

Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего
образования
«ТВЕРСКОЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «ТвГУ»)

На правах рукописи



ФЕДИНА ЮЛИЯ АЛЕКСЕЕВНА

КОЛИЧЕСТВЕННЫЕ МОДЕЛИ В КОРРЕЛЯЦИЯХ «СТРУКТУРА –
СВОЙСТВО» УГЛЕВОДОРОДОВ И ИХ ЗАМЕЩЕННЫХ

Специальность 02.00.04 - Физическая Химия

Диссертации на соискание ученой степени

кандидата химических наук

Научный руководитель:
доктор химических наук, профессор,
Виноградова Марина Геннадьевна

Тверь - 2017

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	5
ГЛАВА 1. ТЕОРЕТИЧЕСКИЕ ОБОСНОВАНИЯ ДЛЯ РЕШЕНИЯ ПРОБЛЕМЫ.	12
1.1. Теория графов для построения моделей «структура-свойство».	12
1.2. Теория Пойя для решения проблем изоморфизмов.	17
1.3. Принципы метода QSPR.	23
1.4. Топологические индексы в качестве молекулярных дескрипторов.	24
1.5. Топологический индекс среднего расстояния в качестве молекулярного дескриптора.	28
1.6. Используемые расчетные методы физико-химических свойств полициклических ароматических углеводородов.	29
ГЛАВА 2. ПОСТРОЕНИЕ QSPR МОДЕЛЕЙ ДЛЯ АЛКАНОВ.	31
2.1. Структура и свойства алканов.	31
2.2. Построение QSPR моделей для расчета температуры кипения алканов с использованием топологических индексов.	37
2.3. Построение QSPR моделей для расчета температуры кипения алканов с применением нового топологического индекса D_{ADTI} .	43
2.4. Построение QSPR моделей для расчета энтальпии образования алканов в газовой фазе с учетом топологических индексов.	47

2.5. Построение QSPR - моделей для расчета энтальпии образования алканов в газовой фазе с учетом индекса среднего расстояния. 54

2.6. Проверка статистической гипотезы дискриминирующей способности индекса среднего расстояния. 57

ГЛАВА 3. ПОСТРОЕНИЕ QSPR МОДЕЛЕЙ ДЛЯ ЦИКЛОАЛКАНОВ. 59

3.1. Структура и свойства циклоалканов. 59

3.2. Построение QSPR моделей для расчета температуры кипения циклоалканов с учетом топологических индексов. 62

3.3. Построение QSPR моделей для расчета температуры кипения циклоалканов с учетом индекса среднего расстояния. 66

3.4. Построение QSPR моделей для расчета энтальпии образования циклоалканов в газовой фазе с учетом топологических индексов. 69

3.5. Построение QSPR - моделей для расчета энтальпии образования циклоалканов в газовой фазе с учетом топологического индекса среднего расстояния. 72

3.6. Проверка статистической гипотезы дискриминирующей способности индекса среднего расстояния. 76

ГЛАВА 4. ПОСТРОЕНИЕ QSPR - МОДЕЛЕЙ ДЛЯ ПОЛИЦИКЛИЧЕСКИХ АРОМАТИЧЕСКИХ УГЛЕВОДОРОДОВ. 78

4.1. Строение и свойства полициклических ароматических углеводородов. 78

4.2. Построение QSPR моделей для расчета температуры плавления ПАУ с учетом топологических индексов. 82

4.3. Построение QSPR - моделей для расчета температуры плавления ПАУ с учетом индекса среднего расстояния.	88
4.4. Построение QSPR моделей для расчета температуры кипения ПАУ с учетом топологических индексов.	91
4.5. Построение QSPR моделей для расчета температуры кипения ПАУ с применением индекса среднего расстояния.	95
4.6. Построение QSPR моделей для расчета коэффициента распределения ПАУ в системе октанол-вода с учетом топологических индексов.	97
4.7. Построение QSPR моделей для расчета коэффициента распределения ПАУ в системе октанол-вода с учетом индекса среднего расстояния.	106
4.8. Построение QSPR моделей для предсказания энтальпии образования ПАУ в газовой фазе.	109
4.9. Построение QSPR-моделей для определения энтальпии образования ПАУ состоящих из бензольных колец.	119
4.10. Проверка статистической гипотезы дискриминирующей способности индекса среднего расстояния.	128
ВЫВОДЫ.	130
ЛИТЕРАТУРА, ОПУБЛИКОВАННАЯ ПО ТЕМЕ ДИССЕРТАЦИИ.	132
СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ	135

ВВЕДЕНИЕ

Очевидно, что предопределение характеристик и поведения химических веществ, исходя из их молекулярной структуры, продолжает оставаться одной из наиболее актуальных и волнующих вопросов современных ученых в области теоретической химии. Выполнение этой задачи первоначально было сконцентрировано в основном в области биологии и токсикологии. В настоящее время активно развиваются методы для прогнозирования физико-химических свойств. На основе фундаментальной теории химии начали устанавливать структуры природных веществ и осуществлять синтез органических соединений. Органические соединения — соединения, имеющие в своём составе углерод, а также водород либо замещающие его элементы (азот, кислород, сера и пр.). Атомы углерода, как правило, имеют тенденцию формировать устойчивые структуры в органических соединениях, в основе которых лежат одинарные, двойные, тройные и кольцевые углерод - углеродные связи. Большинство органических молекул состоит из двух частей: фрагмента, который в ходе реакции остаётся без изменения, и группы, подвергающейся при этом превращениям. Органические соединения входят в состав биологических тканей и являются основой жизни на Земле. Все органические соединения, в зависимости от молекулярной структуры, классифицируются на две группы - циклические и ациклические. Ациклические соединения называют жирными или алифатическими. Они имеют нециклическую или цепную форму. Одними из первых хорошо изученных представителей такого типа были природные жиры, откуда и пришло это название. Циклические соединения подразделяются на группы, одна из которых представлена молекулами, содержащими кольца из углеродных атомов. Эта карбоциклические соединения. Другая группа, где молекулы состоят из колец, содержащих кроме углерода атомы других элементов (азота, серы, кислорода и др.), классифицируются как гетероциклические.

Карбоциклические соединения можно разделить на такие группы, как алициклические (предельные и непредельные), похожие по свойствам на алифатические, и ароматические, которые содержат бензольные кольца. В состав органических соединений кроме углерода и водорода могут также входить и другие элементы, в одиночном виде или в форме функциональных групп – связанных атомов, которые существенно влияют на химические свойства и поведение того или иного класса веществ. Присутствие функциональных групп относит каждое соединение к определенному классу, и такая расстановка значительно упрощает их изучение. Для описания молекул и их химических свойств усиливается тенденция шире использовать математические методы, которые включают исчисление изомеров, применение комбинаторики для описания химических соединений, определение информационного содержания химических графов, поиск зависимости «структура-свойство». Для осуществления этой задачи возникло направление QSPR (Quantitative Structure Property Relationship; дословный перевод – количественное соотношение структура-свойство). В русском языке нет краткой аббревиатуры и используется английский эквивалент. Существует поверье, что все свойства молекулы предопределены ее структурой, а установление количественной зависимости «структура-свойство» нередко помогает доступным способом раскодировать их. Это способствует расширению познаний в том, как структура предопределяет свойства соединений, их поведение и ответную реакцию в виде способности вступать во всевозможные преобразования и формирования новых веществ. Особый интерес к проблеме «структура-свойство» объясняется изобилием новых полученных к данному времени веществ, а также уникальных возможностей и путей их использования. Осуществление и воплощение данной миссии, даже в рамках одного класса соединений, приблизит исследователей к поставленной цели качественного прогнозирования свойств гипотетических молекул, осуществлять синтез новых соединений с интересующим набором свойств. Поиск зависимостей между

характером изменения разнообразных свойств соединений и их структурой представляется чрезвычайно актуальной. Обнаруженные закономерности можно применять для предсказания свойств химических веществ, определения новых соединений с нужным набором характеристик, а также для систематизации молекул определенного гомологического ряда. На текущий момент опубликовано много количественных зависимостей «структура-свойство» (QSPR-моделей), предложенных в основном для гомологичных групп органических соединений. Но вопрос моделирования «структура-свойство», заключающийся в определении свойств разнообразных органических веществ, продолжает оставаться частично открытым. Точные руководящие методики разработки интересующих моделей в определенном виде до настоящего времени не определены, а также, многие модели задействуют параметры (дескрипторы), требующие комплексных непростых расчетов, но не приводящие к заметному улучшению статистических характеристик модели по сравнению с дескрипторами, не требующими экстенсивных затрат. Можно заключить что, изучение проблемы моделирования связи «структура-свойство» и расширение познаний в области построения QSPR-моделей для разнообразных выборок органических соединений выступает весьма интересующей и прогрессивной задачей.

Актуальность проблемы.

Первые шаги в стремлении связать молекулярное строение со свойствами прослеживаются с 19-го века. Одним из успешных достижений того времени было установление влияния молекулярной массы на растворимость воды. В начале 20 столетия Овертон и Майер определили влияние водной токсичности по отношению к распределению в системе жир-вода. На протяжении всего 20-го века это направление прогрессировало. В 1962 году выявился огромный интерес к прогнозированию биологической активности. Первоначально, эта тенденция просматривалась в основном в области медицинской химии и разработки

лекарственных препаратов, но в 1970-х и 1980-х годах, с увеличением экотоксикологических проблем, количественное моделирование «структура-активность» и «структура-свойство» экологической токсичности начало расти, особенно после того, как были вовлечены регулирующие органы. С тех пор это направление продолжает расширяться, более чем 1400 трудов публикуется только в международном журнале QSPR ежегодно с 2011 года.

Аналогично принципам QSAR и QSPR, взаимосвязь «структура-токсичность» (QSTR) и «структура-фармакокинетика» (QSPkR) используются, когда моделирование применяется к токсикологическим или фармакокинетическим системам. QSAR, QSPR, QSTR и QSPkR опираются на предположение, что структурно подобные соединения проявляют аналогичную ответную реакцию на воздействие. Таким образом, эти методы предлагают прогностические и диагностические возможности. Моделирование зависимостей между структурой и свойством соединений позволяют разработать гипотетическую структуру с заданным набором свойств, провести верификацию наличия интересующих характеристик в соединении, проверить присутствие того или иного качества или активности, оценить ответную реакцию всего гомологического ряда на определенное воздействие. QSPR неопределимо в тех ситуациях, когда экспериментальные действия невозможны по многочисленным причинам, включая трудности с очисткой, летучесть, опасность для здоровья. Использование QSPR-моделирования помогает решать наиболее сложные проблемы в поиске новых и изучении синтезированных соединений и требует создания новых алгоритмов и дескрипторов с более высокой дискриминирующей способностью.

Работа выполнена при финансовой поддержке Минобрнауки Российской Федерации в рамках выполнения государственных работ в сфере научной деятельности (проект №4.1325.2014/К), а также НИР №1266, выполненной в рамках базовой части ГЗ в сфере научной деятельности по заданию №220.216

Цель работы. Работа направлена на исследование возможностей разработки и имплементации нового дескриптора QSPR-моделирования в виде топологического индекса для предсказания значений интенсивных и экстенсивных характеристик углеводородов.

Для осуществления поставленной цели были осуществлены следующие **задачи**.

1. Систематизация и вывод изомеров замещения бензола с использованием основных положений теории графов и теоремы Пойя.
2. Определены объекты испытаний: алканы, циклоалканы и полициклические ароматические углеводороды.
3. В качестве экзаменуемых интенсивных и экстенсивных свойств соединений выбраны температура плавления, температура кипения, коэффициент распределения в системе октанол-вода, энтальпия образования в газовой фазе
4. Разработан новый топологический индекс среднего расстояния, основываясь на матрице расстояний, который послужил экспериментальным фактором. Определены контрольные параметры в виде широко известных и протестированных топологических индексов.
5. Созданы контрольная группа и экспериментальная группа QSPR-моделей, построенных с использованием отобранных факторов в качестве независимой переменной величины.
6. Протестирована дискриминирующая способность разработанного индекса посредством сравнительного анализа качества между контрольной и экспериментальной группами QSPR-моделей и проверкой статистических гипотез.
7. Создана и протестирована многопараметровая QSPR-модель для расчета энтальпии образования в газовой фазе группы представителей полициклических ароматических углеводородов
8. Рассчитана энтальпия образования полициклических ароматических углеводородов в газовой фазе с использованием построенной QSPR-модели.

Научная новизна и практическая значимость.

Разработан новый топологический индекс среднего расстояния.

Впервые проведена и проиллюстрирована детальная процедура качественной проверки дискриминирующей способности предложенного индекса.

Проанализирована группа представителей полициклических ароматических углеводородов с примененным структурным критерием соединенных горизонтальных цепей бензольных колец и разработана QSPR-модель для расчета энтальпии образования в газовой фазе для отобранных представителей этого класса.

Интерес к поиску влияния структуры соединений на свойства и поведение продолжает расти. Предлагаются новые дескрипторы и исследуются различные свойства. Разработанный топологический индекс и детальное описание тестирования его эффективности в прогнозировании качеств выбранных объектов ярко продемонстрировал потенциал QSPR-моделирования. Данный индекс может быть использован для прогноза свойств в однопараметровой или в многопараметровой линейной регрессии. Проведенный эксперимент проиллюстрировал возможности разработки новых оптимизированных QSPR-моделей и прогнозирования гипотетических структур с заложенными свойствами.

Достоверность результатов исследования обеспечена применением широко известного и применяемого метода линейно – регрессионного анализа с расчетом статистических величин и проверкой статистической гипотезы по отношению к контрольной группе моделей построенных с учетом многократно тестированных топологических индексов.

Личный вклад автора. Автор предложил и разработал процедуру расчета индекса и принял активное участие в разработке и анализе качества моделей.

Основные положения, выносимые на защиту:

1. Обоснование и систематизация подхода к расчету интенсивных и экстенсивных характеристик углеводородов и их замещенных.
2. Моделирование “структура – свойство” с использованием разработанного топологического индекса среднего расстояния.
3. Прогнозирование и анализ энтальпии образования объектов исследования и анализ тенденции изменения этих величин.

Апробация работы. Материалы диссертации доложены и обсуждены на Международной научной конференции “Моделирование нелинейных процессов и систем” (Москва, 2008); на VI Региональных Менделеевских чтениях (Удомля, 2010); на VII Региональных Менделеевских чтениях (Удомля, 2011); на II Международной конференции “Моделирование нелинейных процессов и систем” (Москва, 2011); на II ежегодной Всероссийской научной Интернет - конференции с международным участием “Химическая наука: современные достижения и историческая перспектива” (Казань, 2014); на III международной научной конференции “Моделирование нелинейных процессов и систем” (Москва, 2015); на VI Международной научной конференции “ХИМИЧЕСКАЯ ТЕРМОДИНАМИКА И КИНЕТИКА” (Тверь, 30 мая – 3 июня 2016 г.)

Публикации. По теме диссертации опубликованы 8 научных статей и 9 тезисов докладов на конференциях.

Объем и структура работы. Диссертация изложена на 130 страницах, иллюстрирована 45 рисунками и включает 20 таблиц. Состоит из введения, обзора основных направлений по теме исследования, глав, описывающих построение моделей и оценку их возможностей, выводов и списка использованной литературы включившего 184 источника.

ГЛАВА 1. ТЕОРЕТИЧЕСКИЕ ОБОСНОВАНИЯ ДЛЯ РЕШЕНИЯ ПРОБЛЕМЫ.

1.1. ТЕОРИЯ ГРАФОВ ДЛЯ ПОСТРОЕНИЯ МОДЕЛЕЙ «СТРУКТУРА-СВОЙСТВО».

