

Г214

М20

На правах рукописи



Малышева Юлия Анатольевна

**ИСПОЛЬЗОВАНИЕ ТОПОЛОГИЧЕСКИХ ИНДЕКСОВ
В КОЛИЧЕСТВЕННЫХ КОРРЕЛЯЦИЯХ
"СТРУКТУРА - СВОЙСТВО"
(на примере органических соединений разных классов)**

Специальность 02.00.04 - физическая химия

Автореферат
диссертации на соискание ученой степени
кандидата химических наук

Тверь - 2001

Работа выполнена на кафедре физической химии
Тверского государственного университета

Научный руководитель : доктор химических наук, профессор,
заслуженный деятель науки РФ
Папулов Ю. Г.

Официальные оппоненты : доктор химических наук,
профессор Белик А.В.

кандидат химических наук,
доцент Левин В.П.

Ведущая организация : Институт органической химии РАН

Защита диссертации состоится "25" октября 2001 г. в 15³⁰ на
заседании Диссертационного Совета Д 212.263.02 при Тверском
государственном университете по адресу: 170002, Тверь, Садовый пер., 35.

43
библиотеке Тверского

001 г.

Лапина Г.П.

Общая характеристика работы

Актуальность темы. Анализ взаимосвязи между структурой и свойствами веществ в настоящее время является неотъемлемой частью химических исследований. Такой анализ, с одной стороны, позволяет прогнозировать свойства малоизученных и еще не синтезированных соединений (с целью отобрать те из них, которые потенциально могут обладать требуемыми свойствами), а с другой стороны углубляет и расширяет наши представления о строении молекул и химической связи.

Существуют методы, позволяющие строить математические модели взаимосвязи «структура-свойство»: квантовомеханические, феноменологические, методы статистической термодинамики, молекулярной механики и др. Все они не исключают, а так или иначе дополняют друг друга, опираясь на разные подходы к количественному описанию структуры химических соединений.

В последние десятилетия при поиске количественных соотношений «структура-свойство»* (и «структура-активность»**) широко применяются топологические индексы (ТИ), как самостоятельно, так и в виде определённых комбинаций с дескрипторами других типов. Несмотря на ограниченность заложенной в них структурной информации (двумерные индексы не учитывают метрические характеристики молекул), ТИ отличаются хорошей корреляционной способностью со многими физико-химическими свойствами и биологической активностью веществ. Простота расчета и возможность получать их непосредственно из структурной формулы делают ТИ особенно полезными на ранних стадиях исследований для предварительной сортировки соединений и оценки свойств гипотетических структур, для которых нет никаких экспериментальных данных.

В настоящее время известно несколько десятков топологических индексов, и продолжают появляться публикации о введении новых ТИ. Такое изобилие индексов при отсутствии их систематического сравнительного анализа затрудняет ориентацию в них и обостряет проблему мотивированного выбора ТИ для конкретного исследования. Довольно часто выбор топологических индексов носит случайный

* Англ. Quantitative Structure-Property Relationships (QSPR).

** Англ. Quantitative Structure-Activity Relationships (QSAR).

характер, и корреляционные зависимости на их основе плохо поддаются физико-химической интерпретации даже на качественном уровне строгости. При использовании того или иного индекса важно знать: представляет ли данный ТИ интерес для корреляций «структура-свойство» или не имеет ценности (не способен отражать важные структурные особенности молекул или дублирует информацию, получаемую с помощью других индексов). Все это и определяет актуальность настоящей работы, посвященной углубленному исследованию корреляционной способности широкого ряда ТИ на примере разных соединений и свойств.

Диссертация выполнена в соответствии с планом научно-исследовательских работ Тверского государственного университета по направлению «Связь свойств веществ со строением молекул: расчетно-теоретическое исследование» (№ гос. регистрации 01.84. 0085361), а также в рамках Российской межвузовской программы «Университеты России – фундаментальные исследования».

Цель работы состоит в проведении систематического исследования и анализа корреляционных возможностей топологических индексов на примере различных физико-химических свойств и биологической активности органических соединений разных классов. В соответствии с этим были поставлены следующие задачи:

- проанализировать современное состояние теоретико-графовых исследований корреляций «структура-свойство» на основе отечественных и зарубежных публикаций;
- подобрать группы соединений с известными значениями свойств и провести расчет топологических индексов для этих соединений;
- построить корреляционные зависимости между топологическими индексами и свойствами рассматриваемых соединений;
- провести анализ полученных зависимостей и выявить основные свойства и практические возможности исследуемых топологических индексов.

Научная новизна. Впервые проведено систематическое сравнительное исследование корреляционной способности большого числа топологических индексов на примере разных соединений и свойств.

Изучено влияние основных факторов (природа и состав рассматриваемых соединений, вид корреляционного уравнения, корреляционная способность ТИ) на качество моделей «структура-свойство». Выявлено, что в большинстве рассмотренных случаев некоторые простые индексы отличаются лучшей корреляционной способностью, чем более сложные ТИ. Отмечено, что низкая дискриминирующая способность некоторых индексов не является помехой в получении точных корреляционных моделей на их основе.

Установлено, что порядок распределения индексов по корреляционной способности сохраняется для разных групп соединений и свойств. Найдено, что наиболее точно во всех рассмотренных случаях описывают корреляционные связи между отдельными ТИ и свойствами квадратичная и степенная функции, использование которых на 20-30% снижает стандартное отклонение по сравнению с линейными зависимостями.

Исследованы корреляции ТИ в рядах неродственных соединений. Установлено, что статистические характеристики моделей «структура-свойство» в этом случае на 15-20% ниже, чем в рядах гомологов; при этом для некоторых свойств (энтальпии образования и токсичности) наблюдается резкое снижение корреляционной способности ТИ (в два раза), а в других случаях (температура кипения, молярный объем) уменьшение КС ТИ незначительно.

