

УДК 519.17:543.42

МЕТРИЧЕСКИЕ СПЕКТРЫ МОЛЕКУЛЯРНЫХ ГРАФОВ

Е.В.Мжельская, В.А.Скоробогатов

В в е д е н и е

В химии широко распространены физические методы исследований соединений, основанные на изучении молекулярных спектров. Масс-спектры отражают "поведение" некоторых осколков молекул, которые можно интерпретировать как подграфы молекулярных графов. Спектры ^{13}C ЯМР и ^1H ЯМР содержат информацию о "поведении в определенных условиях" отдельных атомов углерода или CN- , CH_2- , CH_3- групп. В целом молекулярные спектры воссоздают приближенную картину строения молекул. Метрические характеристики [1,2] также с определенной степенью приближения позволяют описать строение молекулярных графов: дают возможность характеризовать графы в целом, оценивать компактность (или разветвленность) графов, описывать положение отдельных подграфов (фрагментов) в молекулярном графе.

В данной работе предлагается новый тип метрических характеристик графов и молекулярных структур, имеющих вид физических спектров химических соединений.

1. Вершинный метрический спектр графов

В [1,2] определяются множества метрических характеристик, среди которых выделяются интегральные, вершинные, локальные и

относительные. Интегральные характеристики описывают свойства графа в целом. Примерами их служат молекулярные топологические индексы [3]. Вершинные характеристики, например эксцентриситет, дистанция, позволяют изучать метрические свойства вершин графа. Локальные и относительные характеристики отражают метрические свойства подграфов в графе.

Приводимые ниже спектральные метрические характеристики наглядно демонстрируют метрические особенности графов.

Определим вершинный метрический спектр графа следующим образом.

ОПРЕДЕЛЕНИЕ 1. Пусть $X = \{x_1, \dots, x_p\}$ и $Y = \{y_1, \dots, y_p\}$ есть значения некоторых функций вершинных метрических характеристик графа $G = (V, E)$, $|V(G)| = p$ и пара значений (x_i, y_i) соответствует i -й вершине графа G .

Пусть (x_i, y_i) - евклидовы координаты точки, образующей вместе с $(x_i, 0)$ в системе координат (X, Y) отрезок, который мы будем называть "спектральной линией" (или "спектральной полосой") для i -й вершины. Пару значений (x_i, y_i) будем называть координатами или значениями спектральной полосы i -й вершины. Условимся, что если $x_{i_1} = \dots = x_{i_k}$, $k \in \{2, p\}$, то координата y_k , $k = (i_1, \dots, i_k)$, вычисляется по формуле $y_k = y_{i_1} + \dots + y_{i_k}$ и соответствующая спектральная полоса для вершин (i_1, \dots, i_k) имеет кратность k . Тогда все множество спектральных линий определит вершинный метрический спектр графа G .

На рис.1 приведен пример такого спектра для дерева. В качестве координаты Y взят эксцентриситет вершин $e(v)$, а в качестве X - среднеквадратичное отклонение вершин $m_2(v)$ [1]. Для всех полос спектра указаны номера вершин, которым они