Первый и самый важный шаг в QSAR / QSPR моделировании является численно закодировать химические структуры различных молекул таким образом, чтобы построить модель корреляции между химическими структурами различных химических соединений и соответствующих химических и биологических свойств. Следовательно, наиболее точное представление химической формулы (или молекулярного графа) в цифровом формате является одной из основных задач в QSAR / QSPR исследованиях. Есть много методов количественного представления молекулярных структур, в которых топологический индекс является наиболее популярным, так как он может быть получен непосредственно из молекулярных структур и быстро вычислен для большого числа молекул. Теория графов имеет важную значимость для моделирования «структура-свойство». В молекулярном графе, вершина представляет собой атом, а ребро символизирует связь. Таким образом, молекулярные графы могут быть легко выражены с помощью матриц. Химические графы дают возможность прогнозировать химические превращения, пояснять сущность и систематизировать некоторые основные понятия химии: структуру, конфигурацию, конформации, квантовые и статистико-механические взаимодействия молекул, изомерию и т.д. Разработано более 400 видов топологических индексов со дня рождения первого. Топологический индекс может быть использован для оценки структурное сходство и разнообразия.

Рассмотрим и применим некоторые основные понятия о молекулярных графах. Граф – это совокупность точек (вершин) и совокупность пар этих точек

(не обязательно всех), соединенных линиями между собой. Дерево - связный неориентированный граф, не содержащий циклов или контуров. Графы называют изоморфными, если существует взаимно однозначное соответствие между совокупностями их вершин и ребер (дуг).

Для решения задач теории графов и ее приложений графы представляют с помощью матриц (смежности, инцидентности, двустрочных и др.), а также специальных числовых характеристик. Например, в матрице смежности строки и столбцы отвечают номерам вершин графа, а ее элементы принимают значения 0 и 1 (соответственно отсутствие и наличие дуги между данной парой вершин); в матрице инцидентности строки отвечают номерам вершин, столбцы - номерам дуг, а элементы принимают значения 0, + 1 и - 1 (соответственно отсутствие, наличие дуги, входящей в вершину и выходящей из нее). Применение теории графов базируется на построении и анализе различных классов химических графов, учитывающих только характер связи вершин. Дуги (ребра) и вершины этих графов отображают химические понятия, явления, процессы или объекты и соответственно качественные и количественные взаимосвязи либо определенные отношения между ними. В стереохимии органических веществ наиболее часто используют молекулярные графы, которые содержат только те вершины, соответствующие атомам углерода. Составление наборов молекулярных деревьев и установление их изоморфизма позволяют определять молекулярные структуры и находить полное число изомеров алканов, алкенов и алкинов.

Молекулярные графы дают возможность сводить задачи, связанные с кодированием, номенклатурой и структурными особенностями (разветвленность, цикличность и т. п.) молекул различных соединений, к анализу и сопоставлению чисто математических признаков и свойств молекулярных графов и их деревьев, а также соответствующих им матриц. Для выявления количественных корреляций между строением молекул и физико-химическими (в т. ч. фармакологическими) свойствами соединений разработано много топологических индексов молекул

(Винера, Балабана, Хосойи, Плата, Рандича и др.), которые определяют с помощью матриц и числовых характеристик молекулярных деревьев.

Таблица 1.1.1. Топологические индексы и их определения. [136]

Индекс (Ссылка)	Определение
Автокоррелирующие дескрипторы (Moreau and Broto 1980)	$T = (T_0, T_1, T_2) T_n = g(i) * g(j)$, где $g(i)$ и $g(j)$ вклады со стороны атомов i и j .
Индекс Балабана J (Balaban 1982)	$J = \sum (\bar{S}_i \bar{S}_j)^{-1/2} = q \sum (\bar{S}_i \bar{S}_j)^{-1/2}$, где \bar{S}_i и \bar{S}_j средняя величина суммы расстояний вершин V_i и V_j ; $q = n_e / (\mu + 1)$, n_e – количество ребер и μ - количество циклов.
Индекс гибкости связи (Lieth et al. 1996)	$\rho_{KB} = \sum \Phi_i - \Phi + 1$, где Φ_i – гибкость фрагмента и Φ – гибкость всей молекулы.
Центральный индекс (Balaban 1979, and Hu et al. 2003a)	$B = \sum_i \delta_i^2$ $C = 1/2(B - 2n + U)$, где δ_i - вершины на каждом шаге и U величина Кронекера в зависимости от четности числа вершин n . $CI = \sum_{i=1}^n DF_i$, $BF_i = \sum_{j=1}^n L_j * V_j * D_{kj}$ где I_j - внутреннее состояние атома j , V_j - степень вершины атома j и D_{ij} - расстояние между атомами i и j .
Обходной индекс (Lukovits 1996, Razinger 1997)	$\omega = (1/2) \sum_i \sum_j (\Delta)_{ij}$, где Δ_{ij} – элемент обходной матрицы

Таблица 1.1.1.

Индекс (Ссылка)	Определение
Индекс связности граней (Estrada 1995)	$\varepsilon = \sum_r ((\delta(e_i)(\delta(e_j)))_r)^{-0.5}$ где $\delta(e_i)$ и $\delta(e_j)$ – степень ребер e_i и e_j .
Индекс электропии (Yee et al. 1977)	$\varepsilon = \log_2 \frac{N!}{\prod N_i!}$, где N количество атомов в молекуле и N_i среднее количество атомов одного вида.
Индекс E-состояния (Kier and Hall 1990)	$S_i = I_i + \sum_j \Delta I_{ij}$, где I_i - значение внутреннего состояния атома i ; ΔI_{ij} - возмущение I_j на I_i в виде $\Delta I_{ij} = (I_i - I_j)D_{ij}^2$.
Расширенный ЕС индекс (Estrada et al. 1998)	$m_{\varepsilon_i}(G) = \sum_s \prod_i [\delta(e_i)]_s^{-0.5}$, где $\delta(e_i)$ - степень ребра e_i .
Расширенный Индекс Винера (Estrada et al. 1988)	${}^k W = \sum_{i=k-1}^l i(i-1)(i-2)...(i-(k-1))\eta_i$, где η_i число пар вершин на расстоянии i .
Индекс гибкости (Kier 1989)	$\Phi = {}^1 k_\alpha {}^2 k_\alpha / A$, где A число атомов; ${}^1 k_\alpha$ и ${}^2 k_\alpha$ индексы Каппа первого и второго порядка.
Фрагмент WI (Mekenyuan et al. 1988)	$W.EFTI(G) = \sum_{i=1}^N W / EFTI(v_i)$, индекс отражает взаимодействие между отрезанным фрагментом F и остальной частью молекулы ($G - F$).
GSI (Gordon and Scantlebury 1964)	$2N_2 = \sum_{i,j}^n (V_i + V_j) - 2$, где V_i и V_j степень вершины атомов i и j .

Таблица 1.1.1.

Индекс (Ссылка)	Определение
Индекс Харари (Plavsic et al. 1993)	$H = (1/2) \sum_{i=1}^N \sum_{j=1}^N (D^r)_{ij}.$
Индекс Хосои (Hosoya 1971)	$Z(G) = \sum_{K=0}^{n/2} P(G, K),$ <p>где $P(G, K)$ - комбинации, в которых ребра графа могут быть выбраны таким образом, чтобы никакие два из них не являлись смежными.</p>
Индекс гипер-Винер (Klein et al. 1995)	$WW(G) = \frac{1}{2} \sum_{i < j} ([D]_{ij}^2 + [D]_{ij})$
Идентификационные числа (Randić 1977)	$ID = \frac{1}{2} \left[N + \sum_{i=1}^T W_i \right]$
Индекс Каппа (Kier 1985, 1986)	${}^1k = 2^1 P_{\max} {}^1 P_{\min} / ({}^1 P_i)^2,$ ${}^2k = 2^2 P_{\max} {}^2 P_{\min} / ({}^2 P_i)^2,$ ${}^3k = 4^3 P_{\max} {}^3 P_{\min} / ({}^3 P_i)^2,$ <p>где ${}^1 P_i$, ${}^2 P_i$, и ${}^3 P_i$ – количество путей 1, 2 и 3; ${}^1 P_{\min}$ и ${}^1 P_{\max}$ - количество одинарных путей линейного и полного графа; ${}^2 P_{\min}$, и ${}^2 P_{\max}$ - число двухсвязных путей линейного и звездного графа; ${}^3 P_{\min}$, and ${}^3 P_{\max}$ - число трехсвязных путей линейного и двойного звездного графа.</p>
Индекс Кирхгофа (number) (Klein and Randić 1993)	$K_f = \frac{1}{2} \sum_i \sum_j \Omega_{ij},$ <p>где Ω_{ij} - (i, j) элемент матрицы сопротивления.</p>

Таблица 1.1.1.

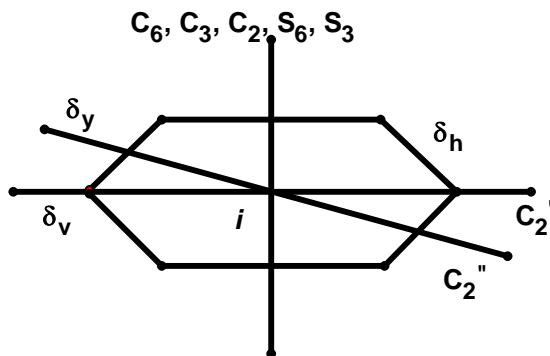
Индекс (Ссылка)	Определение
Индекс молекулярной связанности (Randić 1975, and Kier and Hall 1976)	$\chi = \sum (V_i V_j)^{-1/2},$ ${}^k \chi_p = \sum_{all..edges} (V_i V_j \dots V_k)^{-1/2}$ <p>где $V_i, V_j,$ и V_k – степени вершин атомов $i, j,$ и k.</p>
Молекулярный топологический индекс (Schultz 1989)	$MTI = \sum_{k=1}^n E_k,$ <p>где E_i ряд матрицы, состоящий из $v(A+D)$, где v – степень вершины.</p>
Общая связанность (Bonchev 1997)	$TC(G) = \sum_{k=1}^K A_k (G_k \in G) = \sum_{k=1}^K \sum_{i=1}^{N_k} a_i$
Индекс Квази-Винера (Mohar et al. 1993)	$W^*(G) = N \sum_{i=1}^{N-1} \frac{1}{S_p(L)_i}$ <p>где $S_p(L)$ - спектр матрицы Лапласа.</p>
Сцегед индекс (Khadikar et al. 1995)	$S_z(G) = \sum_{e_{ij} \in E(G)} n_i n_j$
Индекс Винера (Wiener 1947)	$W(G) = 1/2(\text{sum}(\text{sum}(d_{kj})))$
Индексы Загребской группы (Gutman et al. 1975)	$M1 = \sum_{i=1}^n V_i^2 \quad M2 = \sum_{all..edges} V_i V_j$ <p>где $V_i,$ и V_j – степени вершин атомов $i,$ и j.</p>

1.2. ТЕОРИЯ ПОЙЯ ДЛЯ РЕШЕНИЯ ПРОБЛЕМ ИЗОМОРФИЗМОВ.

Определение числа различных графов данного класса — это один из основных вопросов, возникающих о классе графов. Изучая теорию перечисления графов и отвечая на этот вопрос, приходится решать проблемы изоморфизмов для многих классов графов. Теорема перечисления Пойя, когдга впервые стала

широко цениться в начале 1960-х годов, служила основным инструментом для решения проблем изоморфизмов. Воспользуемся теоремой Пойя для вывода изомеров замещения бензола.

Пусть $\Omega = \{1, 2, 3, 4, 5, 6\}$ - множество возможного замещения бензола.



Точечная группа симметрии бензола

$D_{6h}: E, 2C_6, 2C_3, C_2, 3C_2', 3C_2'', i, 2S_6, 2S_3, \sigma_h, 3\sigma_v, 3\sigma_d$ индуцирует на множестве подстановки вида:

$$E \rightarrow f_1^6 = (1)(2)(3)(4)(5)(6) \text{ - шесть единичных циклов}$$

$$C_6 \rightarrow f_6^1 = \begin{cases} (123456), \\ (165432) \end{cases} \text{ - подстановка с одним циклом длины шесть}$$

$$C_3 \rightarrow f_3^2 = \begin{cases} (135)(246), \\ (153)(264) \end{cases} \text{ - подстановка с двумя циклами длины три}$$

$$C_2 \rightarrow f_2^3 = (14)(25)(36) \text{ - подстановка с тремя циклами длины два}$$

$$C_2' \rightarrow f_1^2 f_2^2 = \begin{cases} (1)(4)(26)(35) \\ (2)(5)(13)(46) \\ (3)(6)(24)(15) \end{cases} \text{ - подстановка с двумя единичными циклами и}$$

двумя циклами длины два

$$C_2'' \rightarrow f_2^3 = \begin{cases} (12)(63)(54) \\ (14)(23)(65) \\ (16)(25)(34) \end{cases} \text{ - подстановка с тремя циклами длины два}$$

$$i \rightarrow f_2^3$$

$$S_6 \rightarrow f_6^1$$

$$S_3 \rightarrow f_3^2$$

$$\sigma_h \rightarrow f_1^6$$

$$\sigma_y \rightarrow f_1^2 f_2^2$$

$$\sigma_d \rightarrow f_2^3$$

Цикловой индекс группы подстановок можно представить как многочлен:

$$Z_{D_{6h}} = \frac{1}{12} (f_1^6 + 2f_6^1 + 2f_3^2 + 4f_2^3 + 3f_1^2 f_2^2).$$

Производящая функция числа изомеров замещенных бензола будет иметь вид

$$\begin{aligned} \Phi_{D_6} = & \frac{1}{12} [(h + x + \dots + y)^6 + 2(h^6 + x^6 + \dots + y^6) + 2(h^3 + x^3 + \dots + y^3)^2 + \\ & 4(h^2 + x^2 + \dots + y^2)^3 + 3(h + x + \dots + y)^2 (h^2 + x^2 + \dots + y^2)^2] = h^6 + h^5 x + \\ & \dots + 3h^4 x^2 + \dots + 3h^4 xy + \dots + 3h^3 x^3 + \dots + 6h^3 x^2 y + \dots + 10h^3 xyz + \dots \\ & + 11h^2 x^2 y^2 + \dots + 16h^2 x^2 yz + \dots + 30h^2 xyzu + \dots + 60hxyzuv \end{aligned}$$

Можно заметить что изомеры замещенных бензола можно разбить на 11 семейств по количеству разбиений числа 6 на целые и положительные части:

- | | |
|------------------|-----------------------------|
| 1) h^6 | $6 = 6$ |
| 2) $h^5 x$ | $6 = 5 + 1$ |
| 3) $h^4 x^2$ | $6 = 4 + 2$ |
| 4) $h^4 xy$ | $6 = 4 + 1 + 1$ |
| 5) $h^3 x^3$ | $6 = 3 + 3$ |
| 6) $h^3 x^2 y$ | $6 = 3 + 2 + 1$ |
| 7) $h^3 xy^2$ | $6 = 3 + 1 + 2$ |
| 8) $h^2 x^2 y^2$ | $6 = 2 + 2 + 2$ |
| 9) $h^2 x^2 yx$ | $6 = 2 + 2 + 1 + 1$ |
| 10) $h^2 xyzu$ | $6 = 2 + 1 + 1 + 1 + 1$ |
| 11) $hxyzuv$ | $6 = 1 + 1 + 1 + 1 + 1 + 1$ |

Эти семейства содержат соответственно 1, 1, 3, 3, 3, 6, 10, 11, 16, 30, 60 изомеров.

Число семейств данного типа (например, $h^5 x, h^5 y, x^5 y \dots$) соответственно дается следующим образом:

$$\tau_1 = n + 1, \tau_2 = n(n + 1), \tau_3 = n(n + 1), \tau_4 = \frac{1}{2}(n - 1)n(n + 1),$$

$$\tau_5 = \frac{1}{2}n(n + 1), \tau_6 = (n - 1)n(n + 1), \tau_7 = \frac{1}{6}(n - 2)(n - 1)n(n + 1),$$

$$\tau_8 = \frac{1}{6}(n - 1)n(n + 1), \tau_9 = \frac{1}{9}(n - 2)(n - 1)n(n + 1),$$

$$\tau_{10} = \frac{1}{24}(n - 3)(n - 2)(n - 1)n(n + 1),$$

$$\tau_{11} = \frac{1}{720}(n - 4)(n - 3)(n - 2)(n - 1)n(n + 1), \text{ где } n - \text{ число заместителей разных}$$

видов.

Теперь можно узнать количество изомеров в семействах среди X-, XY-, ...