Показано на примере корреляций ТИ с токсичностью производных бензола, что применение простых топологических индексов позволяет получать корреляционные модели, не уступающие по точности моделям, построенным на основе более сложных и информативных молекулярных дескрипторов (квантово-химических, геометрических и др.).

Практическая значимость. Поскольку метод топологических индексов обладает прогнозирующей силой, полученные в работе корреляционные зависимости могут быть применены для расчета различных свойств органических соединений. Результаты сравнительного анализа корреляционной способности ТИ будут полезны в качестве справочного материала при построении корреляционных моделей «структура-свойство», а также могут быть использованы при чтении спецкурсов для студентов, специализирующихся в области физической органической химии

На защиту выносятся:

- методики построения и анализа корреляционных зависимостей на основе топологических индексов;
- результаты исследования корреляционной способности ТИ на примере физико-химических свойств органических соединений разных классов (алканов, алкенов, спиртов, альдегидов и кетонов, кислот, аминов);
- сравнительный анализ трехмерных и двумерных топологических индексов (на примере алканов);
- результаты изучения корреляционных зависимостей между ТИ и свойствами в рядах неродственных соединений;
- исследование корреляционных возможностей ТИ на примере токсичности производных бензола.

Апробация работы. Основные результаты работы представлялись на следующих конференциях и семинарах:

- 2-й Международной конференции "Математические модели нелинейных возбуждений, переноса, динамики, управления в конденсированных системах и других средах" (г. Тверь, 1996);
- XIII Семинаре по межмолекулярному взаимодействию и конформациям молекул (Тверь, 1997);
- International conference on chemical and biological thermodynamics (India, Amritsar, 1997);
- 3-й Международной научной конференции "Математические модели нелинейных возбуждений, переноса, динамики, управления в конденсированных системах и других средах" (Тверь, 1998);
- I Всероссийской конференции "Молекулярное моделирование" (Москва, 1998).

Публикации. По теме диссертации опубликовано 6 статей и 5 тезисов докладов.

Объем и структура работы. Диссертация состоит из введения, четырех глав, выводов и приложения. Она изложена на 152 страницах, содержит 19 таблиц, 17 рисунков. Список литературы включает 167 наименований.

Содержание работы

Во введении обоснована актуальность темы диссертации, сформулированы цель и основные задачи исследования, дана общая характеристика работы.

В первой главе «Обзор литературы» проанализировано современное состояние теоретико-графовых исследований взаимосвязи «структура–свойство» и показано, что недостаточное внимание уделяется сравнительному анализу корреляционной способности топологических индексов, используемых в этих исследованиях.

Типичный алгоритм, применяемый при изучении зависимостей между особенностями структуры молекул и проявляемыми свойствами, выглядит следующим образом:

- 1) подбор групп соединений с известными значениями свойств для построения модели зависимости «структура – свойство»;
- 2) представление структур молекул отобранных веществ в цифровом виде и расчет дескрипторов для исследуемых соединений;
- 3) построение модели зависимости «структура – свойство» по данным, полученным на предыдущих этапах; при этом в роли зависимой переменной выступает свойство вещества, а в качестве независимых – молекулярные дескрипторы (МД).

В QSPR/QSAR существуют разные подходы к количественному описанию молекулярной структуры химических соединений и используются молекулярные дескрипторы нескольких типов. В частности, выделяют несколько уровней МД: дескрипторы элементного уровня, дескрипторы структурной формулы, дескрипторы электронной структуры, дескрипторы молекулярной формы, дескрипторы межмолекулярных взаимодействий. При этом считается, что каждый последующий уровень включает информационное содержание предыдущего: дескрипторы структурной формулы несут всю информацию, содержащуюся в дескрипторах элементного уровня; дескрипторы электронной структуры включают всю информацию, содержащуюся в МД структурной формулы и т.д.

Современные корреляционные модели «структура – свойство» строятся, как правило, на основе комбинаций молекулярных дескрипторов разных типов. Такой подход позволяет учесть сразу несколько особенностей структуры исследуемых соединений и в то же время дает

возможность выявить те из них, которые наиболее существенны для данного свойства.

В последние десятилетия широкое распространение в QSPR/QSAR получили дескрипторы структурной формулы, называемые топологическими индексами (ТИ). ТИ рассчитываются на основе описания структурной формулы соединения методами теории графов. Несмотря на ограниченность заложенной в них структурной информации (не учитываются метрические свойства молекулы и пространственное расположение атомов), топологические индексы отличаются хорошей корреляционной способностью со многими физико-химическими свойствами и биологической активностью. ТИ применяются как самостоятельно, так и в составе комбинаций с дескрипторами других типов. Как правило, большинство удачных корреляционных моделей обязательно содержит в качестве параметров и топологические индексы. К числу достоинств ТИ относят простоту и быстроту их вычисления, а также эмпирический характер индексов и возможность получать их непосредственно из структурной формулы (последнее особенно существенно, когда речь идет об оценке свойств гипотетических структур, для которых вообще нет никаких экспериментальных данных).

Метод ТИ относится к числу быстро развивающихся и совершенствующихся научных направлений. Задачи расчета и прогнозирования свойств самых разных соединений побуждают к созданию новых эффективных расчетных схем и поиску новых ТИ. В настоящее время известно несколько десятков топологических индексов и число сообщений о введении все новых и новых ТИ не уменьшается. При этом недостаточное внимание уделяется сравнительному анализу корреляционных возможностей различных индексов. Это затрудняет ориентацию в них и обостряет проблему обоснованного выбора ТИ для конкретного исследования. На основании этого в работе поставлена задача - провести сравнительное исследование корреляционной способности широкого ряда ТИ на примере физико-химических свойств и биологической активности органических соединений разных классов.