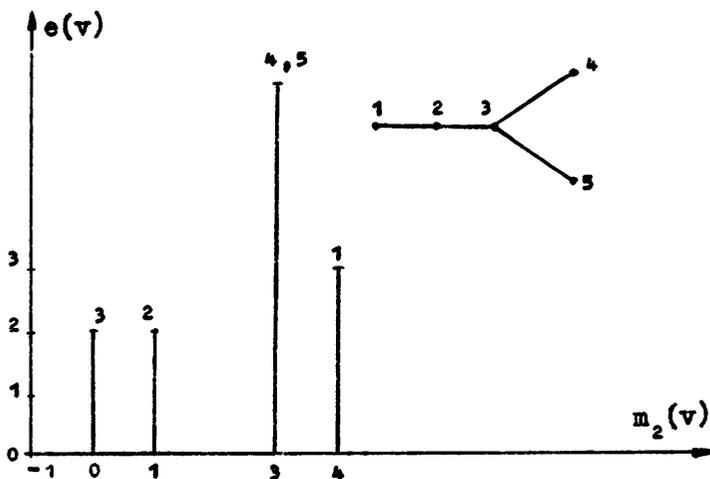


Рис. 1

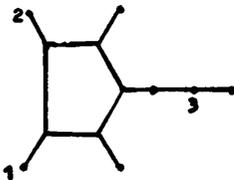
соответствуют. Спектральная полоса для вершин 4 и 5 имеет кратность 2. Нумерация вершин в графе произвольная.

Очевидно, что если граф имеет симметрии [1], то в спектре будут наблюдаться кратные полосы (в приведенном примере вершины 4 и 5 образуют орбиту в графе). Однако спектральные полосы могут быть кратными и тогда, когда будут совпадать координаты неэквивалентных вершин. Такая ситуация будет приводить к тому, что спектральные полосы, соответствующие различным по свойствам вершинам, могут давать вклад в величину одной полосы, как и эквивалентные вершины. Для того, чтобы избежать этой ситуации, можно рассматривать более сложные спектры, подбирая в качестве координат спектра различные сочетания характеристик с целью выделения кратных спектральных полос, соответствующих только эквивалентным вершинам.

Заметим, что вершинные спектры изометричных графов с одинаковыми матрицами слоев [1] совпадают, поскольку из равенства их λ -матриц следует равенство значений вершинных метрических характеристик.

Вершинные спектры транзитивных графов, в частности циклов, содержат одну спектральную линию (ее кратность равна порядку графа), так как все вершины в таком графе являются эквивалентными.

В дальнейшем будем рассматривать спектры в системе координат (X, Y) с функциями $X = \sigma_2(v)d(v) \cdot 10^{-2}$ и $Y = e(v)k \cdot 10^{-1}$, где $\sigma_2(v)$ - среднеквадратичное отклонение вершины, $d(v)$ - дистанция вершины, $e(v)$ - эксцентриситет вершины, k - число вершин с данным значением эксцентриситета. Функции X и Y взяты в качестве координат спектра, так как в них представлены основные метрические характеристики графов: функция X отражает дистанционные свойства графов, а Y - эксцентриситетные. Именно эти два класса определяют метрические характеристики графов. Коэффициенты в функциях X и Y подобраны для удобства изображения спектров в логарифмическом масштабе; иногда они будут опущены.



X	Y	Номера вершин
4.14	1.2	1, 2
4.14	0.5	3

Рис. 2

Отметим, что несмотря на относительную сложность функций, выбранных в качестве координат спектра, имеется пример (рис.2), для которого не выполняется требование кратности спектральных

полос, соответствующих только эквивалентным вершинам. Как видно из таблицы рис.2, значения абсциссы X вершин 1,2 и 3 совпадают. Следовательно, в спектре должна появиться полоса кратности 3 для вершин, которые эквивалентными не являются. Это можно избежать, рассматривая в качестве абсциссы функцию $X' = X \cdot e(v)$, что позволит различить вершины с одинаковыми значениями функции X и устранить появление кратных полос для неэквивалентных вершин.

Поскольку перед нами не стоит цель поиска функций, удовлетворяющих этому условию, то функции X и Y , отражающие по-разному метрические особенности графов, оказываются подходящими для демонстрации различных свойств и возможностей метрических спектров.

Покажем, например, как можно сравнивать вершинные спектры графов одного порядка с равным числом спектральных полос, а результат сравнения использовать для упорядочения графов.

Пусть $\{(x_i^G, y_i^G)\}$, $i = \overline{1, m}$, и $\{(x_i^H, y_i^H)\}$, $i = \overline{1, m}$, есть множества значений спектральных полос метрических спектров графов G и H , упорядоченных по убыванию абсцисс. Тогда принципом упорядочения графов G и H может служить лексикографический порядок на парах (x_i^G, y_i^G) , $i = \overline{1, m}$, и (x_i^H, y_i^H) , $i = \overline{1, m}$: меньший номер имеет граф, вектор абсцисс спектральных полос которого является лексикографически большим.