замещенных бензола, отраженных в таблице 1.2.1 и 1.2.2:

$$\gamma_1 = \tau_1, \gamma_2 = \tau_2, \gamma_3 = 3\tau_3, \gamma_4 = 3\tau_4, \gamma_5 = 3\tau_5, \gamma_6 = 6\tau_6,$$

$$\gamma_7 = 10\tau_7, \gamma_8 = 11\tau_8, \gamma_9 = 16\tau_9, \gamma_{10} = 30\tau_{10}, \gamma_{11} = 60\tau_{11}.$$

Общее число типов рассчитывается формулой:

$$\tau = C_{n+6}^6 = \frac{1}{720}(n + 6)(n + 5)(n + 4)(n + 3)(n + 2)(n + 1).$$

Общее число изомеров бензола выражается следующим образом:

$$\gamma = \tau_1 + \tau_2 + 3\tau_3 + 3\tau_4 + 3\tau_5 + 6\tau_6 + 10\tau_7 + 11\tau_8 + 16\tau_9 + 30\tau_{10} + 60\tau_{11}.$$

Таблица 1.2.1. Распределение типов замещения бензола.

n	Зам. Бензола	Общее число типов	Число типов по семействам										
			1	2	3	4	5	6	7	8	9	10	11
1	X	7	2	2	2	-	2	-	-	-	-	-	-
2	XY	28	3	6	6	3	3	6	-	1	-	-	-
3	XYZ	84	4	12	12	12	6	24	4	4	6	-	-
4	XYZU	210	5	20	20	30	10	60	20	10	30	5	-
5	XYZUV	462	6	30	30	60	15	120	60	20	90	30	1
6	XYZUVW	924	7	42	42	105	21	210	140	35	210	105	7
...
n	XYZUVW	τ	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	τ_7	τ_8	τ_9	τ_{10}	τ_{11}

Таблица 1.2.2. Распределение изомеров замещения бензола.

Зам. бензола	Общее число изомеров	Число изомеров замещения										
		1	2	3	4	5	6	7	8	9	10	11
X	13	2	2	6	0	3	0	0	0	0	0	0
XY	92	3	6	18	9	9	36	0	11	0	0	0
XYZ	430	4	12	36	36	18	144	40	44	96	0	0
XYZU	1505	5	20	60	90	30	360	200	110	480	150	0
XYZUV	4291	6	30	90	180	45	720	600	220	1440	900	60
XYZUVW	10528	7	42	126	315	63	1260	1440	385	3360	3150	420

1.3. ПРИНЦИПЫ МЕТОДА QSPR.

Количественные соотношения структура-активность (QSARs) или структура-свойство (QSPRs) представляют математические модели, которые пытаются связать структурно-зависимую характеристику соединения с его биологической или физико-химической активностью. Очевидной становится практическая ценность общих топологических подходов для решения химических задач. Рассмотрим теоретическую основу методики построения количественных моделей связи «структура-свойство» для органических соединений. Для определения взаимосвязи между структурой и свойством применяются многочисленные методы математического моделирования, уместность которых зависит от природы исследуемых химических соединений и экзаминируемых свойств. QSPR занимает значительное место для решения этой задачи. В настоящее время QSPR – это компьютерное правило, описывающее свойство вещества через дескрипторы. Дескриптор химической структуры – это число или набор чисел, которые характеризуют структуру органического соединения. QSPR работает по следующей схеме. Первоначально из группы соединений с известными физико-химическими характеристиками выбирают два набора: один тренировочный и другой тестовой. Все соединения обладают определенным свойством, полученным из соответствующего эксперимента и выраженным численным значением. Далее количественные значения отраженного свойства соотнесены с конкретной структурой. На следующем этапе выбирается набор дескрипторов и методами статистики устанавливается корреляция величин «свойства» для соединений тренировочного набора с дескрипторами и получается так называемое QSPR - уравнение. Это уравнение отражает функциональную зависимость свойств от дескрипторов. Правильность этого уравнения проверяется на тестовом наборе структур. Для этого вычисляют дескрипторы для каждой тестовой структуры, рассчитывают свойства по QSPR - уравнению и сравнивают

их с экспериментальными значениями. Если для тестового набора наблюдается хорошее совпадение экспериментальных и предсказанных величин, то данное QSPR - уравнение можно применить для прогнозирования свойств новых, еще не синтезированных структур.

1.4. ТОПОЛОГИЧЕСКИЕ ИНДЕКСЫ В КАЧЕСТВЕ МОЛЕКУЛЯРНЫХ ДЕСКРИПТОРОВ.

Для выбора дескрипторов используют многочисленные направления. В настоящее время среди этих подходов специальное место занимают топологические подходы, использующие только информацию, содержащуюся в структурной формуле исследуемого соединения [15]. В таких подходах химическую структурную формулу представляют молекулярным графом, вершины которого соответствуют атомам (ядрам), а ребра химическим связям молекулы. При этом, как правило, рассматриваются только скелетные атомы (атомы водорода обычно не включаются в граф) и связи между ними. Каждый молекулярный граф можно представить либо матрицей, либо полиномом, либо числовым индексом [17]. Представление структурной формулы в виде числового значения, часто называемого топологическим индексом (ТИ), может осуществляться несколькими способами [18–20]. Наиболее часто применяется построение матрицы смежности $A(G)$ и матрицы расстояний $D(G)$ на графе. В таблице 1.1.1 перечислены многие топологические индексы, вычисляемые на основании матриц. [136] Топологические индексы (структурные дескрипторы) включают информацию о размере и форме молекулы, о соединении атомов и структурных групп в ней и их взаимном расположении. Кроме того, с их помощью можно учитывать особенности электронного и пространственного строения молекул, выбирая соответствующим образом веса вершин и ребер молекулярного графа. В настоящее время топологические индексы используются для кодирования

химической информации, оценки реакционной способности молекул, при планировании химического эксперимента, для количественного описания химических структур при анализе связи между структурой молекулы, ее свойствами, а также активностью. Способность различать очень близкие по строению и составу соединения делает ТИ одними из наиболее распространенных молекулярных дескрипторов в корреляционном анализе. Их преимущество состоит в том, что они не требуют сложных экспериментальных определений или квантово-химических расчетов, а вычисляются непосредственно из структурной формулы соединений. Это особенно важно, когда речь идет об оценке свойств гипотетических структур, для которых нет никаких данных. Однако имеется и очевидный недостаток — учитываются не все особенности молекулярного строения. Топологическим индексам трудно дать физико-химическую интерпретацию, поскольку их значения получают путем формальных операций над графами. Тем не менее, корреляционные соотношения, полученные с их помощью, даже без достаточно ясного физического смысла, могут оказаться полезными в решении различных практических вопросов [21]. Существует точка зрения [22], что нет необходимости какой-либо физической или физико-химической интерпретации топологических (или графово-теоретических) индексов, так как они являются математическими абстракциями, но в то же время должна существовать связь между топологией молекулы и физико-химическими свойствами соединения.

В работе была изучена корреляционная возможность ряда топологических индексов. Любой эксперимент нуждается в контрольной модели для валидации результатов. С этой целью, были использованы давно известные и многократно тестированные топологические индексы: индекс Винера, модифицированный индекс Винера, индекс связности Рандича, индекс Балабана и индекс Гутмана. Для улучшения статистических характеристик QSPR-моделей применялись модификации выше указанных индексов. Индекс Винера W определен как

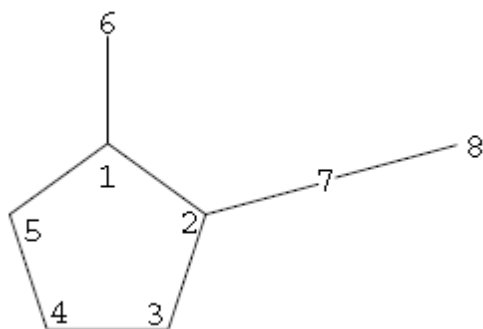
полусумма топологических расстояний между всеми N атомами в молекулярном графе и рассчитывался по формуле, указанной в таблице 1.1.1.

Индекс связности Рандича представляет собой математически закодированную информацию о числе атомов в молекуле, их связи между собой, о степени разветвления молекулы и может быть рассчитан для различных уровней связанности атомов молекулы между собой [23–25]. Индекс связанности $\chi = \chi(G)$ графа G описан в таблице 1.1.1. Ссылки на индекс Рандича датируются 1975 годом, что говорит о многочисленных исследованиях с его применением. Известно о высокой корреляции различных физико-химических характеристик органических соединений с индексом Рандича. По всей вероятности, модели с применением индекса Рандича будут служить независимым экспертом для проверки надежности нового дескриптора.

Балабан предложил обладающий высокой дискриминирующей способностью топологический индекс, основанный на матрице расстояний молекулярного (химического) графа. Этот молекулярный дескриптор – индекс Балабана J – широко используется при исследованиях количественных корреляций «структура-свойство» или «структура-активность». Индекс Балабана рассчитывается как средняя сумма расстояний связанности по формуле [23-25] и описан в таблице 1.1.1. Индекс Балабана, аналогично индексу Рандича, цитируется в литературе с 1982 года. Это подтверждает уместность такого выбора для создания контрольной модели для проверки качества разработанного индекса среднего расстояния.

Для исследований физико-химических характеристик циклоалканов был использован индекс Гутмана, вычисляемый по формуле [24, 128], указанной в таблице 1.1.1. Милан Рандич изучал корреляцию этого индекса с физико-химическими свойствами циклоалканов и предложил модификацию для колец с нечетным количеством атомов углерода, что значительно улучшило результаты. Индекс Гутмана, Szeged (Sz) индекс, учитывает количество атомов, находящихся

ближе к рассматриваемому связывающему ребру с каждой стороны. Предусматривается для атомов, равноудаленных от пары связанных атомов, поровну разделить это количество атомов. Например, для 1-метил-2-этилциклопентана модифицированный индекс Винера и индекс Гутмана, Szeged (Sz) индекс, рассчитываются следующим образом:



Пара вершин (<i>i. j</i>)	$N_1 \times N_2$	Пара вершин (<i>i. j</i>)	$N_1 \times N_2$
(1, 2)	$3,5 \times 4,5 = 15,75$	(1, 2)	$3 \times 4 = 12$
(1, 5)	$5,5 \times 2,5 = 13,75$	(1, 5)	$5 \times 2 = 10$
(1, 6)	$7 \times 1 = 7$	(1, 6)	$7 \times 1 = 7$
(2, 3)	$5,5 \times 2,5 = 13,75$	(2, 3)	$5 \times 2 = 10$
(2, 7)	$6 \times 2 = 12$	(2, 7)	$6 \times 2 = 12$
(3, 4)	$5 \times 3 = 15$	(3, 4)	$4 \times 2 = 8$
(4, 5)	$3,5 \times 4,5 = 15,75$	(4, 5)	$2 \times 3 = 6$
(7, 8)	$7 \times 1 = 7$	(7, 8)	$7 \times 1 = 7$

Модифицированный индекс Винера 100

Индекс Гутмана Szeged (Sz) 72

Милан Рандич отметил, что исследователи не всегда могут найти качественную корреляцию с использованием известных топологических индексов и начинают выводить новые формулы [158]. Он заключил, что более целесообразно использовать модификацию существующих индексов, которые были многократно апробированы. В работе был использован индекс Гутмана для

четных колец и модифицированный индекс Винера для нечетных колец в исследовании циклоалканов.

Матрица Лапласа графа G определяется по формуле $L(G) = D(G) - A(G)$, где $A(G)$ матрица смежности и $D(G)$ диагональная матрица степеней вершин графа G . Инвариант графа на основе матрицы Лапласа вычисляется по формуле [191]:

$$LEL(G) = \sum_{i=1}^n \sqrt{\mu_i}, \quad (1)$$

где μ_i - собственные значения матрицы. Инвариант графа на основе матрицы применен для создания контрольной группы QSPR-моделей полициклических ароматических углеводородов.

1.5. ТОПОЛОГИЧЕСКИЙ ИНДЕКС СРЕДНЕГО РАССТОЯНИЯ В КАЧЕСТВЕ МОЛЕКУЛЯРНОГО ДЕСКРИПТОРА.

Разработан новый топологический индекс среднего расстояния с использованием матрицы расстояний. Формула расчета включает суммирование элементов матрицы расстояний и деление её на количество атомов углерода, входящих в состав соединения. Иными словами, предложенный индекс равен средней величине дистанций от любого атома ко всем атомам в структуре, описанной молекулярным графом. Этот молекулярный дескриптор использован для моделирования «структура – свойство» и создания экспериментальной группы.

$$D_{ADTI} = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N D_{ij}, \quad (2)$$

где N – количество атомов углерода в соединении, D_{ij} – элементы матрицы расстояний молекулярного графа G , представляющие наименее короткую дистанцию между вершинами i и j графа G .

Пример расчета. Построим матрицу расстояний для 2-метилпропана.

0	1	2	2
1	0	1	1
2	1	0	2
2	1	2	0

Сумма элементов матрицы равна 18. Делим эту величину на 4 (количество атомов углерода) и получаем 4,5. Средняя сумма расстояний от любого атома углерода до остальных атомов углерода в молекуле равна 4,5 единиц атомных связей.

1.6. ИСПОЛЬЗУЕМЫЕ РАСЧЕТНЫЕ МЕТОДЫ ФИЗИКО-ХИМИЧЕСКИХ СВОЙСТВ ПОЛИЦИКЛИЧЕСКИХ АРОМАТИЧЕСКИХ УГЛЕВОДОРОДОВ.

Полициклические ароматические углеводороды привлекли заметное внимание в последние десятилетия. Получение точных экспериментальных значений ПАУ осложняется проблемой очистки соединений при подготовке к исследованиям. Отсутствие экспериментальных данных подчеркивает практическую значимость надежных математических алгоритмов в определении величин физико-химических свойств ПАУ. Предлагаются многочисленные методы расчета характеристических величин, так как экспериментальные данные получены для немногих соединений этого класса. Для определения термодинамических величин ПАУ наиболее широко используется методы С. Бенсона и его дополненная версия С. Стейна, Н. Моисеевой и О. Дорофеевой, Д. Армитажа и К. Берда [66]. С. Стейн оставил группу, описанную С. Бенсоном [C_B (H)], и добавил три новых группы, определяющие атомы углерода, которые принадлежат к двум или трем ароматическим кольцам: [$C_{BF} - (C_B)_2(C_{BF})$], [$C_{BF} - (C_B)(C_{BF})_2$], и [$C_{BF} - (C_{BF})_3$]. Н. Моисеева и О. Дорофеева расширили свой

аддитивный метод, предложив три новые группы для углерода в пятичленных кольцах непредельных полициклических ароматических углеводородов. Д. Армитаж и К. Берд также вывели величину для $[C_{BF} - (C_{BF})_3]$ группы, принадлежащей к одному пятичленному кольцу, но другим способом. Надежность разработанных QSPR-моделей в работе оценена путем сравнения с вышеупомянутыми методами, которые послужили контрольной группой.

ГЛАВА 2. ПОСТРОЕНИЕ QSPR МОДЕЛЕЙ ДЛЯ АЛКАНОВ.

2.1. СТРУКТУРА И СВОЙСТВА АЛКАНОВ.

Для статистических расчетов использован программный пакет Fathom Dynamic Data Software, MS Excel и ActivStat. В работе использованы экспериментальные значения физико-химических характеристик структур из опубликованных литературных изданий [193 -195] и из базы данных Миланской Исследовательской Группы по Хемометрике и QSAR. В работе были построены корреляционные зависимости между структурными дескрипторами и физико-химическими параметрами для обучающего и тестового наборов исследуемых соединений. Эксперименты показали, что топологические индексы в зависимости от способа расчета и их порядка по-разному коррелируют с физико-химическими параметрами веществ.

