of equilateral triangles (lattice MOPs, or LMOPs). Weak dual of LMOPs are isomorphic and similar to molecular graphs of isomers of conjugated polyene hydrocarbons (CPH). For this reason the LMOPs lend themselves to prognosticating properties of CPH. We suggest a new approach to search of multiple structure-property relationship (QSPR) correlations between topological indices (TI) of LMOPs from some given set and the physico-chemical parameters of isomers (CPH). The essence of such an approach lies in the initial selection of TI suitable for prognostication of properties of unbranched isomers CPH with subsequent search for structure-property correlations between already selected TI and physico-chemical parameters of isomers CPH. Such a methodology allows to eliminate unsuitable TIs, as well as physico-chemical parameters that poorly correlate with the 3D Wiener index. The fitness of the suggested approach has been proven through a computational experiment. Within the framework of the suggested approach we studied the TI of LMOP-graphs of $n = 8, 10, \dots, 28$ orders, whose weak duals are isomorphic and geometrically similar to the molecular graphs of CPH model adopted in S.J. Cyvin (1996). From the multitude of Balaban, Wiener, Schultz, Randich, Harary and Zagreb group TIs simple J and modified J' Balaban indices were selected.

Keywords: *planar graph, maximal outerplane graph, topological index, 3D Wiener index, application of graph theory.*

References

1. **Harary F.** *Graph Theory*. Reading, Addison-Wesley, 1969.
2. **Diudea M.V., Gutman I., Jantschi L.** *Molecular Topology*. Huntington, New York, Nova, 2001.
3. **Wiener H.** Structural Determination of Paraffin Boiling Points. *Journal of the American Chemical Society*, 1947, vol. 69, no. 1, pp. 17–20.
4. **Todeschini R., Consonni V.** *Molecular Descriptors for Chemoinformatics*. Vol. I & II. Weinheim, Wiley, 2009.
5. *Polyene*. Available at: <https://en.wikipedia.org/wiki/Polyene>, accessed: 05.11.2021.
6. **Moss G.P.** Basic terminology of stereochemistry (IUPAC Recommendations 1996). *Pure and Applied Chemistry*, vol. 68, iss. 12, pp. 2193–2222.
7. **Craig N.C., Demaison J., Groner P., Rudolph H.D.** Electron delocalization in polyenes: A semiexperimental equilibrium structure for (3E) 1,3,5-hexatriene and theoretical structures for (3Z,5Z), (3E,5E), and (3E,5Z) 1,3,5,7-octatetraene. *The Journal of Physical Chemistry A*, 2015, vol. 119, no. 1, pp. 195–204.
8. **Panchenko Y.N., De Mare G.R., Aroca R., Bock C.W.** All-trans- and t,T,t,C,t,T,t-deca-1,3,5,7,9-pentaenes: Ab initio structures, vibrational analyses, and some regularities in the series of related molecules. *Structural Chemistry*, 2000, vol. 11, no. 2–3, pp. 195–204.
9. **Dobrynin A.A.** Wiener index of trees: Theory and applications. *Acta Applicandae Mathematicae*, 2001, vol. 66, no. 3, pp. 211–249.
10. **Cyvin S.J., Lloyd E.K., Cyvin B.N., Brunvoll J.** Chemical relevance of a pure combinatorial problem: Isomers of conjugated polyenes. *Structural Chemistry*, 1996, vol. 7, no. 3, pp. 183–186.