Так, для КК[6]-графов ката-конденсированных полибензолов [4], изображенных на рис. 3, упорядочение выглядит следующим образом: $\langle G, H \rangle$, поскольку первые пять значений абсцисс метрических спектров графов G и H совпадают, а шестое значение больше для графа G ($x_6^G = 18.89227$, $x_6^H = 16.34318$). Кратность всех спектральных полос равна 2. Нумерация полос соответствует нумерации вершин в графах*).

*) Изображение дано в логарифмическом масштабе.

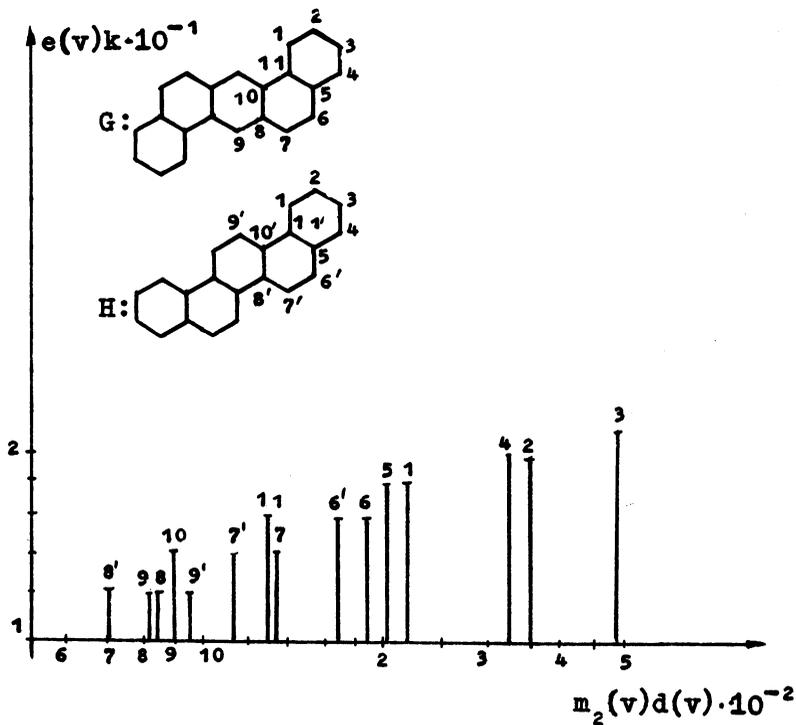


Рис. 3

Рассмотренные графы G и H являются Z -эквивалентными. Понятие Z -эквивалентности было введено в [5] и основано на совпадении у $KK[6]$ -графов вектора $Z = \{d; r; l; L; B\}$, получаемого из матрицы слоев [1], элементами которого являются: d - диаметр, r - радиус; l - число всех попарно различных строк в матрице слоев; $L = (l_1, \dots, l_n)$ - вектор значений, l_j - число строк одинаковой длины, $n = d - r + 1$; $B = (b_1, \dots, b_n)$ - вектор значений, b_j - число вершин, имеющих строку одной и той же длины, т.е. равного эксцентриситета.

Отметим, что множества значений $\{v_i^G\}$ и $\{v_i^H\}$, $i = \overline{1, m}$, вершинных спектров Z -эквивалентных графов G

и H в выбранной системе координат совпадают. Это следует из равенства векторов $V^G = (b_1^G, \dots, b_m^G)$ и $V^H = (b_1^H, \dots, b_m^H)$, элементы которых равны числу вершин с данным значением эксцентриситета, что и означает равенство координат $\{y_i^G\}$ и $\{y_i^H\}$, $i = \overline{1, m}$. На рис.3 дана иллюстрация этого свойства: значения ординат спектральных полос графов G и H для вершин с одинаковой нумерацией совпадают.