Алканы представляют собой химические соединения, которые играют очень важную роль в нашей жизни. Они дают доступ к таким важным для нас товарам, как изделия из пластика, бензин, природный газ. Топливо очень важно для нас. Нам нужно топливо для автомобилей, для обогрева домов, и для многих других целей. Алканы являются простейшими из семейства углеводородов - соединений, содержащих только углерод и водород. Они содержат только углерод - водородные и углерод - углеродные простые связи и поэтому называются насыщенными углеводородами. Эта группа соединений составляет гомологический ряд с общей молекулярной формулой C_nH_{2n+2} , где n равно любому целому числу. Алканы - это углеводороды, которые имеют только sp^3 атомы углерода. Каждый атом углерода в молекуле окружен одиночными ковалентными связями и соединен с четырьмя атомами в тетраэдрической геометрии с величиной валентных углов $\sim 109.5^\circ$. Из-за этой геометрии и небольшой разницы в электроотрицательности между водородом и углеродом, алканы являются

неполярными молекулами. Все алканы с 4 или более атомами углерода в них имеют структурные изомеры. Это означает, что есть две или более различные структурные формулы, которые можно получить для каждой молекулярной формулы. Первые четыре представителя алканов являются газами при комнатной температуре. В твердом состоянии можно встретить молекулы с 17 и более атомами углерода.

Алканы содержат сильные углерод - углеродные простые связи и сильные углерод - водородные связи. Углерод - водородные связи являются лишь очень слабо полярными и поэтому нет никаких частей молекул, которые несут какое-либо значительное количество положительного или отрицательного заряда, к которому могут притягиваться другие частицы. Например, многие органические реакции инициируются ионом или полярной молекулой, притянутой к части органической молекулы, которая несет в себе определенный положительный или отрицательный заряд. Этого не происходит с алканами, поскольку их молекулы не имеют такого разделения заряда. Результирующий эффект состоит в том, что алканы имеют довольно ограниченный набор реакций. Они подвергаются горению, реагируют с некоторыми из галогенов, преодолев углерод – водородные связи; или же разрушаются углерод - углеродные связи.

Свойства алканов достаточно изучены, поэтому они являются хорошими объектами исследований и проверки достоверности предлагаемых прогнозирующих методов и моделей.

Таблица 2.1.1. Список алканов.

Номер	Название соединения	Номер	Название соединения
1	Этан	37	3-метилгептан
2	Пропан	38	2-метилгептан
3	2-МетилПропан	39	2,2,3,3-тетраметилпентан

Таблица 2.1.1.

Номер	Название соединения	Номер	Название соединения
4	Бутан	40	2,3,3,4-тетраметилпентан
5	2,2-диметилпропан	41	Октан
6	2-МетилБутан	42	2,2,3,4-тетраметилпентан
7	Пентан	43	2,3диметил-3-этилпентан
8	2,2-диметилбутан	44	2,2,4,4-тетраметилпентан
9	2,3-диметилбутан	45	3,3,4-триметилгексан
10	3-метилпентан	46	3,3-диэтилпентан
11	2-метилпентан	47	2,2-диметил-3-этилпентан
12	Гексан	48	2,3,3-триметилгексан
13	2,2,3-триметилбутан	49	2,4-диметил-3-этилпентан
14	3,3-диметилпентан	50	2,2,3-триметилгексан
15	2,2-диметилпентан	51	2,3,4-триметилпгексан
16	2,3-диметилпентан	52	2,4,4-триметилгексан
17	2,4-диметилпентан	53	3-метил-3-этилгексан
18	3-этилпентан	54	2,2,4-триметилгексан
19	3-метилгексан	55	2,3,5-триметилгексан
20	2-метилгексан	56	4,4-диметилгептан
21	Гептан	57	2,2,5-триметилгексан
22	2,2,3,3-тетраметилбутан	58	3,3-диметилгептан
23	2,3,3-триметилпентан	59	3,4-диметилгептан
24	2,2,3-триметилпентан	60	2-метил-4-этилгексан
25	3-метил-3-этилпентан	61	3,5-диметилгептан
26	2,3,4-триметилпентан	62	2,3-диметилгептан
27	2,2,4-триметилпентан	63	2,4-диметилгептан
28	3,3-диметилгексан	64	4-этилгептан

Таблица 2.1.1.

Номер	Название соединения	Номер	Название соединения
29	2-метил-3-этилпентан	65	2,2-диметилгептан
30	3,4-диметилгексан	66	2,5-диметилгептан
31	2,3-диметилгексан	67	3-этилгептан
32	2,2-диметилгексан	68	2,6-диметилгептан
33	2,4-диметилгексан	69	4-метилоктан
34	3-этилгексан	70	3-метилоктан
35	2,5-диметилгексан	71	2-метилоктан
36	4-метилгептан	72	Нонан

Таблица 2.1.2. Данные для алканов: $T_{кип}$ – температура кипения, $\Delta_f H^\circ_{(г,298К)}$ – энтальпия образования в газовой фазе, W - индекс Винера, J - индекс Балабана, χ - индекс Рандича, D_{ADTI} – индекс среднего расстояния.

Номер	$T_{кип}, ^\circ C$	$\Delta_f H^\circ_{(г,298К)}, кДж/моль$	W	J	χ	D_{ADTI}
1	-88,5	-83,45	1	1	2	1,00
2	-44,5	-104,67	4	1,633	1,41	2,67
3	-10,5	-134,67	9	2,3238	1,73	4,50
4	-0,5	-125,66	10	1,9747	1,91	5,00
5	9,5	-167,95	16	3,0237	2	6,40
6	27,9	-153,68	18	2,5395	2,27	7,20
7	36,5	-146,77	20	2,1906	2,41	8,00
8	49,7	-186,1	28	3,1685	2,56	9,33
9	58,1	-178,3	29	2,9935	2,64	9,67
10	63,2	-172,1	31	2,7542	2,81	10,33
11	60,2	-174,8	32	2,6272	2,77	10,67
12	68,7	-167,03	35	2,3391	2,91	11,67

Таблица 2.1.2.

Номер	$T_{кип}, ^\circ\text{C}$	$\Delta_f H^\circ_{(г, 298\text{К})}, \text{кДж/моль}$	W	J	χ	D_{ADTI}
13	80,9	-204,5	42	3,5412	2,94	12,00
14	86	-201,2	44	3,3604	3,12	12,57
15	79,2	-205,9	46	3,1545	3,06	13,14
16	89,8	-198	46	3,1442	3,18	13,14
17	80,5	-201,7	48	2,9532	3,13	13,71
18	93,5	-189,3	48	2,9923	3,35	13,71
19	91,8	-191,3	50	2,8318	3,31	14,29
20	90	-194,6	52	2,6783	3,27	14,86
21	98,4	-187,1	56	2,4475	3,41	16,00
22	106	-225,5	58	4,0204	3,25	14,50
23	114,6	-216,3	62	3,7083	3,5	15,50
24	110,5	-220	63	3,6233	3,48	15,75
25	118,2	-214,9	64	3,5832	3,68	16,00
26	113,4	-217,3	65	3,4642	3,55	16,25
27	99,3	-224	66	3,3889	3,41	16,50
28	112	-220	67	3,3734	3,62	16,75
29	115,6	-211	67	3,3549	3,72	16,75
30	118,7	-212,8	68	3,2925	3,72	17,00
31	115,3	-213,8	70	3,1708	3,68	17,50
32	107	-224,6	71	3,1118	3,52	17,75
33	109,4	-219,2	71	3,0988	3,66	17,75
34	118,9	-210,7	72	3,0744	3,85	18,00
35	108,4	-222,5	74	2,9278	3,63	18,50
36	117,7	-212	75	2,9196	3,81	18,75
37	118,8	-212,5	76	2,8621	3,81	19,00

Таблица 2.1.2.

Номер	$T_{кип}, ^\circ\text{C}$	$\Delta_f H^\circ_{(г, 298\text{К})}, \text{кДж/моль}$	W	J	χ	D_{ADTI}
38	117,6	-215,4	79	2,7158	3,77	19,75
39	140,3	-237,1	82	4,1447	3,81	18,22
40	141,5	-236,1	84	4,0137	3,89	18,67
41	125,8	-208,7	84	2,5301	3,91	21,00
42	133	-236,9	86	3,8776	3,85	19,11
43	141,6	-	86	3,9192	4,06	19,11
44	122,7	-241,8	88	3,7464	3,71	19,56
45	140,5	-	88	3,8024	4,04	19,56
46	146,2	-231,8	88	3,8247	4,24	19,56
47	133,8	-	88	3,7929	4,02	19,56
48	137,7	-	90	3,7021	4	20,00
49	136,7	-	90	3,6776	4,09	20,00
50	131,7	-	92	3,5887	3,98	20,44
51	139	-	92	3,5758	4,09	20,44
52	126,5	-	92	3,5768	3,98	20,44
53	140,6	-	92	3,6174	4,18	20,44
54	126,5	-	94	3,4673	3,95	20,89
55	131,3	-	96	3,3766	4,03	21,33
56	135,2	-	96	3,4311	4,12	21,33
57	124	-254	98	3,2807	3,92	21,78
58	137,3	-	98	3,3301	4,12	21,78
59	140,1	-	98	3,3248	4,22	21,78
60	133,8	-	98	3,3074	4,2	21,78
61	136	-	100	3,223	4,2	22,22
62	140,5	-	102	3,1553	4,18	22,67

Таблица 2.1.2.

Номер	$T_{кип}$, °С	$\Delta_f H^\circ_{(г,298К)}$, кДж/моль	W	J	χ	D_{ADTI}
63	133,5	-	102	3,1513	4,16	22,67
64	141,2	-230,5	102	3,1753	4,35	22,67
65	132,7	-	104	3,073	4,06	23,11
66	136	-	104	3,0608	4,16	23,11
67	143	-	104	3,0922	4,35	23,11
68	135,2	-	108	2,9147	4,13	24,00
69	142,5	-	108	2,9548	4,31	24,00
70	143	-	110	2,8766	4,31	24,44
71	142,5	-	114	2,7467	4,27	25,33
72	150,6	-228,7	120	2,5951	4,41	26,67

2.2. ПОСТРОЕНИЕ QSPR - МОДЕЛЕЙ ДЛЯ РАСЧЕТА ТЕМПЕРАТУРЫ КИПЕНИЯ АЛКАНОВ С ИСПОЛЬЗОВАНИЕМ ТОПОЛОГИЧЕСКИХ ИНДЕКСОВ.

Проведем анализ соотношений между температурой кипения алканов и топологическими индексами и построим линейно-регрессионную модель “структура-свойство”. Разобьем перечисленные в таблице 2.1.1 соединения алканов на тренировочный и контрольный набор и проанализируем зависимость между температурой кипения ($T_{кип}$) и индексом Рандича (χ) (рис. 2.1.).

В результате исследования построена модель позволяющая вычислять температуру кипения с учетом индекса Рандича. Эта модель имеет высокую корреляцию $r = 0,98$ и среднеквадратичное стандартное отклонение $s = 8,96$ °С. График отклонений показывает флуктуацию значений выше и ниже предсказанной температуры кипения. Нет определенной тенденции, отвергающей линейную функциональную зависимость. Анализ отклонений подтверждает уместность

линейной модели для выражения зависимости между температурой кипения алканов и индексом Рандича.

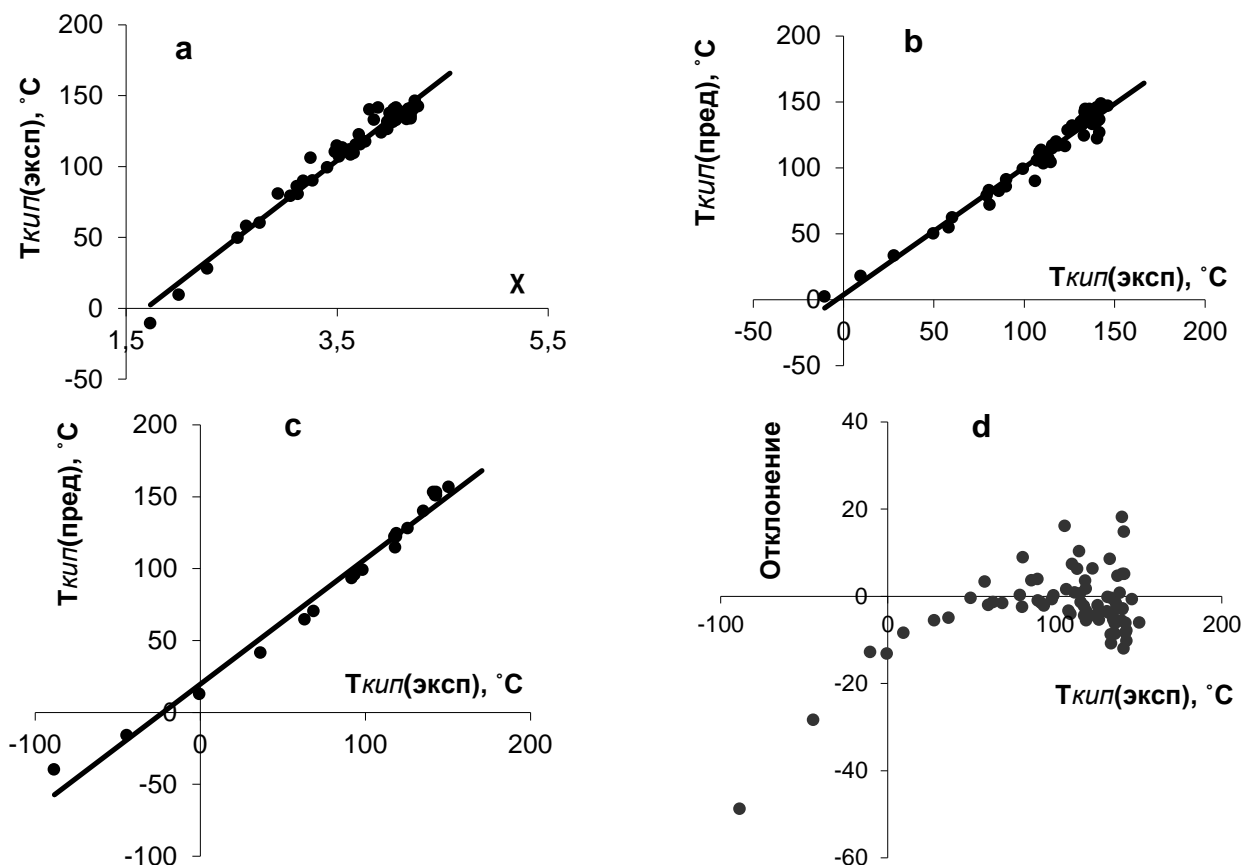


Рисунок. 2.1.

QSPR – зависимость температуры кипения алканов ($T_{кип}$) от величина индекса Рандича (χ): а) соотношение между экспериментальной $T_{кип}$ и индексом Рандича для тренировочного набора; б) соотношение между экспериментальной и предсказанной $T_{кип}$ алканов для тренировочного набора; в) соотношение между экспериментальной и предсказанной $T_{кип}$ алканов для тестового набора; г) анализ отклонений между экспериментальными и предсказанными величинами.

Исследуем взаимосвязь между температурой кипения ($T_{кип}$) алканов и индексом Винера (W) (рис. 2.2.) и построим линейно-регрессионную модель “структура-свойство”. Проследуем аналогичной процедуре по отбору тренировочного и контрольного набора структур.