11. **Markov M.** On the vertex separation of maximal outerplanar graphs. *Serdica Journal of Computing*, 2008, vol. 154, no. 5, pp. 207–238.
12. **Uznanski Dan.** "Grid." from *MathWorld — A Wolfram Web Resource*, created by Eric W. Weisstein. Available at: <https://mathworld.wolfram.com/Grid.html> (accessed: 05.11.2021).
13. **Ponarin Ya.P.** *Elementarnaya geometriya* [Elementary geometry]. Vol. 1. *Planimetriya, preobrazovaniya ploskosti* [Planimetry, plane transformations]. Moscow, Publ. House of Moscow Center for Continuous Mathematical Education, 2004. (In Russ.).
14. **Nosov Yu.L.** O maksimal'nykh vneshneploskikh grafakh s dvumya simplitsial'nymi vershinami [Maximal outerplane graphs with two simplicial vertices]. *Prikladnaya diskretnaya matematika* [Applied Discrete Mathematics], 2015, vol. 29, no. 3, pp. 95–109. (In Russ.).
15. **Nosov Yu.L.** Extremal by Wiener index maximal outerplane graphs with two simplicial vertices. *Chelyabinsk Physical and Mathematical Journal*, 2019, vol. 4, no. 3, pp. 285–322. (In Russ.).
16. **Draper N.R., Smith H.** *Applied Regression Analysis*. New York, Chichester, Brisbane, Toronto, Singapore, Wiley & Sons, 2009.
17. **Bogdanov B., Nikolić S., Trinajstić N.** On the three-dimensional wiener number. *Journal of Mathematical Chemistry*, 1989, vol. 3, no. 7, pp. 299–309.
18. *CRC Standard Mathematical Tables and Formulas*. 33rd ed./ Ed. D. Zwillinger. Boca Raton, London, New York, Chapman and Hall/CRC, 2018.
19. **Asratian A.S., Oksimets N.** Graphs with hamiltonian balls. *Australian Journal of Combinatorics*, 1998, vol. 17, no. 4, pp. 185–198.
20. **Konstantinova E.V., Vidyuk M.V.** Discriminating tests of information and topological indices. Animals and trees. *Journal of Chemical Information and Computer Sciences*, 2003, vol. 43, no. 6, pp. 1860–1871.
21. **Balaban A.T., Ionescu-Pallas N., Balaban T.S.** Asymptotic values of topological indices J and J' (average distance sum connectivities) for infinite acyclic graphs. *MATCH Communications in Mathematical and in Computer Chemistry*, 1985, vol. 17, pp. 121–146.
22. *Shestiugol'naya reshyotka* [Hexagonal lattice]. Available at: URL: <https://google-nfo.org/1117811/1/shestiugolnaya-reshetka.html> (accessed: 25.11.2021). (In Russ.).
23. *Wolfram Mathematica*. Available at: <https://www.wolfram.com/mathematica> (accessed: 25.11.2021)
24. **Lewars E.G.** *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics*. Dordrecht, Heidelberg, London, New York, Springer, 2011.
25. **Hanwell M.D., Curtis D.E., Lonie D.C., Vandermeersch T., Zurek E., Hutchison G.R.** Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics*, 2012, vol. 4, no. 17, pp. 1–17.
26. **Halgren T.A.** Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *Journal of Computational Chemistry*, 1996, vol. 17, no. 5–6, pp. 450–519.
27. **Granovsky A.A.** *Firefly v. 8*. <http://classic.chem.msu.su/gran/firefly/index.html> (accessed: 02.02.2019).
28. **Schmidt M.W., Baldrige K.K., Boatz J.A., Elbert S.T., Gordon M.S., Jensen J.H., Koseki S., Matsunaga N., Nguyen K.A., Su S., Windus T.L., Dupuis M., Montgomery J.A.** General atomic and molecular electronic structure system. *Journal of Computational Chemistry*, 1993, vol. 14, no. 11, pp. 1347–1363.
29. **Becke A.D.** Density-functional thermochemistry. III. The role of exact exchange *Journal of Chemical Physics*, 1993, vol. 98, no. 7, pp. 5648–5652.
30. **Grimme S., Ehrlich S., Goerigk L.** Effect of the damping function in dispersion corrected density functional theory. *Journal of Computational Chemistry*, 2011, vol. 32, no. 7, pp. 1456–1465.

31. **Weigend F., Ahlrichs R** Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Physical Chemistry Chemical Physics*, 2005, vol. 7, no. 18, pp. 3297–3305.
32. **Neese F.** The ORCA program system. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2005, vol. 2, no. 1, pp. 73–78.
33. **Zhurko G.A.** *Chemcraft – graphical program for visualization of quantum chemistry computations*. Ivanovo, Russia, 2005. <https://chemcraftprog.com> (accessed: 15.04.2021).
34. **Furtula B., Lekishvili G., Gutman I.** A graph theoretical approach to cis/trans isomerism. *Journal of the Serbian Chemical Society*, 2014, vol. 79, no. 7, pp. 805–813.

Article received 09.06.2021.

Corrections received 11.01.2022.