2. Вершинный метрический спектр химических структур

Химические структуры могут быть представлены помеченными графами. Метки вершин соответствуют атомам или их группам, метки ребер - связям. В качестве меток можно рассматривать как символы (например, CH_3 , NO_2 , S и т.д.), так и количественные характеристики - веса. Под весом ребра (связи) можно понимать, например, кратность химической связи, длину химической связи и т.д. Вес вершины может быть представлен любыми количественными данными, характерными для соответствующего атома или группы атомов, например, значения атомных весов или ван-дер-ваальсовы радиусы атомов. В [6] в качестве весов вершин и ребер предлагается использовать величины, в которых учтены числа всех валентных электронов внутренних связей (для атомов) и числа всех электронов атомов, соединенных данной связью (для связей).

Пусть $G = (V, E)$ - помеченный граф, вершины и ребра которого имеют вес. Будем называть помеченный граф G , соответствующий химической структуре F , молекулярным графом.

Метрический спектр химической структуры определяется с учетом весов вершин и ребер молекулярного графа.

Одним из предлагаемых способов такого учета является следующий.

2.1. Молекулярный метрический спектр с учетом весов атомов и связей в качестве коэффициентов.

ОПРЕДЕЛЕНИЕ 2. Пусть $\{(x_i, y_i)\}$, $i = \overline{1, p}$, есть множество значений вершинного спектра графа G (без учета кратности полос), соответствующего химической структуре F , и пусть $\alpha(g_i)$ - вес атома g_i структуры F , $i = \overline{1, p}$, $\beta_j(g_i)$ - вес j -й связи атома g_i , $i = \overline{1, p}$, $j = \overline{1, s_i}$. Будем называть значениями (или координатами) спектральной полосы (спектральной линии) i -й вершины молекулярного графа G следующие величины:

$$\left. \begin{aligned} x(g_i) &= \frac{\sum_{j=1}^{s_i} \beta_j(g_i)}{\alpha(g_i) + 1} x_i, \\ y(g_i) &= y_i, \quad i = \overline{1, p}. \end{aligned} \right\} \quad (1)$$

Если $x(g_{i_1}) = \dots = x(g_{i_k})$ и $\alpha(g_{i_1}) = \dots = \alpha(g_{i_k})$, $\beta_j(g_{i_1}) = \dots = \beta_j(g_{i_k})$, $k \in \overline{2, p}$, то координата $y(g_k)$, $k = (i_1, \dots, i_k)$, вычисляется по формуле $y(g_k) = y(g_{i_1}) + \dots + y(g_{i_k})$ и соответствующая спектральная полоса для атомов $(g_{i_1}, \dots, g_{i_k})$ имеет кратность k . Тогда все множество спектральных линий определит молекулярный метрический спектр химической структуры F .

В данном случае веса атомов и связей учитываются в качестве коэффициентов при определении множества абсцисс спектра.

На рис. 4 изображены спектры молекулярных графов метилбензола и нитробензола, полученные на основе формул (1) с использованием значений весов атомов и связей из [6]. Систему координат, как и ранее, определяют функции $X = m_2(v)d(v)$, $Y = e(v)k$. Можно заметить, что наблюдается некоторая "преемственность" спектральных линий для этих графов. Такая ситуация объясняется следующим образом: положение атомов 1, 2, 3 в струк-