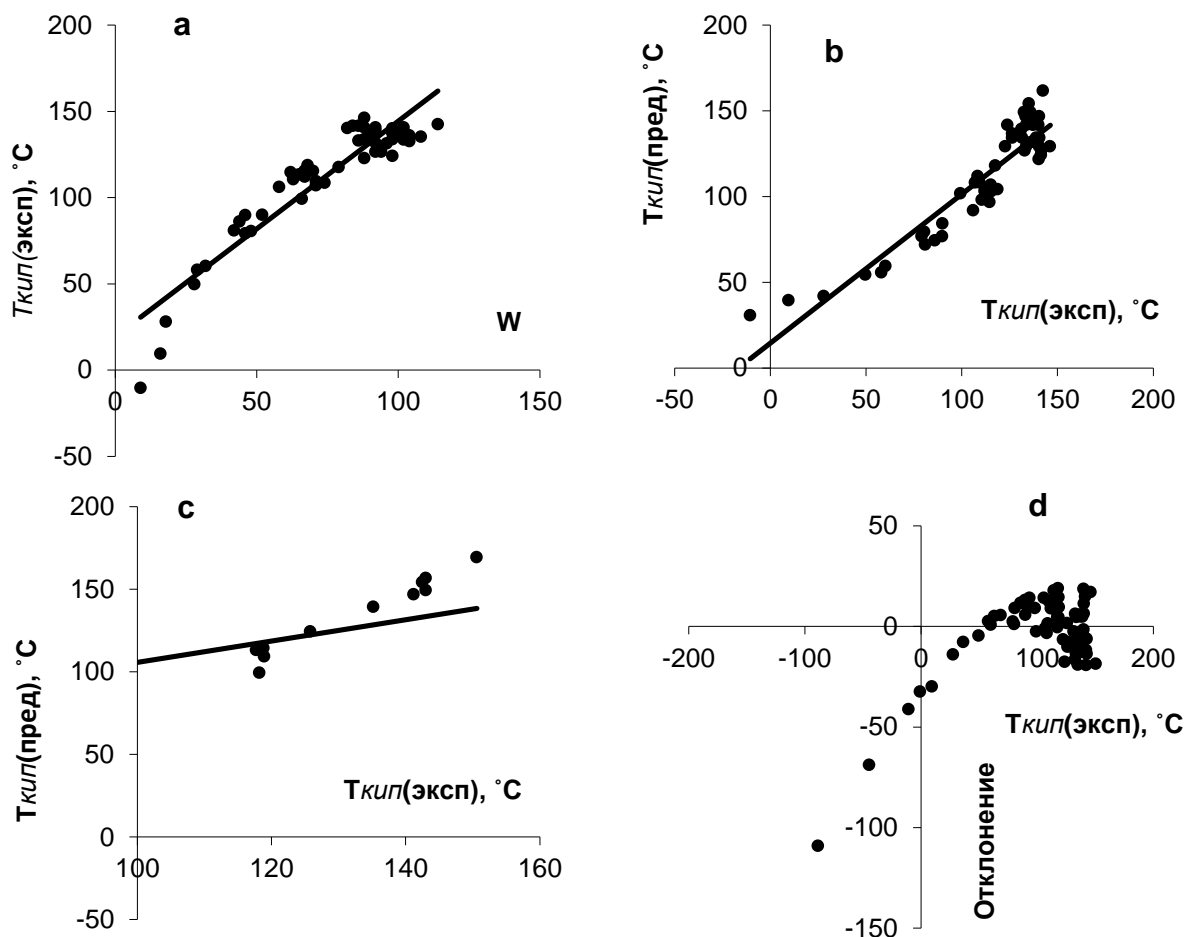


Рисунок. 2.2.

QSPR – зависимость температуры кипения алканов ($T_{кип}$) от величины индекса Винера (W): а) Соотношение между экспериментальной температурой кипения алканов и индексом Винера для тренировочного набора; б) Соотношение между экспериментальной и предсказанной температурой кипения алканов для тренировочного набора; в) Соотношение между экспериментальной и предсказанной температурой кипения алканов для тестового набора; г) Анализ отклонений между экспериментальными и предсказанными значениями.

В результате моделирования построена взаимосвязь, дающая возможность вычислить температуру кипения алканов с учетом индекса Винера. Линейная

регрессия имеют высокую корреляцию $r = 0,93$ и среднеквадратичное стандартное отклонение $s = 19,55$ °С. Не смотря на это, графическая нелинейная зависимость ярко подчеркивает существование функциональной зависимости отличной от линейной. Анализ отклонений подтверждает этот вывод.

В связи с вышеизложенным произведем логарифмическую трансформацию индекса Винера и протестируем соотношение между полученными величинами и температурой кипения алканов (рис. 2.3.).

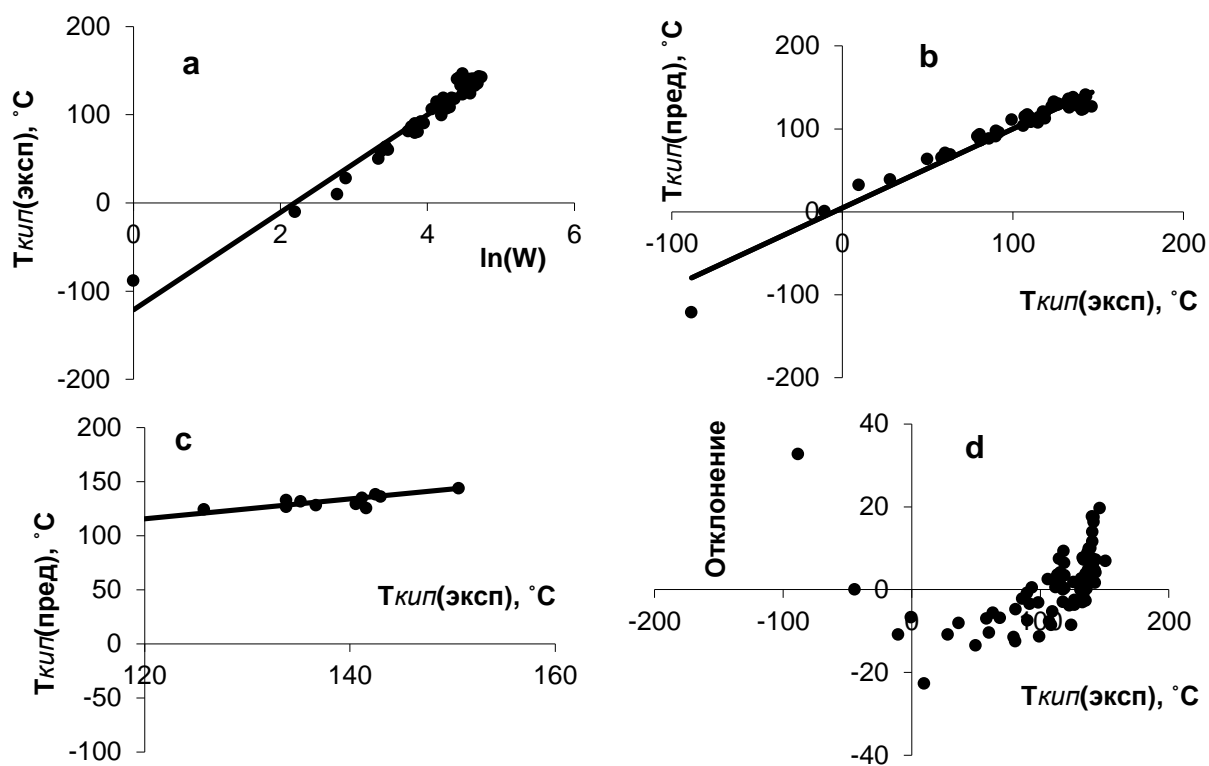


Рисунок 2.3.

QSPR – зависимость, построенная для температуры кипения алканов ($T_{кип}$) с учетом логарифма индекса Винера ($\ln(W)$): а) Соотношение между $T_{кип}$ алканов и логарифма индексом Винера для тренировочного; б) Соотношение между экспериментальной и предсказанной $T_{кип}$ алканов для тренировочного набора; с) Соотношение между экспериментальной и предсказанной $T_{кип}$ алканов для тестового набора; д) Анализ отклонений

между экспериментальными и предсказанными значениями.

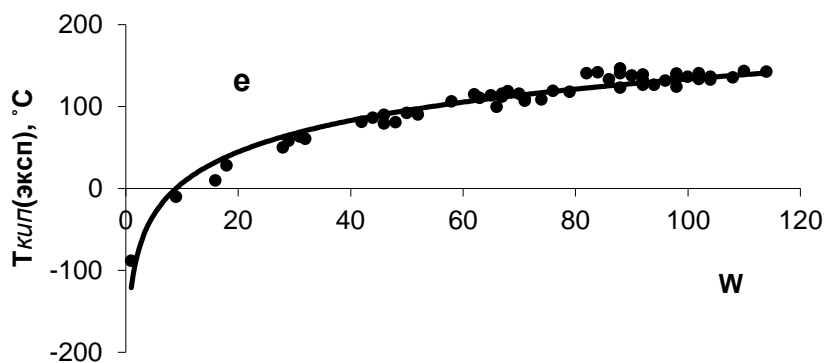


Рисунок 2.3.

QSPR – зависимость, построенная для температуры кипения алканов ($T_{кип}$) с учетом логарифма индекса Винера ($\ln(W)$): е) Температура кипения алканов представлена как логарифмическая функция в которой аргументом выступает индекс Винера. (продолжение)

Новая модель позволяет предсказывать температуру кипения алканов с учетом логарифма индекса Винера. У этой модели повысился коэффициент корреляции $r = 0,98$ и уменьшилось среднеквадратичное стандартное отклонение $s = 8,9$ °C. Анализ отклонений подтверждает, что логарифмическая функциональная зависимость имеет более высокую прогностическую способность.

Исследуем соотношения между температурой кипения алканов и индекса Балабана используя тренировочный и контрольный наборы структур (рис. 2.4.).

Построенная линейная регрессия для расчета температуры кипения алканов с учетом индекса Балабана имеет коэффициент корреляции $r = 0,58$ и стандартное отклонение $s = 32,9$ °C и демонстрирует слабую предсказательную способность. График отклонений указывает, что нет определенной тенденции противоречащей линейной модели. Но природа отклонений свидетельствует о наличии кластеров.

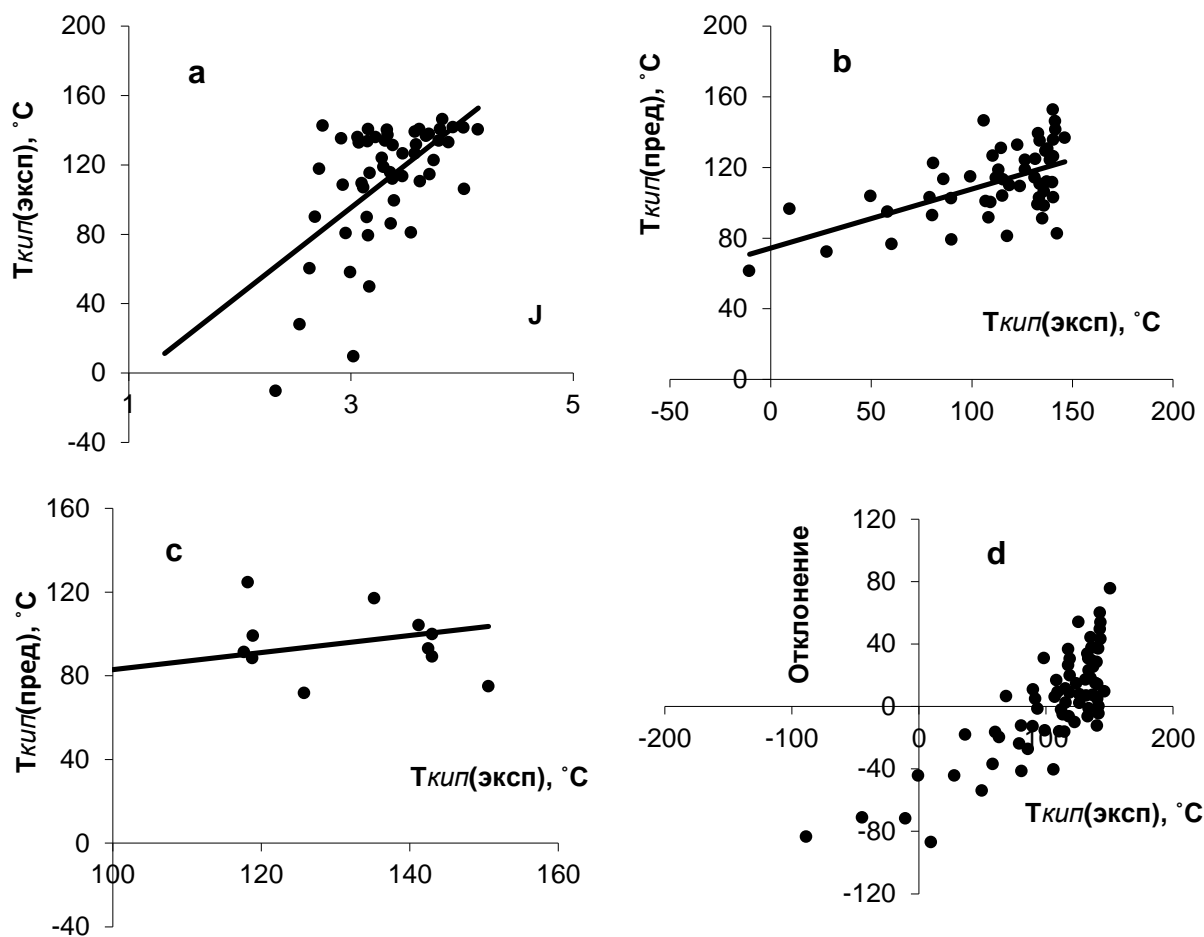


Рисунок 2.4.

QSPR – зависимость, построенная для температуры кипения алканов ($T_{кип}$) с учетом индекса Балабана(J): а) Соотношение между экспериментальной температурой кипения алканов и индексом Балабана для тренировочного набора; б) Соотношение между экспериментальной и предсказанной температурой кипения алканов для тренировочного набора; в) Соотношение между экспериментальной и предсказанной температурой кипения алканов для тестового набора; д) Анализ отклонений между экспериментальными и предсказанными значениями.

2.3. ПОСТРОЕНИЕ QSPR - МОДЕЛЕЙ ДЛЯ РАСЧЕТА ТЕМПЕРАТУРЫ КИПЕНИЯ АЛКАНОВ С ПРИМЕНЕНИЕМ НОВОГО ТОПОЛОГИЧЕСКОГО ИНДЕКСА D_{ADTI} .

Построим двухмерный график для зависимости температуры кипения алканов от структуры, отраженной с учетом топологического индекса среднего расстояния D_{ADTI} , используя обучающую выборку (рис. 2.5.).

Статистические параметры построенной модели $R^2 = 0,870$, $r = 0,93$, $s = 16,53^\circ\text{C}$ указывают на тесную взаимосвязь между величиной топологического индекса D_{ADTI} и температурой кипения, но большой разброс на графике отклонений и характерная прослеживаемая кривая свидетельствуют о необходимости поиска другой функциональной зависимости.

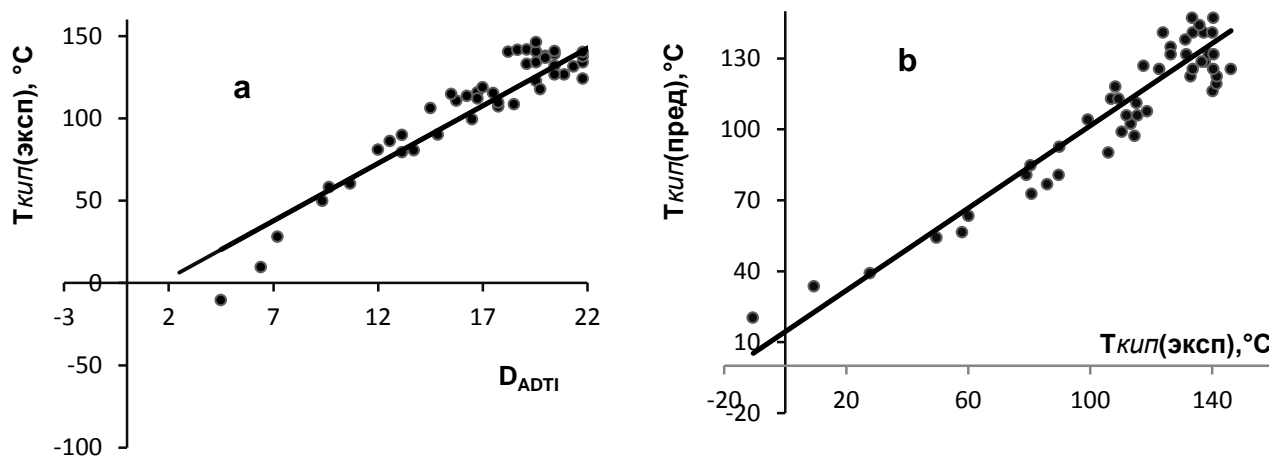


Рисунок 2.5.

QSPR – зависимость, построенная для температуры кипения алканов ($T_{кип}$) с учетом индекса D_{ADTI} : а) Соотношение между экспериментальной $T_{кип}$ и индексом D_{ADTI} для обучающего набора; б) Соотношение между экспериментальной и предсказанной $T_{кип}$ алканов для тренировочного набора.