Во второй главе «Методика построения зависимостей «структура-свойство» дается описание основных этапов получения корреляционных зависимостей, необходимых для проведения сравнительного исследования и анализа разных ТИ. Рассмотрены, в частности, основные понятия и определения, используемые при теоретико-

графовом описании структуры молекулы, соотношения и расчетные формулы для вычисления топологических индексов, виды корреляционных зависимостей между ТИ и свойствами соединений, способы оценки точности корреляционных моделей.

При топологическом описании молекул структурную формулу соединения представляют в виде графа, вершины которого соответствуют атомам, а ребра - химическим связям. Веса вершин и ребер характеризуют типы соответствующих атомов и связей. При этом обычно рассматриваются только скелетные атомы (со "стертыми" атомами водорода) и связи между ними.

Теоретико-графовое описание молекул хорошо отражает целостность, относительные размеры молекул и характер связывания атомов в ней (наличие цепей, циклов, разветвлений), но не учитывает метрические отношения (длины связей, валентные и азимутальные углы). Графы можно представлять в цифровом виде - в форме квадратных матриц. Наиболее часто применяются матрицы смежности A и матрица расстояний D . Элементы матрицы смежности вершин a_{ij} равны или единице, или нулю в зависимости от того, связана ли ребром вершина i графа G с вершиной j или не связана. При этом число единиц в i -й строке или j -м столбце матрицы равно степени вершины. Каждый элемент D_{ij} матрицы расстояний представляет число ребер, соединяющих вершину i с вершиной j наикратчайшим путем.

Помимо упомянутых матриц смежности и расстояний используют также и другие матрицы: матрицу связности S , матрицу смежности ребер A^e , расширенную матрицу расстояний E , Матрицу Винера W , путевую матрицу P , реберную матрицу расстояний D^e , геометрическую матрицу G , топографическую матрицу T и др.

Так как вид матрицы зависит от способа нумерации вершин (или ребер) графа, для количественного описания структуры молекул используют не сами матрицы, а их инварианты, называемые топологическими индексами (инвариант графа - это число, определяемое по его топологической матрице и не зависящее от способа нумерации вершин графа). Обычно ТИ получают с помощью различных математических операций (сложение, вычитание, умножение, деление, возведение в степень и др.) над элементами матриц, описывающих молекулу в целом или ее отдельные структурные фрагменты.

Выбор топологических индексов для конкретного исследования считается самым сложным среди этапов построения корреляционной зависимости «структура-свойство». Предложено много ТИ, но отсутствуют какие-либо определенные правила или рекомендации, регламентирующие выбор индексов, лучше всего подходящих в данном случае. Обычно исследователю приходится полагаться на собственную интуицию и перечень необходимых качеств ТИ: индекс должен быть удобным в использовании, универсальным, адекватно описывать структуру соединения, иметь ясный физический смысл и хорошо коррелировать с рассматриваемым свойством. В настоящее время для оценки качества ТИ применяются два основных критерия:

- дискриминирующая способность (ДС) (способность "различать" соединения некоторого класса); для оценки дискриминирующей способности можно использовать формулу $DC=n/N$, где n - число различных значений индекса в массиве из N структур;
- корреляционная способность (КС); один ТИ считается "лучше" другого, если полученное с его помощью корреляционное уравнение является более точным; при этом точность корреляционного уравнения оценивается с помощью стандартных статистических характеристик - коэффициента корреляции r , стандартного отклонения s и др.

Качество корреляционных моделей зависит от многих факторов: корреляционной и дискриминирующей способностей ТИ, вида корреляционного уравнения, природы исследуемых соединений и свойств и др. Влияние этих факторов на точность корреляций требует внимательного изучения на примере различных соединений и свойств.

В третьей главе «Исследование топологических индексов на примере алканов» изучена корреляционная способность широкого ряда ТИ на примере физико-химических свойств алканов (энтальпии образования, температуры кипения, молярного объема). Выбор именно этих соединений и свойств обусловлен надежностью и полнотой экспериментальных данных о них, поскольку теоретико-графовой подход относится к числу феноменологических методов и требуется достаточно большое количество надежных реперных величин.

Для изучения были выбраны тридцать наиболее известных и применяемых в QSPR/QSAR топологических индексов. Рассмотрены, в

частности, следующие ТИ: углеродное число (n), индекс Винера (W), молекулярно-топологический индекс Шульца (МТИ), индекс полной смежности Бартона (A'), индексы Загребской группы ученых (M_1, M_2), индексы связности ($\chi^1, \chi^2, \chi^{3p}, \chi^{3c}, \chi^{4p}, \chi^{4pc}, \chi^{4c}$), числа путей определенной длины ($p_1, p_2, p_3, p_4, p_5, p_6, p_7$) и число троек смежных ребер (R), реберный и квадратичный аналоги индекса Винера (W^1 и W' , соответственно), индекс Балабана (J), индекс Хосойи (Z), число Харари (H) и др.

Для исследуемых соединений (первые сорок алканов с числом атомов углерода от 1 до 8) были рассчитаны значения топологических индексов и построены различные корреляционные зависимости между ТИ и свойствами. Были исследованы, в частности:

- двухпараметрические зависимости на основе отдельных ТИ, при этом в качестве аппроксимирующих функций были испробованы:

1. $y=ax+b$	7. $y=a+b/n(x)$
2. $y=x/(a+bx)$	8. $y=ax/(b+x)$
3. $y=ab^x$	9. $y=a+bx^{1/2}$
4. $y=a\exp(bx)$	10. $y=ax^2/(b+x)$
5. $y=1/(a+b\exp(-x))$	11. $y=ax^2+bx+c$
6. $y=ax^b$	

- многопараметрические корреляционные зависимости на основе комбинаций разных ТИ в виде множественной линейной регрессии $y=a_0+\sum a_i x_i$, с разным числом параметров.