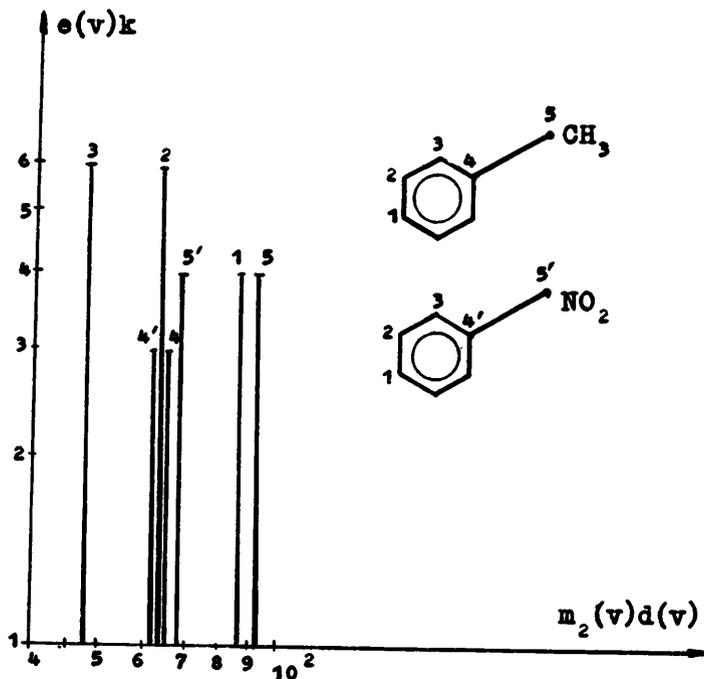


Рис. 4

туре метилбензола и их окружение ничем не отличается от расположения соответствующих атомов в нитробензоле. Спектральные линии в этом случае различаются только для групп CH_3 и NO_2 и атомов углерода, связанных с этими группами. Кратность спектральных полос с номерами 2 и 3 равна 2. Нумерация полос соответствует нумерации вершин в графах. Изображение дано в логарифмическом масштабе.

Такое определение молекулярного метрического спектра является наиболее простым, поскольку учитывает влияние друг на друга лишь соседних атомов. При этом во внимание не берутся атомы, находящиеся от данного на расстоянии более единицы.

Этого недостатка лишен второй способ определения значений спектральных полос молекулярного метрического спектра.

2.2. Молекулярный метрический спектр на основе матрицы расстояний для помеченных графов. Данный способ определения состоит в использовании метрических характеристик, вычисляемых непосредственно из матрицы расстояний для помеченных графов, которая определяется следующим образом.

Пусть $G = (V, E)$ - молекулярный граф химической структуры (помеченный граф, вершины и ребра которого имеют вес). Под расстоянием d_{ij} между вершинами $i, j \in V(G)$ молекулярного графа G понимается минимальный вес (i, j) -цепи кратчайшей длины. Вес цепи определяется как суммарный вес вершин и ребер, составляющих данную цепь^{*)}.

ОПРЕДЕЛЕНИЕ 3. Матрицей расстояний $D(G) = \|d_{ij}\|$ молекулярного графа G с p вершинами называется $(p \times p)$ -матрица, в которой d_{ij} есть расстояние между вершинами i и j , а d_{ii} - вес i -й вершины.

Используя построенную таким образом матрицу расстояний, можно непосредственно рассчитать метрические характеристики для молекулярного графа^{**)}. Например, дистанция молекулярного графа вычисляется по формуле:

$$D(G) = \frac{1}{2} \sum_{i, j=1}^p d_{ij}.$$

Определение молекулярного метрического спектра в данном случае совпадает с определением вершинного метрического спектра для обыкновенного графа, где в качестве координат X и Y

*) Вес цепи может быть определен и другим способом, что зависит от физического смысла, вкладываемого в веса вершин и ребер.

***) Формулы для вычисления метрических характеристик представлены в [1].

взяты значения некоторых функций вершинных метрических характеристик, полученные по матрице расстояний.