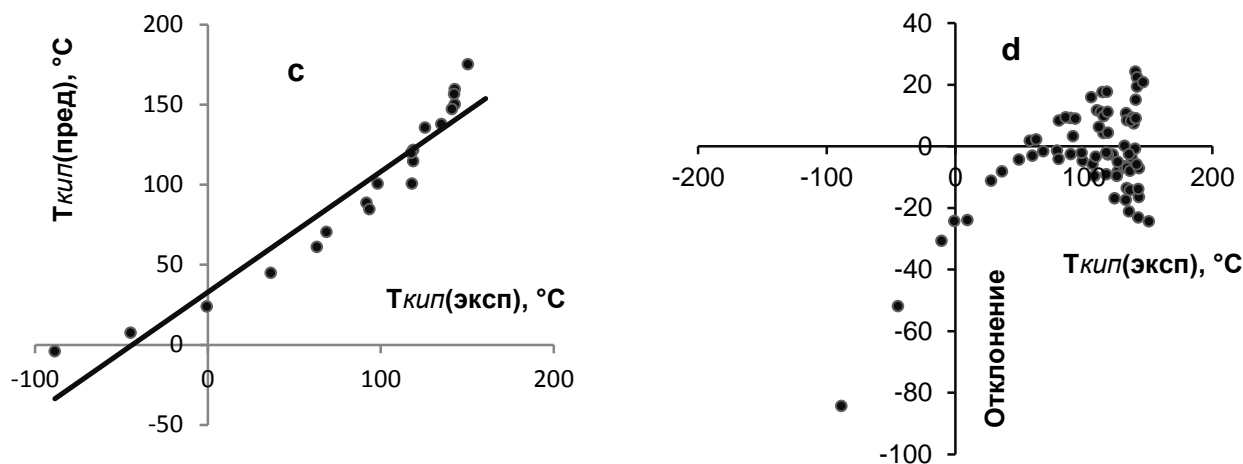


Рисунок 2.5.

QSPR – зависимость, построенная для температуры кипения алканов ($T_{кип}$) с учетом индекса D_{ADTI} : с) Соотношение между экспериментальной и предсказанной $T_{кип}$ алканов для тестового набора d) Анализ отклонений между экспериментальными и предсказанными значениями. (продолжение)

Произведем трансформацию данных с использованием логарифма индекса D_{ADTI} и построим графическую зависимость (рис. 2.6.).

Результаты исследования показали улучшение в статистических показателях модели “структура-свойство” $R^2 = 0,934$, $r = 0,97$, $s = 12,23$ °C. На графике отклонений четко вырисовывается одна точка, выпадающая из общего тренда. Это отклонение между экспериментальной и рассчитанной величиной температуры кипения этана. Молекулярный граф этана имеет очень простой вид, так как молекула состоит только из двух атомов углерода, чем объясняется заметное различие в ответной реакции на процесс моделирования его свойств. Этот факт квалифицирует этан как аутлайер – экстремальную величину, не вписывающуюся в наблюдаемую тенденцию. При исключении этого объекта из исследовательского набора, среднеквадратичное стандартное отклонение снижается до 8,56°C а

коэффициент детерминации $R^2 = 0,986$ и коэффициент корреляции $r = 0,99$ повышаются, указывая на улучшение прогностической способности модели.

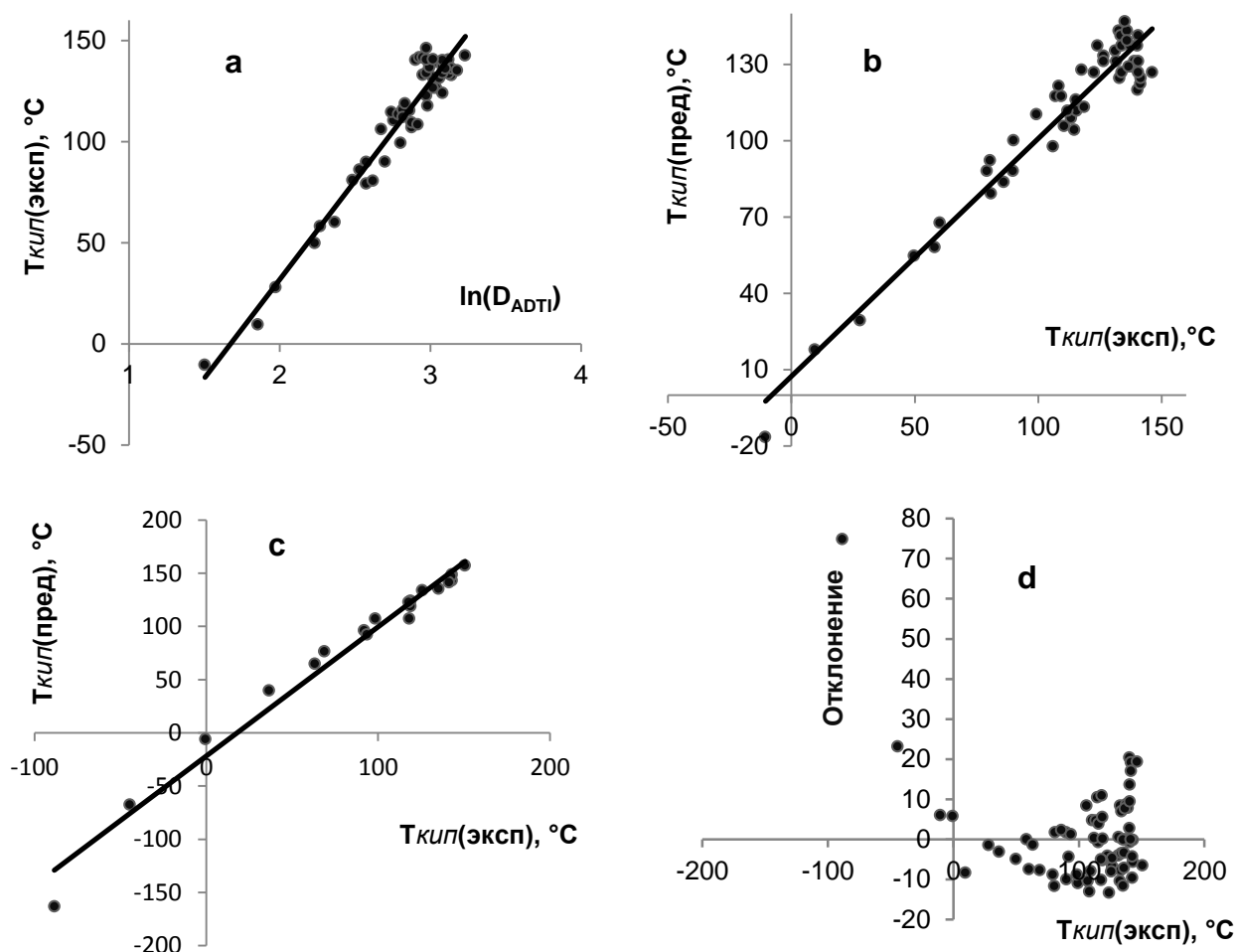


Рисунок 2.6.

QSPR – зависимость, построенная для температуры кипения алканов ($T_{кип}$) с учетом индекса D_{ADTI} в виде логарифмической функции: а) Соотношение между экспериментальной температурой кипения алканов и индексом среднего расстояния D_{ADTI} для тренировочного набора; б) Соотношение между экспериментальной и предсказанной температурой кипения алканов для обучающего набора; в) Соотношение между экспериментальной и предсказанной температурой кипения алканов для контрольного . д) Анализ отклонений между экспериментальными и предсказанными значениями.

Построим и исследуем двухмерный график зависимости между температурой кипения алканов и индексом среднего расстояния D_{ADTI} , выступающим независимой переменной величиной в полиномиальной функции (рис. 2.7.).

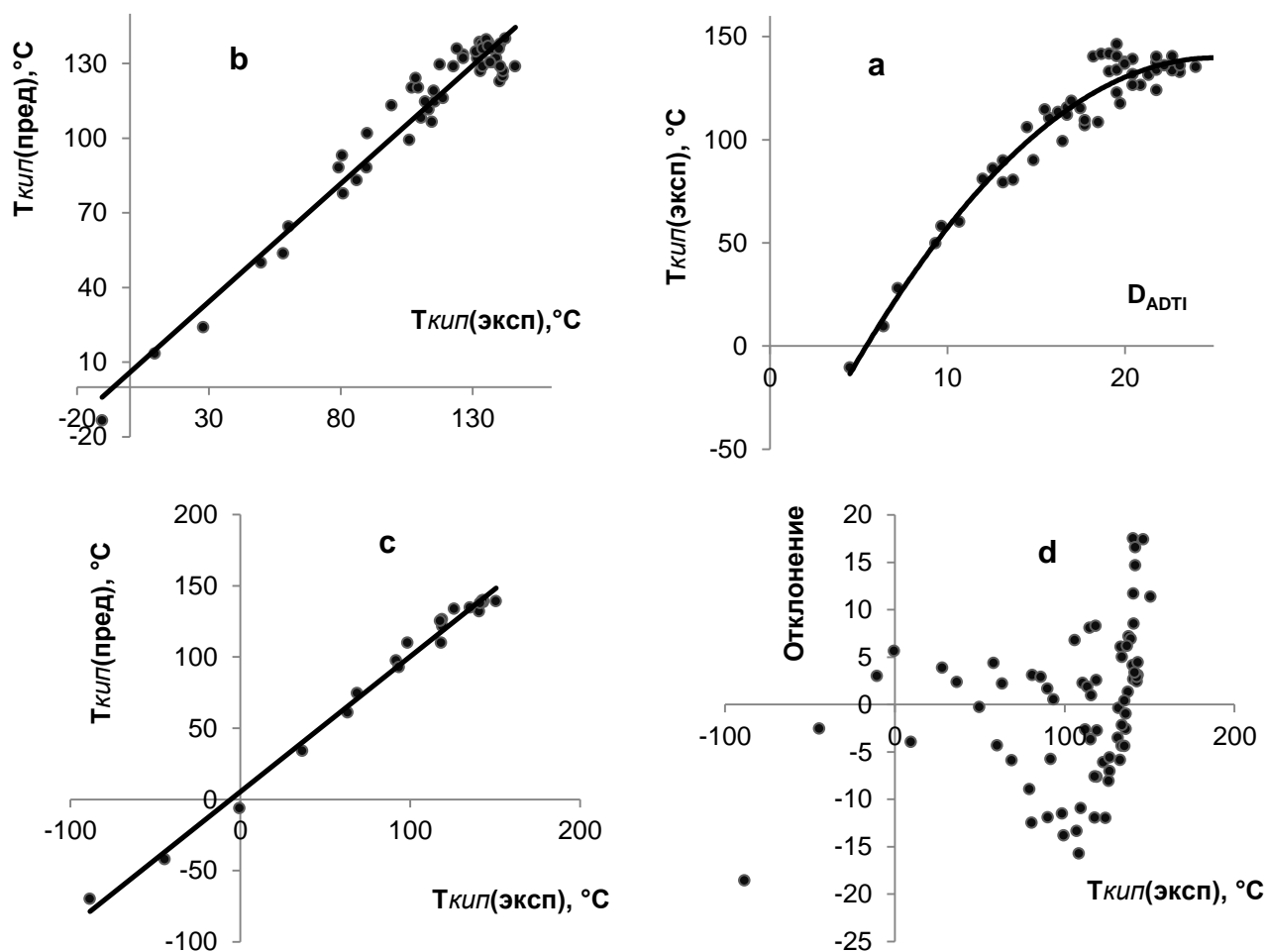


Рисунок 2.7.

QSPR – зависимость, построенная для температуры кипения алканов ($T_{кип}$) с учетом индекса D_{ADTI} в виде полинома функции: а) Соотношение между экспериментальной $T_{кип}$ алканов и индексом среднего расстояния D_{ADTI} для тренировочного; б) Соотношение между экспериментальной и предсказанной $T_{кип}$ алканов для тренировочного набора; в) Соотношение между экспериментальной и предсказанной температурой кипения алканов для тестового набора; д) Анализ отклонений между экспериментальными и предсказанными значениями.

Данная QSPR - модель отражает тесную связь между температурой кипения алканов и индексом среднего расстояния D_{ADTI} . Об этом свидетельствуют высокий коэффициент корреляции $r = 0,97$ и коэффициент детерминации $R^2 = 0,946$. Невысокое среднеквадратичное стандартное отклонение $s = 7,84$ °C и график отклонений подтверждают уместность выбранной функциональной зависимости и указывают на высокую предсказательную возможность построенной QSPR – модели. Сравнивая прогностическую способность созданных моделей «структура-свойство» для определения температуры кипения алканов можно отметить, что молекулярные дескрипторы индекс Рандича, логарифм индекса Винера и индекс среднего расстояния, выступающий аргументом в полиномиальной функции, предоставляют возможность наиболее точно прогнозировать температуру кипения алканов.

2.4. ПОСТРОЕНИЕ QSPR - МОДЕЛЕЙ ДЛЯ РАСЧЕТА ЭНТАЛЬПИИ ОБРАЗОВАНИЯ АЛКАНОВ В ГАЗОВОЙ ФАЗЕ С УЧЕТОМ ТОПОЛОГИЧЕСКИХ ИНДЕКСОВ.

Рассмотрим графическую взаимосвязь между энтальпией образования алканов в газовой фазе ($\Delta_f H^\circ_{(г,298K)}$) и индексом Рандича (χ) для тренировочного и контрольного набора структур (рис. 2.8.) и построим линейно-регрессионную модель “структура-свойство”.

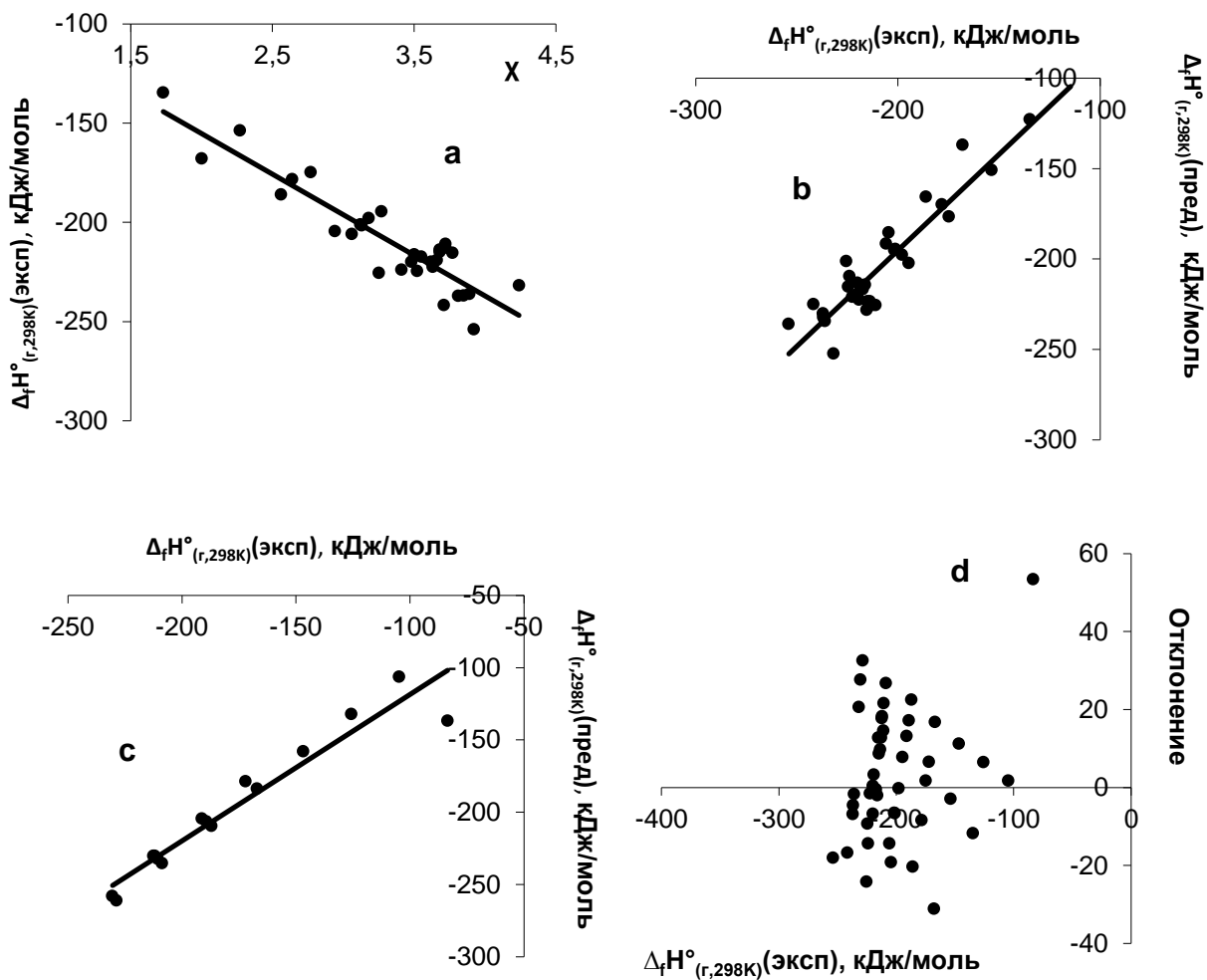


Рисунок 2.8.