Зависимости строились как для всего набора алканов, так и отдельно для изомеров октана для того, чтобы проследить как влияет состав рассматриваемых соединений на корреляционную способность ТИ. Отдельно было проведено сравнительное исследование корреляций со свойствами для трехмерных индексов и их двумерных аналогов. Точность корреляционных уравнений оценивались на основе статистических параметров - коэффициента корреляции r и стандартного отклонения s . Расчет ТИ, построение корреляционных моделей и их анализ проводились с помощью стандартной программы Windows 98 - электронных таблиц Microsoft Excel.

В результате исследования было установлено, что из тридцати рассмотренных ТИ ряд индексов отличается стабильно высокой корреляционной способностью (табл. 1): самые высокие коэффициенты корреляции с энтальпией образования имеют индексы n, A', p_1 ($r=0,987$), с

температурой кипения – χ^1 , n , A' , p_1 ($r=0,991=0,983$), с молярным объемом – n , A' , p_1 ($r=0,993$). Индексы n , A' , p_1 просто рассчитываются и имеют достаточно ясный физический смысл (n – число атомов углерода в молекуле соединения, p_1 – число связей, A' – сумма валентностей всех атомов, входящих в молекулярный граф), но очень низкую дискриминирующую способность (табл. 1). Поэтому они не пригодны для корреляций со свойствами изомеров. Вместе с тем, использование более информативных ТИ с высокой дискриминирующей способностью (ДС) еще не гарантирует лучшие результаты при установлении взаимосвязей структура – свойство, а низкая ДС не является помехой для получения точных корреляционных уравнений (табл. 2). Например индексы (p_1 , p_2 , ...) с низкой дискриминирующей способностью (ДС~0,2, где $0 \leq ДС \leq 1$) коррелируют со свойствами не хуже, а в ряде случаев лучше, чем более информативные индексы связности (χ^1 , χ^2 , ...) с ДС~0,9. Поэтому представляется разумным одновременное использование ТИ с разной информативностью.

Таблица 1. Дискриминирующая (ДС) и корреляционная (КС) способности топологических индексов алканов

Индекс	ДС	КС			Индекс	ДС	КС		
		$-\Delta_f H^\circ$	Ткип	Вмол			$-\Delta_f H^\circ$	Ткип	Вмол
n	0,2	0,987	0,983	0,993	S_1	0,625	0,970	0,906	0,946
W	0,9	0,904	0,893	0,913	S_2	0,625	0,977	0,922	0,957
МТІ	0,95	0,908	0,895	0,917	K_3	0,2	0,962	0,931	0,953
A'	0,2	0,987	0,983	0,993	χ^1	0,95	0,943	0,991	0,978
p_1	0,2	0,987	0,983	0,993	χ^2	0,95	0,914	0,786	0,874
p_2	0,3	0,935	0,824	0,894	χ^{3p}	0,9	0,807	0,820	0,787
p_3	0,25	0,843	0,826	0,818	χ^{3c}	0,575	0,487	0,272	0,366
p_4	0,175	0,695	0,696	0,770	χ^{4p}	0,75	0,662	0,686	0,744
p_5	0,125	0,392	0,426	0,465	χ^{4c}	0,2	0,303	0,115	0,173
p_6	0,075	0,183	0,260	0,255	χ^{4pc}	0,65	0,498	0,403	0,539
R	0,175	0,559	0,368	0,453	W^3	1	0,894	0,891	0,909
M_1	0,425	0,982	0,911	0,956	МТІ ³	1	0,912	0,901	0,921
M_2	0,7	0,957	0,888	0,928	H	1	0,989	0,960	0,981
W^c	0,9	0,852	0,853	0,874	J	1	0,889	0,851	0,869
W'	1	0,797	0,811	0,833	Z	0,9	0,817	0,844	0,847

Таблица 2. Корреляционные модели «структура-свойство» алканов

Переменные	$-\Delta_f H^\circ$, кДж/моль			Ткип, °С			Вмол, мл/моль		
	Вид зав.	r	s	Вид зав.	r	s	Вид зав.	r	s
W	6	0,963	9,42	6	0,994	5,45	6	0,992	3,20
МТІ	11	0,966	9,04	6	0,993	5,93	6	0,993	2,97
A'	11	0,984	6,16	11	0,991	6,72	6	0,996	2,38
M_1	11	0,989	5,13	11	0,939	17,10	11	0,968	6,50
S_2	6	0,995	3,55	11	0,972	11,77	6	0,989	3,93
K_3	6	0,983	6,38	6	0,990	7,00	6	0,996	2,38
χ^1	6	0,942	11,65	11	0,997	4,06	8	0,980	5,14
p_1	11	0,984	6,16	11	0,991	6,72	6	0,996	2,38
χ^1, χ^2	МЛР	0,997	2,73	МЛР	0,993	6,07	МЛР	0,995	2,69
p_1, p_2	МЛР	0,993	4,16	МЛР	0,993	5,87	МЛР	0,992	3,41
S_1, S_2	МЛР	0,986	5,89	МЛР	0,996	4,41	МЛР	0,993	3,03
p_1, W	МЛР	0,988	5,49	МЛР	0,987	8,01	МЛР	0,992	3,23
$p_1, p_2, p_3, p_4, p_5, p_6, R$	МЛР	0,998	2,25	МЛР	1,000	1,65	МЛР	0,999	1,46
$\chi^1, \chi^2, \chi^{3p}, \chi^{3c}, \chi^{4p}, \chi^{4c}, \chi^{4pc}$	МЛР	0,998	2,38	МЛР	0,997	3,90	МЛР	0,998	1,97
W, МТІ, $\chi^1, \chi^2, p_1, p_2, R$	МЛР	0,999	2,06	МЛР	1,000	1,27	МЛР	0,999	1,39

Как показал анализ, точность корреляционных моделей зависит от вида аппроксимирующей функции (табл. 3). Среди одиннадцати испробованных зависимостей во всех рассмотренных случаях лучшими оказались квадратичная $y=ax^2+bx+c$ ($r_{cp}=0,981$) и степенная $y=ax^b$ ($r_{cp}=0,982$) функции, худшими $y=1/(a+b \exp(-x))$ ($r_{cp}=0,702$) и $y=ax^2/(b+x)$ ($r_{cp}=0,369$).