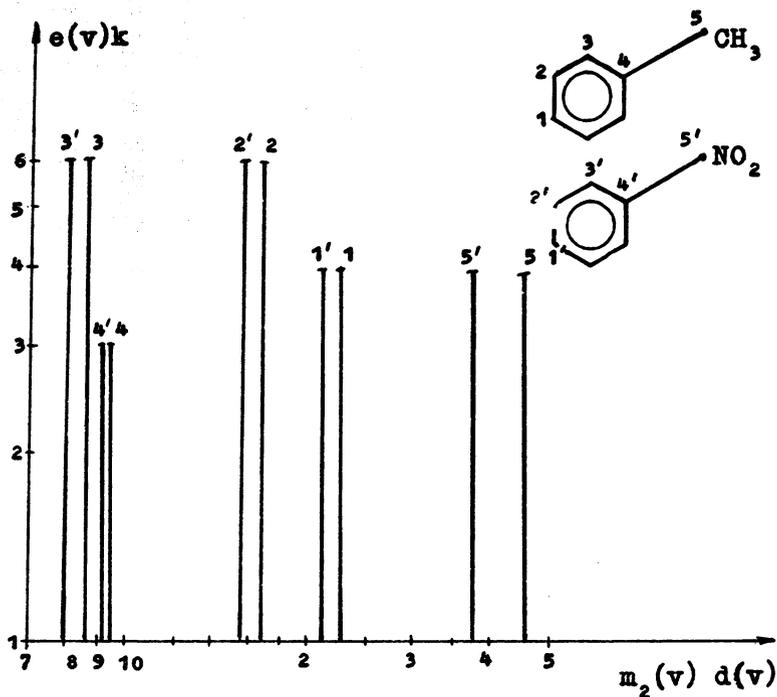


Рис. 5

На рис. 5 изображены спектры молекулярных графов метилбензола и нитробензола. Систему координат, как и раньше, определяют функции $X = m_2(v) d(v)$, $Y = e(v)k$, веса связей взяты из [6]. Как видно из рис. 5, метрические характеристики, полученные по матрице расстояний, более точно учитывают веса в графе: в отличие от спектров, изображенных на рис. 4, спектральные линии всех вершин молекулярных графов метилбензола и нитробензола различны. Это объясняется тем, что в первом случае для определения значений молекулярного метрического спектра учитываются лишь веса ближайшего окружения, в то время как по матрице рас-

стояний вершинные характеристики получаются с учетом весов всех вершин и ребер.

Оба определения молекулярного метрического спектра могут быть использованы для изучения свойств молекулярных графов химических структур, а также для поиска корреляций с физическими спектрами, поскольку дают возможность с разной степенью точности учитывать веса вершин и ребер рассматриваемых графов.

3. Спектры химических структур, определяемых фрагментами

В §1 и 2 рассматривались вершинные метрические спектры. В качестве координат их спектральных линий использовались значения функций вершинных метрических характеристик. Ниже рассматриваются молекулярные метрические спектры, координаты спектральных линий которых соответствуют значениям метрических характеристик подграфов (фрагментов) молекулярного графа (химической структуры).

3.1. Метрический спектр относительного положения фрагментов в структуре. Как уже упоминалось ранее, среди множества метрических характеристик выделяют относительные характеристики [2], которые связывают с подграфами. Например, дистанция подграфа $H \subseteq G$ определяется как

$$D_G(H) = \sum_{\substack{v \in V(H) \\ u \in V(G)}} d(u, v),$$

а нормированное значение $D(H, G) = \frac{1}{D(G)} D_G(H)$ является относительной дистанцией подграфа H в G . Аналогичные формулы можно получить для других метрических характеристик [2].

Рассматривая некоторое множество подграфов $\{H_i\}$, $i = \overline{1, n}$, в графе G , получим соответствующее множество значений их относительных метрических характеристик $\{m_k(H_i, G)\}$,

$i = \overline{1, n}, k = \overline{1, r}$. В качестве подграфов $\{H_i\}$ могут быть выбраны, например, ребра. В этом случае будем иметь относительные реберные характеристики, которые позволяют сравнивать относительное положение ребер.

Применение относительных метрических характеристик может быть обосновано тем, что свойства химических соединений зависят не только от наличия в соединении некоторого фрагмента, но и от положения этого фрагмента в молекуле, взаимного расположения фрагментов, формы самой молекулы [2].