Зависимость энтальпии образования алканов в газовой фазе $\Delta_f H^\circ_{(r,298K)}$ от индекса Рандича: а) Зависимость между энтальпией образования алканов для тренировочного набора; б) Соотношение между экспериментальной и предсказанной энтальпией образования алканов для тренировочного набора; в) Соотношение между экспериментальной и предсказанной энтальпией образования алканов для тестового набора; д) Анализ отклонений между экспериментальными и предсказанными значениями.

Данная модель “структура-свойство” для расчета зависимости энтальпии образования алканов и индекса Рандича имеет высокий коэффициент корреляции

$r = 0,92$ и среднеквадратичное стандартное отклонение $s = 16,5$ кДж/моль. Статистические показатели указывает на высокую прогностическую способность. График отклонений демонстрирует уместность линейной модели с незначительными кластерами.

Построим графическую взаимосвязь между энтальпией образования алканов в газовой фазе $\Delta_f H^\circ_{(г,298К)}$ и индексом Винера (W) для тренировочного и тестового набора структур (рис. 2.9.) и определим линейно-регрессионную модель “структура-свойство”.

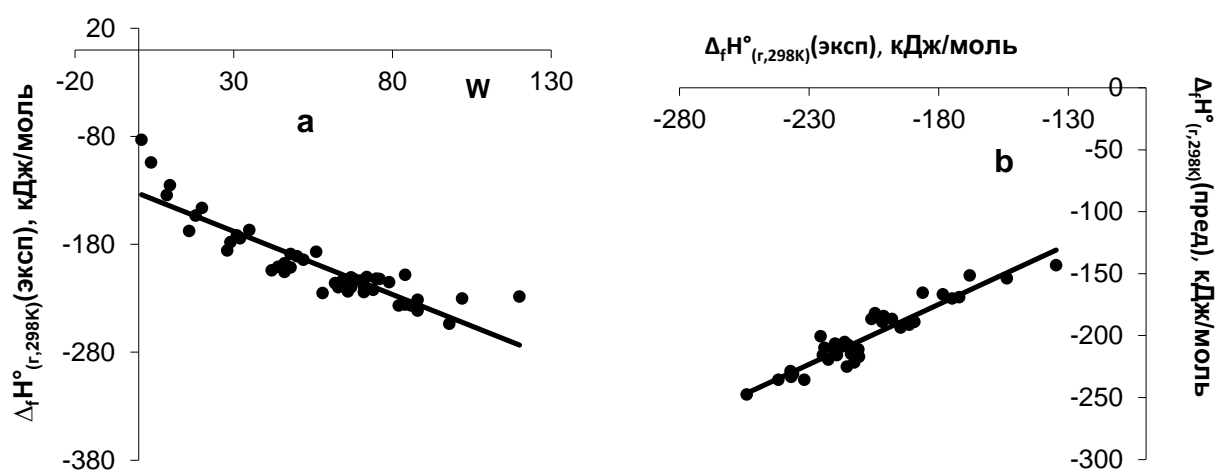


Рисунок 2.9.

Зависимость энтальпии образования алканов в газовой фазе ($\Delta_f H^\circ_{(г,298К)}$) и индекса Винера (W): а) Зависимость между экспериментальной энтальпией образования алканов и индексом Винера для тренировочного набора; б) Соотношение между экспериментальной и предсказанной энтальпией образования алканов для тренировочного набора.

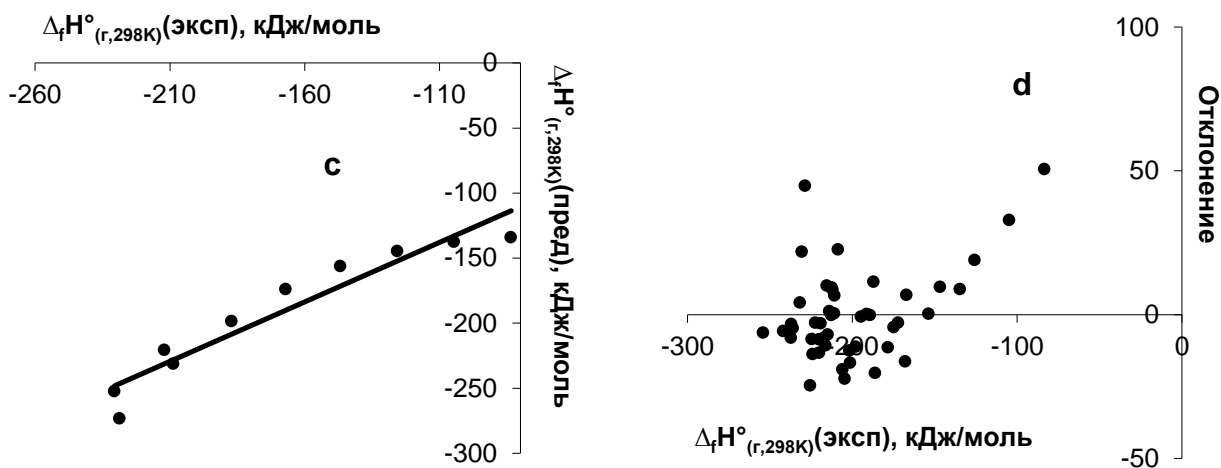


Рисунок 2.9.

Зависимость энтальпии образования алканов в газовой фазе ($\Delta_f H^\circ_{(r,298K)}$) и индекса Винера (W): с) Соотношение между экспериментальной и предсказанной энтальпией образования алканов для тестового набора; d) Анализ отклонений между экспериментальными и предсказанными значениями. (продолжение)

Полученная модель «структура-свойство» имеет высокий коэффициент корреляции $r = 0,9$ и умеренное среднеквадратичное стандартное отклонение $s = 15,8$ кДж/моль. Анализ отклонений указывает на несоответствие линейной зависимости и необходимость модификации молекулярного дескриптора путем трансформации значений.

Протестируем нелинейную функциональную зависимость для построения модели, отражающей величину энтальпии образования в газовой фазе в ряду алканов ($\Delta_f H^\circ_{(r,298K)}$). Построим графическую зависимость между величиной энтальпии образования алканов и логарифмом индекса Винера (рис. 2.10.).

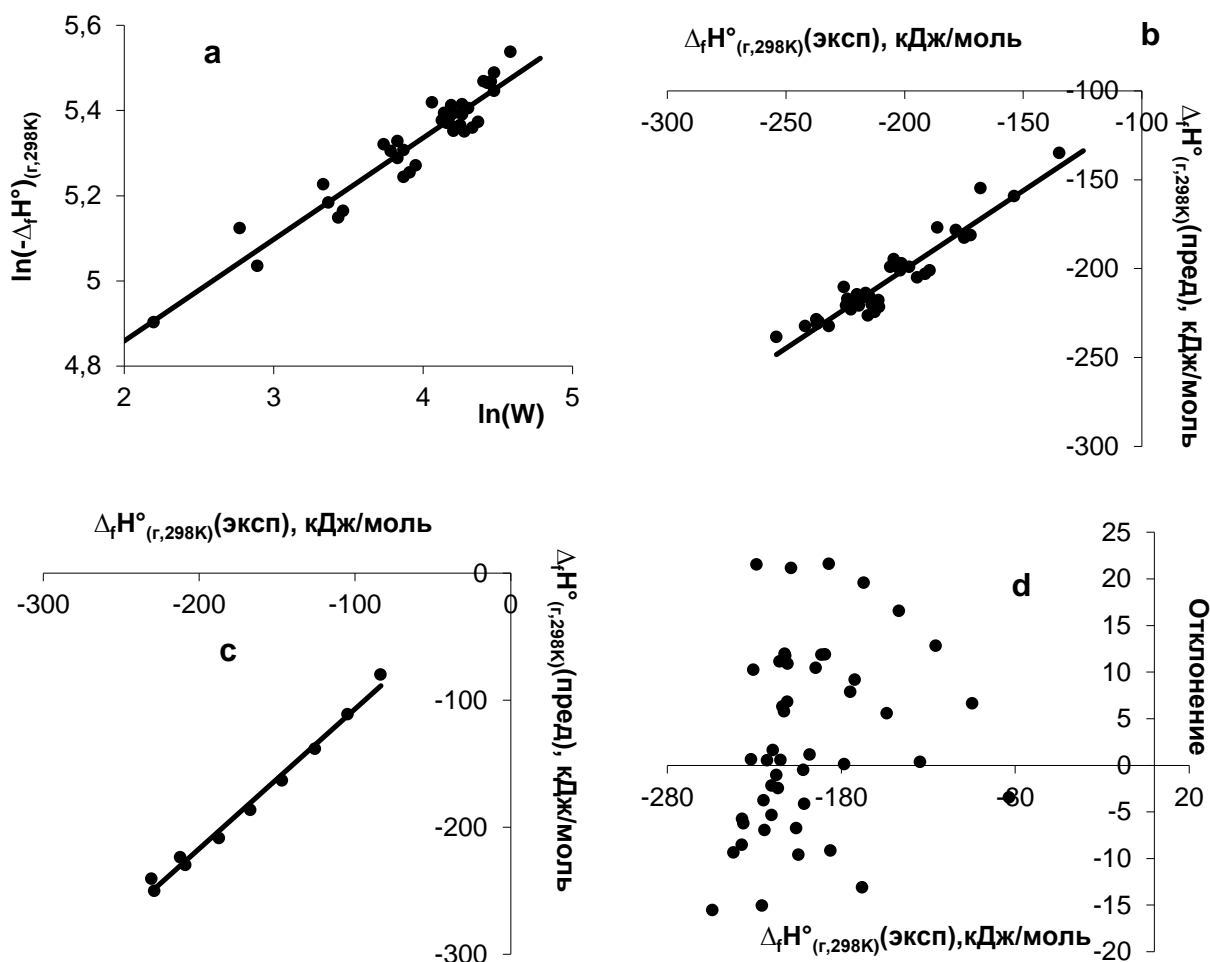


Рисунок 2.10.

QSPR – зависимость, построенная для логарифма величины энтальпии образования алканов в газовой фазе ($\ln(-\Delta_f H^\circ_{(r,298K)})$) с учетом логарифма индекса Винера ($\ln(W)$): а) Соотношение между логарифмом величины экспериментальной энтальпии образования алканов и логарифмом индекса Винера для тренировочного набора; б) Соотношение между экспериментальной и предсказанной энтальпией образования алканов для тренировочного набора; в) Соотношение между экспериментальной и предсказанной энтальпией образования алканов для тестового набора; д) Анализ отклонений между экспериментальными и предсказанными значениями.

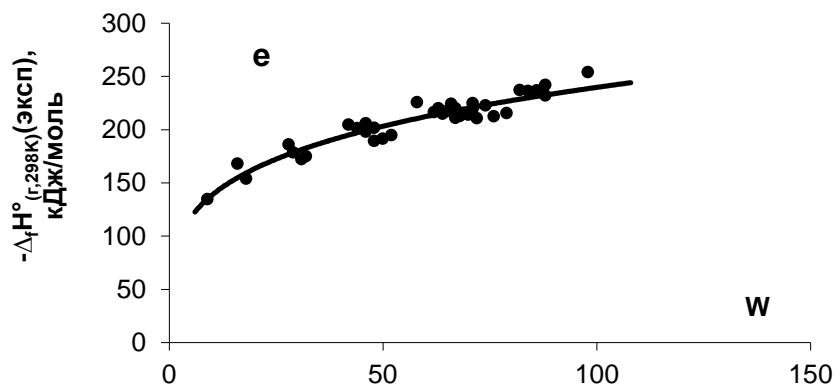


Рисунок 2.10.

QSPR – зависимость, построенная для логарифма величины энтальпии образования алканов в газовой фазе ($\ln(-\Delta_f H^\circ_{(г,298К)})$) с учетом логарифма индекса Винера ($\ln(W)$): е) Соотношение между энтальпией образования алканов и индексом Винера в виде показательной функции. (продолжение)

Построенная модель «структура-свойство» выражает зависимость между энтальпией образования алканов и индексом Винера как степенная функция. Коэффициент корреляции повысился $r = 0,98$, а среднеквадратичное стандартное отклонение уменьшилось $s = 9,8$ кДж/моль. Статистические характеристики демонстрируют высокую предсказательную способность зависимости.

Построим зависимость между энтальпией образования алканов в газовой фазе ($\Delta_f H^\circ_{(г,298К)}$) и индексом Балабана (J) для тренировочного и контрольного наборов соединений (рис. 2.11.).

Полученная модель “структура-свойство” представляет соотношение между энтальпией образования алканов в газовой фазе и индексом Балабана. Она имеет коэффициент корреляции $r = 0,74$, среднеквадратичное стандартное отклонение $s = 19,7$ кДж/моль, что является показателем умеренной взаимной зависимости и посредственной предсказательной способности.

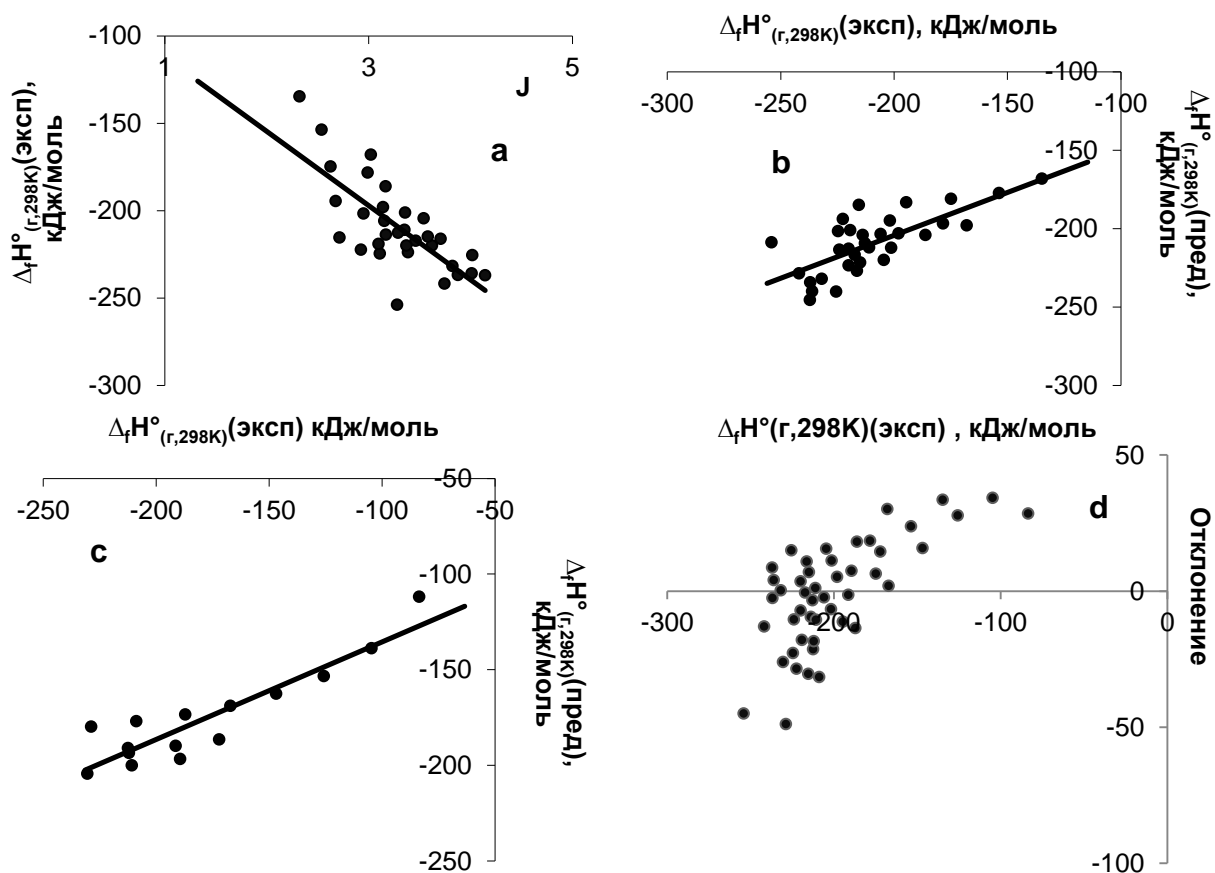


Рисунок 2.11.