Таблица 3. Коэффициенты корреляции для уравнений разного вида

Свойство	№ формулы										
	1	2	3	4	5	6	7	8	9	10	11
$-\Delta_f H^\circ$	0,956	0,966	0,934	0,934	0,674	0,974	0,958	0,961	0,900	0,428	0,976
Ткип	0,957	0,966	0,939	0,939	0,703	0,982	0,973	0,962	0,895	0,182	0,982
Вмол	0,964	0,983	0,941	0,941	0,730	0,989	0,980	0,981	0,907	0,497	0,985

Использование нескольких индексов в виде множественной линейной регрессии $y = a_0 + \sum a_i(\text{ТИ})_i$ значительно эффективнее, чем корреляции одиночных индексов (табл. 2); для одиночных ТИ $r_{\text{ср}} = 0,982$, $r_{\text{max}} = 0,996$; для регрессии - $r_{\text{ср}} = 0,990$, $r_{\text{max}} = 1$. Использование двух независимых индексов вместо их суммы, например (p_1 , p_2) вместо ($p_1 + p_2$), приводит к уменьшению стандартной ошибки на 30-40%.

С ростом числа параметров в корреляционном уравнении его точность увеличивается (табл. 2). Среди моделей с одинаковым числом ТИ в качестве параметров лучшая точность достигается в случае применения индексов, имеющих наиболее высокие коэффициенты корреляции с исследуемым свойством (табл. 1).

Трехмерные индексы Винера и Шульца рассчитывались с учетом пространственной структуры молекул и отличаются от двумерных аналогов более высокой дискриминирующей способностью (табл. 1). В то же время, точность корреляционных моделей, построенных на их основе, не превышает точность моделей на основе двумерных индексов.

По результатам проведенного исследования были составлены таблицы и диаграммы, отражающие основные свойства и практические возможности рассмотренных ТИ, а также влияние различных факторов на точность корреляций. Представленные в них данные могут быть полезны в качестве справочного материала при выборе топологических индексов для исследования и построении корреляционных зависимостей.

В четвертой главе «Исследование топологических индексов на примере соединений разных классов» проведено исследование корреляционной способности топологических индексов на примере нескольких классов органических соединений, отличающихся от рассмотренных в третьей главе алканов более сложной структурой (наличием разных функциональных групп с атомами азота, кислорода, кратными связями). Были рассмотрены, в частности: алкены, спирты, альдегиды и кетоны, кислоты, амины, производные бензола. Исследование проводилось на примере как физико-химических свойств (энтальпии образования, температуры кипения, молярного объема), так и биологической активности (токсичности) соединений. В процессе выполнения исследования был проведен отбор групп соединений с известными значениями свойств, рассчитаны значения топологических индексов для всех соединений, построены корреляционные зависимости «структура – свойство».

Анализ полученных зависимостей показал, что лучше всего с физико-химическими свойствами соединений разных классов коррелируют простые топологические индексы A' , p_1 , M (A' - индекс полной смежности Бартона, равный сумме степеней всех вершин молекулярного графа; p_1 - число путей длины один; M - молекулярная масса соединения), прямо пропорциональные числу атомов в молекуле соединения. Более сложные ТИ (W , MTI), хотя и отличаются более высокой дискриминирующей способностью, но коррелируют в среднем на 8-10% хуже (рис. 1).

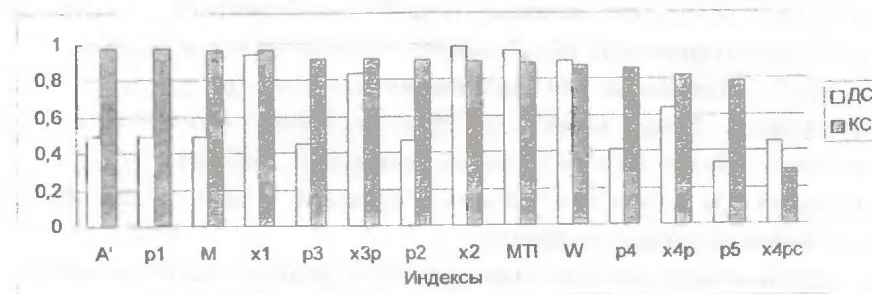


Рис. 1. Соотношение дискриминирующей (ДС) и корреляционной (КС) способностей ТИ.

Общая картина корреляций для разных классов соединений примерно одинакова: сохраняется порядок распределения индексов по корреляционной способности (A' , p_1 , M , x^1 , p^3 , x^3p , p^2 , x^2 , MTU , W , p_4 , x^{4p} , p_5 , x^{4pc} , R , x^{3c}) и характер корреляции свойств (лучше коррелирует с ТИ молярный объем - $r_{\text{ср}} = 0,958$, немного хуже температура кипения - $r_{\text{ср}} = 0,925$, хуже всего энтальпия образования - $r_{\text{ср}} = 0,918$). При этом с температурами кипения рассмотренных соединений лучше других ТИ коррелирует индекс связности χ^1 , отражающий степень ветвления молекулы (он принимает максимальные значения для наименее разветвленных структур и минимальные - для наиболее разветвленных).