Пусть $f = \{f_i, i = \overline{1, n}\}$ - множество фрагментов, $\mathcal{F} = \{F_j, j = \overline{1, l}\}$ - множество молекулярных структур и $\mathcal{M} = \{m_k(f_i, F_j), i = \overline{1, n}, j = \overline{1, l}, k = \overline{1, r}\}$ - множество значений относительных характеристик для фрагментов $f_i \subseteq F_j$. Тогда можно рассмотреть следующие фрагментные молекулярные метрические спектры.

ОПРЕДЕЛЕНИЕ 4. Пусть $f = \{f_i, i = \overline{1, m}\}$, $f_i \subseteq F$, есть множество различных фрагментов, входящих в химическую структуру F , и $m_{k_1}(f_i, F)$ и $m_{k_2}(f_i, F), i = \overline{1, m}; k_1, k_2 \in \overline{1, r}$, есть значения некоторых относительных метрических характеристик. Будем называть значениями (или координатами) спектральной полосы (спектральной линии) i -го фрагмента пару значений $(m_{k_1}(f_i, F), m_{k_2}(f_i, F))$. Тогда все множество спектральных линий фрагментов определит фрагментный молекулярный метрический спектр химической структуры F .

Определенный таким образом спектр, обозначим его фрагментный спектр l , характеризует относительное положение различных фрагментов в некоторой структуре F .

ОПРЕДЕЛЕНИЕ 5. Пусть $m_{k_1}(f, F_i)$ и $m_{k_2}(f, F_i), i = \overline{1, l}$, есть значения относительных характеристик некоторого фрагмента $f \subseteq F_i, i = \overline{1, l}$, входящего во все химические структуры множества $\mathcal{F} = \{F_i, i = \overline{1, l}\}$. Будем называть значениями (или коор-

динатами) спектральной полосы (спектральной линии) фрагмента f i -й структуры пару значений $(m_{k_1}(f, F_i), m_{k_2}(f, F_i))$.

Тогда все множество спектральных линий определит фрагментный молекулярный метрический спектр химических структур F_i , $i = 1, 1$. В таком спектре, назовем его фрагментный спектр 2, имеется возможность оценивать относительное положение одного и того же фрагмента в различных структурах.

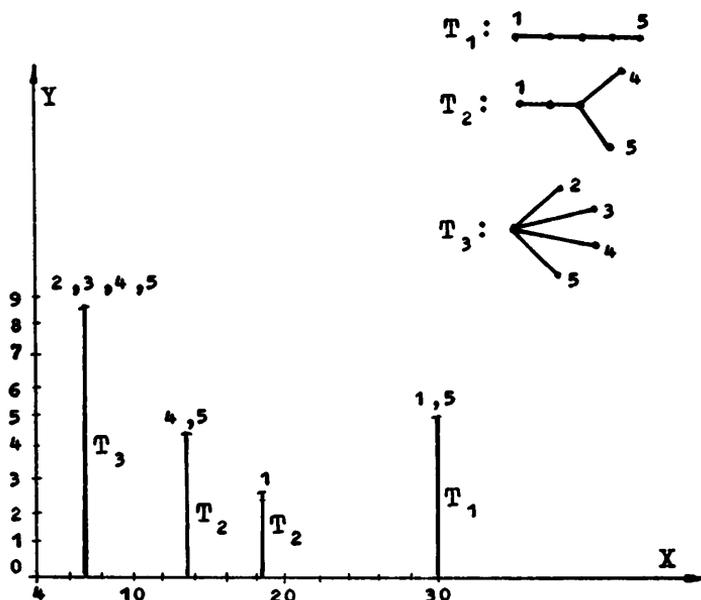


Рис. 6

На рис.6 представлен метрический фрагментный спектр 2 молекулярных графов, порядок которых равен пяти. В качестве фрагментов рассматривались группы CH_3 . Систему координат определяют функции $X = m_2(v, G)d(v, G) \cdot 10^2$, $Y = e(v, G)k \cdot 10$, где $m_2(v, G)$, $d(v, G)$, $e(v, G)$ являются относительными метрическими характеристиками вершины v в графе G . Кратные спектральные линии, как и ранее, соответствуют эквивалентным фраг-

ментам (в данном случае эквивалентным вершинам). Неэквивалентные фрагменты дают различные спектральные полосы (для графа T_2 это полосы, соответствующие вершинам 4, 5 и 1).