Сравнение статистических параметров корреляционных уравнений разных видов подтвердило сделанные в третьей главе выводы о зависимости точности корреляций от вида функции, используемой при аппроксимации. Во всех рассмотренных случаях наиболее точно описывают корреляционную связь между ТИ и свойствами степенная ($y = ax^b$) и квадратичная ($y = ax^2 + bx + c$) функции, наименее точны функции вида $y = 1/(a + b \exp(-x))$ и $y = ax^2/(b + x)$. Использование нелинейных зависимостей

(вместо линейных) снижает стандартное отклонение на 22% для энтальпии образования, на 17% для температуры кипения, на 30% для молярного объема (в среднем по всем группам соединений).

Как и в случае алканов, применение комбинации из нескольких топологических индексов в виде множественной линейной регрессии эффективнее корреляционных моделей на основе отдельных ТИ. В некоторых случаях нелинейные корреляции отдельных индексов также отличаются достаточно высокой точностью: например, корреляции χ^1 с температурами кипения кислот ($r=0,999$, $s=2,6^\circ\text{C}$) и индекса r_1 с молярными объемами алкенов ($r=0,997$, $s=2,2\text{мл/моль}$). Точность регрессионного уравнения увеличивается с ростом числа параметров.

Было установлено, что комбинации индексов (r_1 , r_2 , r_3 , ..., R), отличающихся более низкой дискриминирующей способностью по сравнению с более информативными индексами связности (χ^1 , χ^2 , ...), практически для всех рассмотренных соединений и свойств дают более точные корреляционные уравнения.

Был исследован вопрос о применимости топологических индексов в корреляциях со свойствами неродственных соединений. Для этого были построены корреляционные зависимости между ТИ и свойствами различных серий органических соединений, содержащих одновременно молекулы с разными функциональными группами. Были последовательно рассмотрены наборы веществ, включающие соединения двух, трех, четырех и пяти разных классов (табл. 4). Анализ полученных зависимостей показал, что коэффициенты корреляции ТИ в рядах неродственных соединений в среднем на 15-20% ниже, чем для гомологов (рис.2).

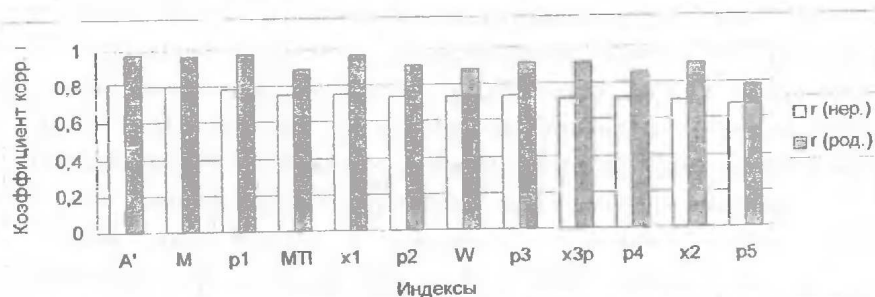


Рис. 2. Средние значения коэффициентов корреляции ТИ со свойствами в рядах родственных и неродственных соединений.

Таблица 4. Коэффициенты корреляции ТИ со свойствами (в рядах неродственных соединений)

Свойство	Соединения	Коэффициенты корреляции r					
		A'	МП	W	χ^1	r_1	M
$-\Delta_f H^\circ$, кДж/моль	Спирты	0,959	0,834	0,824	0,940	0,959	0,959
	Спирты, алкены	0,578	0,433	0,380	0,342	0,356	0,432
	Спирты, алк., кетоны	0,657	0,485	0,413	0,383	0,428	0,488
	Сп., алк., кет., кислоты	0,793	0,590	0,486	0,363	0,489	0,539
	Сп., алк., кет., кисл., амины	0,806	0,612	0,502	0,388	0,514	0,553
$T_{\text{кип}}$, $^\circ\text{C}$	Спирты	0,964	0,947	0,938	0,978	0,964	0,964
	Спирты, алкены	0,907	0,771	0,745	0,787	0,794	0,835
	Спирты, алк., кетоны	0,886	0,806	0,775	0,816	0,829	0,860
	Сп., алк., кет., кислоты	0,923	0,812	0,775	0,757	0,826	0,856
	Сп., алк., кет., кисл., амины	0,926	0,820	0,785	0,777	0,840	0,864
$V_{\text{моль}}$, мл/моль	Спирты	1,000	0,929	0,920	0,996	1,000	1,000
	Спирты, алкены	0,879	0,836	0,850	0,963	0,969	0,943
	Спирты, алк., кетоны	0,811	0,849	0,867	0,968	0,966	0,946
	Сп., алк., кет., кислоты	0,700	0,792	0,851	0,972	0,949	0,928
	Сп., алк., кет., кисл., амины	0,722	0,783	0,848	0,974	0,953	0,937

В случае энтальпии образования наблюдается резкое падение корреляционной способности ТИ (в два раза); для температуры кипения и молярного объема уменьшение не столь значительно (6% и 1%, соответственно). Было показано, что на основе различных комбинаций ТИ можно получить корреляционные зависимости средней точности для энтальпии образования ($r = 0,972$, $s = 41$ кДж/моль) и температуры кипения ($r = 0,966$, $s = 18^\circ\text{C}$) и более точные для молярного объема ($r = 0,998$, $s = 2,45$ мл/моль). При этом использование температуры кипения в качестве одного из параметров корреляционного уравнения позволяет на 30% снизить стандартное отклонение для энтальпии образования ($r = 0,986$, $s = 29,2$ кДж/моль).

Корреляции топологических индексов с биологической активностью изучались на примере токсичности производных бензола LC_{50} (LC_{50} - однократная доза, которая поражает 50% испытываемых организмов, мг/кг). Набор из пятидесяти соединений включал моно-, ди- и тризамещенные бензола с разными функциональными группами ($-\text{NH}_2$, $-\text{NO}_2$, $-\text{CH}_3$, $-\text{Cl}$, $-\text{Br}$, $-\text{OH}$) в качестве заместителей.

При исследовании корреляционных зависимостей между ТИ и токсичностью разных производных бензола было обнаружено, что коэффициенты корреляции для всего набора соединений имеют низкие значения ($r_{cp}=0,312$, $r_{max}=0,634$), а для рядов родственных соединений - в среднем в два раза выше ($r_{cp}=0,642$, $r_{max}=0,895$) (табл. 5, рис. 3). Лучше всего коррелирует с LC_{50} молекулярная масса ($r_{cp}=0,691$, $r_{max}=0,895$), остальные ТИ несколько хуже ($r_{cp}=0,57$, $r_{max}=0,6\pm 0,88$).

Таблица 5. Коэффициенты корреляции ТИ со свойствами разных производных бензола.

Индексы	Все производные (50)			Хлорпроизводные (14)			Фенолы (12)			Толуолы (14)		
	LC_{50}	Ткип	V_{mol}	LC_{50}	Ткип	V_{mol}	LC_{50}	Ткип	V_{mol}	LC_{50}	Ткип	V_{mol}
W	0,179	0,351	0,687	0,602	0,930	0,587	0,545	0,874	0,906	0,647	0,866	0,862
МТИ	0,192	0,473	0,351	0,262	0,708	0,620	0,482	0,850	0,574	0,531	0,755	0,862
М	0,634	0,503	0,175	0,895	0,949	0,338	0,624	0,336	0,248	0,613	0,907	0,865
χ^1	0,293	-0,024	0,664	0,882	0,900	0,562	0,662	0,748	0,922	0,457	-0,129	0,235
χ^2	0,303	-0,022	0,681	0,863	0,862	0,586	0,649	0,769	0,935	0,443	-0,120	0,198
χ^{3c}	0,319	0,015	0,707	0,835	0,835	0,594	0,635	0,775	0,952	0,465	-0,053	0,282
$\chi_{зам}$	0,330	0,110	0,522	0,790	0,903	0,448	0,718	0,731	0,872	0,494	0,160	0,727
p1	0,289	0,386	0,666	0,718	0,977	0,524	0,660	0,805	0,909	0,682	0,746	0,861
p2	0,289	0,386	0,666	0,718	0,977	0,524	0,660	0,805	0,909	0,682	0,746	0,861
R	0,289	0,386	0,666	0,718	0,977	0,524	0,660	0,805	0,909	0,682	0,746	0,861

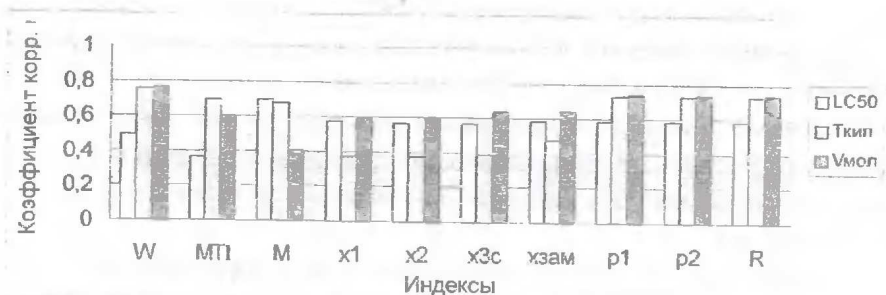


Рис. 3. Коэффициенты корреляции ТИ со свойствами производных бензола (средние значения для разных групп соединений).

Точность корреляционных моделей, построенных на основе различных комбинаций ТИ в виде множественной линейной регрессии невысока: для лучшей модели из десяти ТИ $r=0,722$, $s=0,46$ (табл. 6).

Таблица 6. Модели «структура-свойство» для производных бензола

Свойство	Индексы	Разные производные бензола		Только хлорпроизводные бензола	
		r	s	r	s
LC_{50} мг/кг	W, p1, p2, R	0,488	0,54	0,838	0,24
	МТИ, p1, p2, R	0,299	0,59	0,896	0,19
	М, p1, p2, R	0,652	0,47	0,896	0,19
	W, M, p1, p2, R	0,691	0,45	0,903	0,20
	$\chi^1 \cdot \chi^2 \cdot \chi^{3c} \cdot \chi_{зам}$	0,414	0,56	0,912	0,18
	W, МТИ, М, $\chi^1 \cdot \chi^2 \cdot \chi^{3c} \cdot \chi_{зам} \cdot p1 \cdot p2, R$	0,722	0,46	0,949	0,14
	T, W, МТИ, М, $\chi^1 \cdot \chi^2 \cdot \chi^{3c} \cdot \chi_{зам} \cdot p1 \cdot p2, R$	0,828	0,41		

Введение температуры кипения в уравнение регрессии в качестве переменной позволило на 10% улучшить статистические характеристики: $r=0,828$, $s=0,41$. Для сравнения: по литературным данным использование более информативных молекулярных дескрипторов других типов (квантово-химических, геометрических и др.) приводит к получению корреляционной модели с $r=0,827$.

Анализ корреляционных зависимостей между ТИ и свойствами разных органических соединений показывает, что топологические индексы, несмотря на ограниченность заложенной в них структурной информации, обладают довольно высокой корреляционной способностью и могут быть полезны в прогнозировании различных свойств органических соединений.

Выводы

- Проведено систематическое исследование корреляционных зависимостей между топологическими индексами и свойствами органических соединений. Исследование впервые выполнено для широкого ряда ТИ на примере как физико-химических свойств, так и биологической активности органических соединений разных классов (алканов, алкенов, спиртов, альдегидов и кетонов, кислот, аминов, производных бензола).