В общем случае можно определить спектр для множества структур F и множества фрагментов f и рассматривать относительное положение фрагментов для всего множества химических структур.

3.2. Метрический спектр взаимного положения фрагментов в структуре. Рассмотренные фрагментные матрические спектры основаны на относительных характеристиках подграфов в графе. Для оценки взаимного положения подграфов H, H_1 в графе G в [2] вводятся характеристики взаимного положения. Например, относительная дистанция подграфа H_1 по отношению к H есть функция

$$D(H_1, H) = \frac{1}{D(G)} D_H(H_1),$$

где
$$D_H(H_1) = \sum_{v \in V(H_1)} D_H(v),$$

$$D_H(v) = \sum_{u \in V(H)} d_G(u, v).$$

Такие характеристики позволяют оценить положение выбранного семейства подграфов, например, по отношению к центру графа или другим его частям. Для выбранной совокупности подграфов H_1, H_2, \dots, H_n ; H получим n значений характеристик $m_1(H_1, H)$ и $m_2(H_1, H)$, $i = 1, n$, которые можно рассматривать в качестве координат спектральных линий и определять спектр аналогично описанным ранее спектрам фрагментов молекулярных структур. Варьируя относительные метрические характеристики при выборе системы координат, можно рассматривать различные метрические спектры взаимного положения подграфов в графе или, если говорить о химических структурах, различные метрические спектры взаимного положения фрагментов в структуре.

Таким образом, методика определения относительного положения подграфов в графе дает математические средства метрической характеристики молекулярного графа с учетом его структурных особенностей и весов, а методика определения метрических спектров позволяет отражать наглядно эти особенности.

В целом молекулярные метрические спектры, как и физические спектры, дают некоторые общие представления о строении химической структуры. Например, и в спектре ^{13}C ЯМР, и в метрическом спектре содержатся спектральные линии эквивалентных вершин. Имеющиеся соответствия в свойствах спектров дают реальную возможность поиска корреляций между метрическими и физическими спектрами.

Л и т е р а т у р а

1. СКОРОБОГАТОВ В.А., ХВОРОСТОВ П.В. Анализ метрических свойств графов //Методы обнаружения закономерностей с помощью ЭВМ. - Новосибирск, 1981. - Вып. 91: Вычислительные системы. - С. 3-20.

2. SKOROBOGATOV V.A., DOBRYNIN A.A. Metric Analysis of Graphs //Math. Chem. (MATCH). - 1988. -N. 23. - P. 105-151.

3. СТАНКЕВИЧ М.И., СТАНКЕВИЧ И.В., ЗЕФИРОВ И.С. Топологические индексы в органической химии //Успехи химии. - 1988. - Т. 57. - С. 337-366.

4. МЖЕЛЬСКАЯ Е.В. Графы полициклических соединений //Вопросы алгоритмического анализа структурной информации. - Новосибирск, 1987. - Вып.119: Вычислительные системы. -С. 71-90.

5. СКОРОБОГАТОВ В.А., МЖЕЛЬСКАЯ Е.В., МЕЙРМАНОВА Н.М. Изучение метрических характеристик ката-конденсированных полибензолов //Алгоритмический анализ графов и его приложения.-Новосибирск, 1988. - Вып. 127: Вычислительные системы. -С. 40-91.

6. БАРИШ М., ЯШАРИ Дж., ЛАЛЛ Р. и др. Матрица расстояний для молекул, содержащих гетероатомы //Химические приложения топологии и теории графов: Перевод с англ. Под ред. Р.Кинга.-М.: Мир, 1987. -С. 259-265.

Поступила в ред.-изд.отд.

25 июля 1989 года