- Подобраны группы органических соединений с известными значениями свойств; выполнен расчет топологических индексов для всех рассматриваемых соединений, построены корреляционные зависимости между ТИ и свойствами и проведен их сравнительный анализ.
- Установлено, что качество корреляционных моделей зависит от ряда факторов (природа и состав рассматриваемых соединений, вид корреляционного уравнения, корреляционная способность ТИ) и выявлен характер этой зависимости. Показано, что низкая дискриминирующая способность некоторых ТИ не является помехой для получения точных корреляционных моделей на их основе. Установлено, что порядок распределения индексов по корреляционной способности примерно сохраняется для разных групп соединений и свойств.
- Изучен характер зависимости между точностью корреляционного уравнения и его видом. Найдено, что наиболее точно описывают корреляционные связи между отдельными ТИ и свойствами соединений нелинейные функции вида $y=ax^b$ и $y=ax^2+bx+c$. При этом использование нелинейных корреляционных моделей на 20-30% снижает стандартное отклонение по сравнению с линейными зависимостями. Показано, что применение комбинаций нескольких ТИ в виде множественной линейной регрессии значительно эффективнее корреляционных моделей на основе отдельных индексов.
- Исследованы корреляции между ТИ и свойствами соединений в негомологичных рядах. Установлено, что коэффициенты корреляции для негомологичных соединений в среднем на 15-20% ниже чем для гомологов; при этом в случае энтальпии образования и токсичности наблюдается резкое снижение корреляционной способности ТИ (почти в два раза), а в случае температуры кипения и молярного объема уменьшение не столь значительное (на 6% и 1%, соответственно).
- Рассмотрена на примере физико-химических свойств алканов корреляционная способность трехмерных индексов и установлено, что, несмотря на более высокую дискриминирующую способность и информативность трехмерных индексов по сравнению с их

двумерными аналогами, точность корреляционных моделей на их основе не превышает показатели двумерных ТИ.

- На примере корреляций ТИ с токсичностью производных бензола показано, что использование топологических индексов, несмотря на ограниченность заложенной в них структурной информации, позволяет получать корреляционные модели, не уступающие по точности зависимостям, построенным на основе более сложных и информативных молекулярных дескрипторов (геометрических, квантово-химических, стереоэлектронных).

Основное содержание диссертации отражено
в следующих публикациях:

1. Папулов Ю.Г., Виноградова М.Г., Смоляков В.М., Куплок Л.В., Мальшева Ю.А. Использование трехмерного числа Винера в корреляциях структура-свойство // Уч. зап. тверского гос. университета. Тверь: ТвГУ, 1996. Т.1. С.186.
2. Papulov Yu.G., Vinogradova M.G., Malysheva Yu.A. Use of topological indices in stucture-property correlations: history, present situation and prospects // Abstracts of international scientific conference "Mathematical models of non-linear excitation, transport, dynamics, control in condensed systems and other mediums". Tver, 1996. P. 25.
3. Папулов Ю.Г., Виноградова М.Г., Мальшева Ю.А., Ботов А.Б. Топологические индексы в корреляциях «структура-свойство» //Труды II Международной научной конференции "Математические модели нелинейных возбуждений, переноса, динамики, управления в конденсированных системах и др. средах". Тверь:ТГТУ,1997. С. 160-163.
4. Vinogradova M.G., Papulov Yu.G., Smolyakov V.M., Malysheva Yu.A., Polyakov M.N. Application of graph theory to calculation thermodynamical properties of hydrocarbons // Abstracts of Intern. conf. on chemical and biological thermodynamics. India, Amritsar, 1997, P. 35.
5. Мальшева Ю.А., Папулов Ю.Г., Виноградова М.Г., Ботов А.Б., Смоляков В.М. Свойства и строение органических молекул. 1. Трехмерные топологические индексы алканов // Журнал структурной химии. 1998. Т.39, №3. С.482-492.

6. *Мальшиева Ю.А., Папулов Ю.Г., Виноградова М.Г., Давыдова И.Г.* Свойства и строение органических молекул. 2. Теоретико-графовое изучение алкенов и спиртов // Журнал структурной химии. 1998. Т.39, №3. С. 490-499.
7. *Мальшиева Ю.А., Папулов Ю.Г.* Использование разных видов матриц молекулярного графа для описания структуры конформационно гибких соединений // Уч. зап. ТвГУ. 1998. Т.4. С.43-47.
8. *Мальшиева Ю.А., Папулов Ю.Г., Березкин А.В.* Сравнительное изучение методов расчета физико-химических свойств алканов // Свойства веществ и строение молекул. Тверь: ТвГУ, 1998. С. 20-33.
9. *Ботов А.Б., Мальшиева Ю.А., Виноградова М.Г.* Использование топологичных индексов для расчета физико-химических свойств спиртов, аминов и алкенов // Свойства веществ и строение молекул. Тверь: ТвГУ, 1998. С. 34-39.
10. *Мальшиева Ю.А., Папулов Ю.Г., Виноградова М.Г.* Количественные корреляции «структура-свойство»: теоретико-графовое изучение //Тезисы докладов III Международной научной конференции "Математические модели нелинейных возбуждений, переноса, динамики, управления в конденсированных системах и других средах". Тверь: ТГТУ, 1998. С. 89.
11. *Мальшиева Ю.А., Папулов Ю.Г., Ботов А.Б., Березкин А.В.* Теоретико-графовый подход в исследованиях корреляций «структура – свойство» //Тезисы докладов I Всероссийской конференции "Молекулярное моделирование". М: ГЕОХИ РАН, 1998. С. 31.

Технический редактор Т.В.Малахова
Изд. лиц. ЛР № 020268 от 03.04.1997 г.
Подписано в печать 11.09.2001. Формат 60 x 84 1/16.
Бумага офсетная. Печать офсетная.
Усл.печ.л. 1,5. Уч.-изд.л. 1,19.
Тираж 100 экз. Заказ № 385.
Тверской государственный университет,
Редакционно-издательское управление.
Адрес: Россия, 170000, г. Тверь, ул. Желябова, 33.
Тел. РИУ: (0822) 42-60-63.