QSPR – зависимость, построенная для энтальпии образования алканов в газовой фазе ($\Delta_f H^\circ_{(г,298К)}$) с учетом индекса Балабана (J): а) Соотношение между величиной экспериментальной энтальпии образования алканов и индексом Балабана для тренировочного набора; б) Соотношение между экспериментальной и предсказанной энтальпией образования алканов для тренировочного набора; в) Соотношение между экспериментальной и предсказанной энтальпией образования алканов для тестового набора; д) Анализ отклонений между экспериментальными и предсказанными значениями.

2.5. ПОСТРОЕНИЕ QSPR - МОДЕЛЕЙ ДЛЯ РАСЧЕТА ЭНТАЛЬПИИ ОБРАЗОВАНИЯ АЛКАНОВ В ГАЗОВОЙ ФАЗЕ С УЧЕТОМ ИНДЕКСА СРЕДНЕГО РАССТОЯНИЯ.

Произведем анализ графической зависимости между энтальпией образования алканов в газовой фазе с учетом индекса среднего расстояния (рис. 2.12.).

Коэффициент корреляции $r = 0,94$ и среднеквадратичное стандартное отклонение $s = 14,86$ кДж/моль свидетельствуют, что построенная модель «структура-свойство» обладает хорошей прогностической способностью.

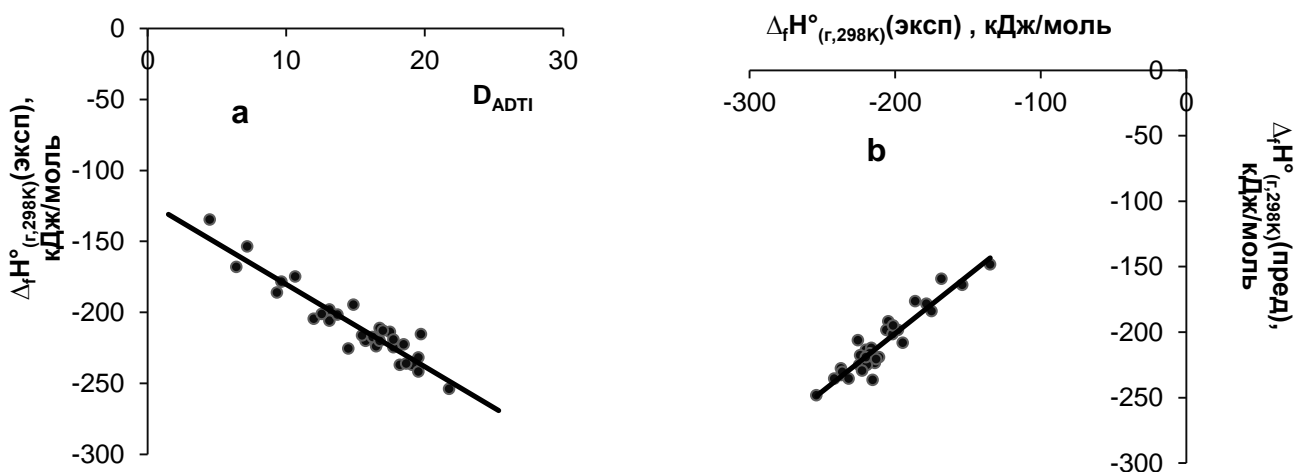


Рисунок 2.12.

QSPR – зависимость, построенная для энтальпии образования алканов в газовой фазе ($\Delta_f H^\circ_{(г,298К)}$) с учетом индекса среднего расстояния (D_{ADTI}): а) Соотношение между величиной экспериментальной энтальпии образования алканов и индексом среднего расстояния для тренировочного набора; б) Соотношение между экспериментальной и предсказанной энтальпией образования алканов для тренировочного набора.

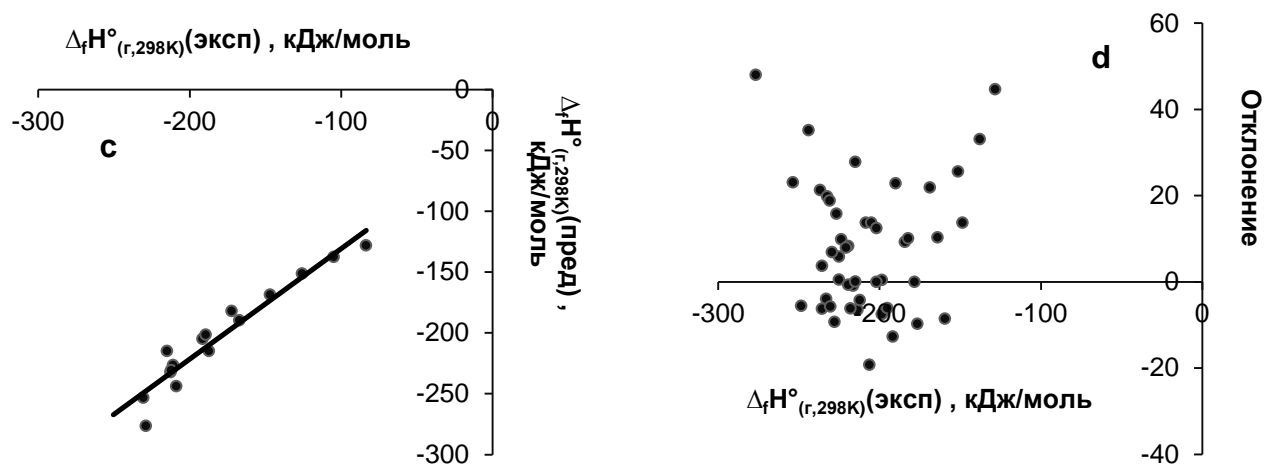


Рисунок 2.12.

QSPR – зависимость, построенная для энтальпии образования алканов в газовой фазе ($\Delta_f H^\circ_{(г,298К)}$) с учетом индекса среднего расстояния (D_{ADTI}): c) Соотношение между экспериментальной и предсказанной энтальпией образования алканов для тестового набора; d) Анализ отклонений между экспериментальными и предсказанными значениями. (продолжение)

Проверим наличие взаимосвязи между энтальпией образования алканов в газовой фазе и индексом среднего расстояния в виде логарифмической функции (рис. 2.13.).

Полученная QSPR модель «структура-свойство» описывает тесное взаимодействие между структурой и энтальпией образования алканов в газовой фазе. Следует отметить, что значения переменных для этана являются экстремальными и выделяются из общей тенденции. Это можно предположительно объяснить короткой цепью атомов в молекуле. Исключение этана из группы исследуемых объектов ведет к повышению коэффициента корреляции с $r = 0,96$ до $r = 0,97$ и уменьшению среднеквадратичного стандартного отклонения с $s = 14,70$ кДж/моль до $s = 11,97$ кДж/моль.

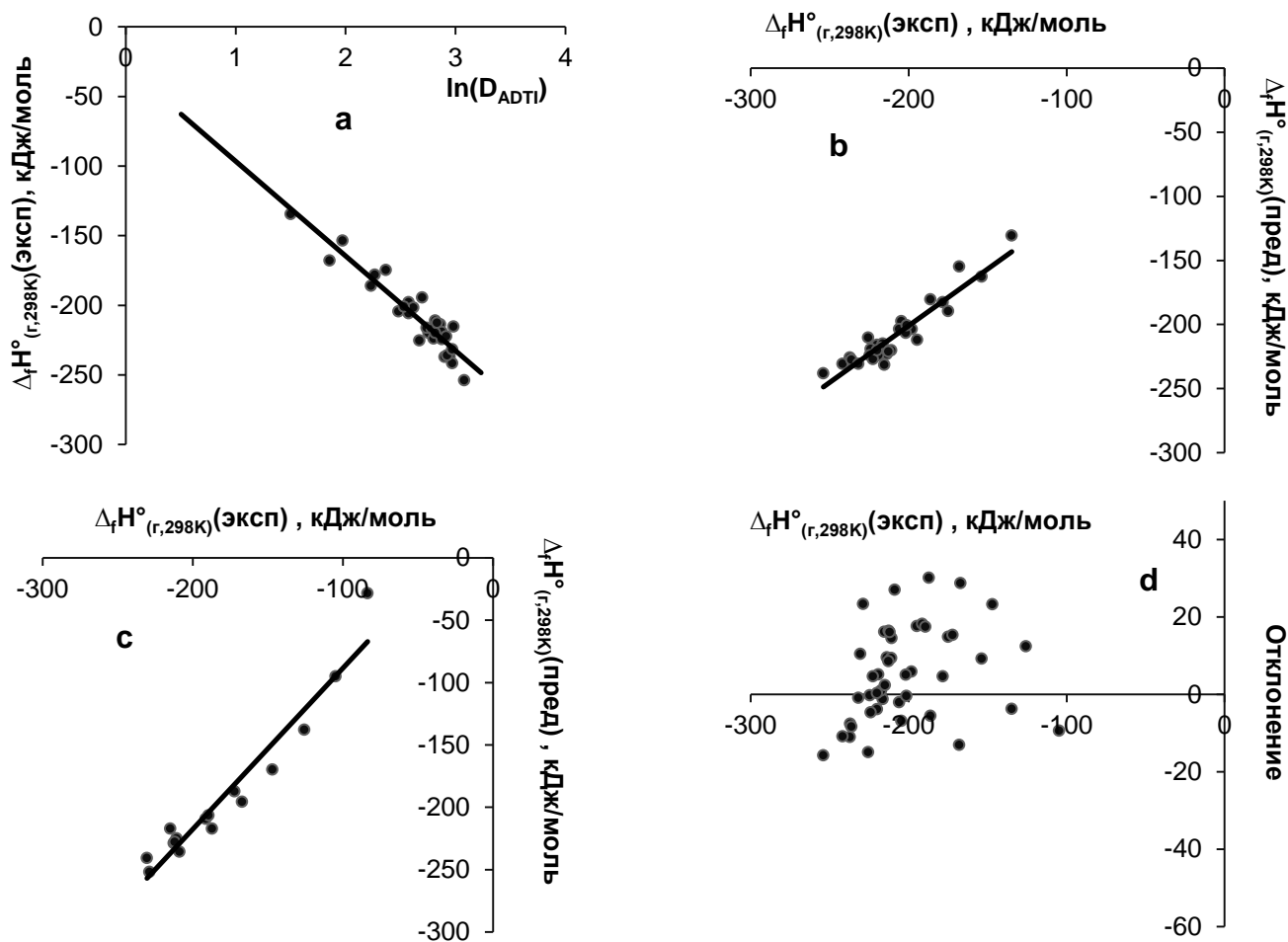


Рисунок 2.13.

QSPR – зависимость, построенная для энтальпии образования алканов в газовой фазе ($\Delta_f H^\circ_{(r,298K)}$) с учетом логарифма индекса среднего расстояния ($\ln(D_{\text{ADTI}})$): а)

Соотношение между величиной экспериментальной энтальпии образования алканов и индексом среднего расстояния для тренировочного набора; б)

Соотношение между экспериментальной и предсказанной энтальпией образования алканов для обучающего набора; с)

Соотношение между экспериментальной и предсказанной энтальпией образования алканов для тестового набора; д)

Анализ отклонений между экспериментальными и предсказанными значениями.

Таблица 2.5.1. QSPR - Уравнения полученные для алканов, статистические характеристики: R^2 - коэффициент детерминации, r - коэффициент корреляции, s - стандартное отклонение.

№	QSPR-модель	R^2	r	s
1	$T_{кип} = 57,6\chi - 97,303$	0,9658	0,98	8,96 °C
2	$T_{кип} = 1,2484W + 19,428$	0,8698	0,93	21,4 °C
3	$T_{кип} = 55,337\ln(W) - 121,17$	0,9531	0,98	8,8 °C
4	$T_{кип} = 50,188J - 55,264$	0,3338	0,58	32,88 °C
5	$T_{кип} = 6,981D_{ADTI} - 11,11$	0,870	0,93	16,53°C
6	$T_{кип} = 97,59\ln(D_{ADTI}) - 163,3$	0,934	0,97	12,23 °C
7	$T_{кип} = -0,360(D_{ADTI})^2 + 18,11D_{ADTI} - 87,69$	0,946	0,97	7,84 °C
8	$\Delta_f H^\circ_{(г,298K)} = -40,882\chi - 73,484$	0,8415	0,92	16,5 кДж/моль
9	$\Delta_f H^\circ_{(г,298K)} = -1,1713W - 132,88$	0,8053	0,9	15,8 кДж/моль
10	$\Delta_f H^\circ_{(г,298K)} = -80W^{0,2382}$	0,9106	0,95	9,8 кДж/моль
11	$\Delta_f H^\circ_{(г,298K)} = -42,52J - 69,439$	0,5473	0,74	19,7 кДж/моль
12	$\Delta_f H^\circ_{(г,298K)} = -5,788D_{ADTI} - 122,3$	0,882	0,94	14,86 кДж/моль
13	$\Delta_f H^\circ_{(г,298K)}(эксп) = -68,08\ln D_{ADTI} - 28,43$	0,884	0,94	14,70 кДж/моль

2.6. ПРОВЕРКА СТАТИСТИЧЕСКОЙ ГИПОТЕЗЫ ДИСКРИМИНИРУЮЩЕЙ СПОСОБНОСТИ ИНДЕКСА СРЕДНЕГО РАССТОЯНИЯ.

Для оценки дискриминирующей способности топологического индекса среднего расстояния при расчете температуры кипения алканов выдвинем нулевую гипотезу при уровне значимости $\alpha = 0,05$: Стандартное отклонение температуры кипения алканов экспериментальной группы не имеет существенного различия с контрольной группой. Альтернативная гипотеза предполагает

существенное различие между экспериментальными и контрольными стандартными отклонениями.

$$H_0: \mu_{\text{эксп.}} = \mu_{\text{контр.}} \quad H_A: \mu_{\text{эксп.}} \neq \mu_{\text{контр.}}$$

При проведении t теста двух выборок получаем величину p -значимости 0,4076, которая не позволяет отклонить нулевую гипотезу. Если альтернативная гипотеза верна и отличие значений стандартного отклонения в экспериментальной и контрольной группах статистически существенно, то вероятность случайности комбинации значений стандартного отклонения в экспериментальной и контрольной группах в данных выборках слишком велика. Следовательно, отличие стандартного отклонения в экспериментальной и контрольной группах несущественно при уровне значимости $\alpha = 0,05$.

Для оценки дискриминирующей способности топологического индекса среднего расстояния при расчете энтальпии образования алканов выдвинем нулевую гипотезу при уровне значимости $\alpha = 0,05$: Стандартное отклонение энтальпии образования алканов экспериментальной группы не имеет существенного различия с контрольной группой. Альтернативная гипотеза утверждает существенное различие между экспериментальными и контрольными стандартными отклонениями.

$$H_0: \mu_{\text{эксп.}} = \mu_{\text{контр.}} \quad H_A: \mu_{\text{эксп.}} \neq \mu_{\text{контр.}}$$

При проведении t теста двух выборок получаем величину p -значимости 0,7671, которая не позволяет отвергнуть нулевую гипотезу. Если альтернативная гипотеза подлинна и расхождение значений стандартного отклонения в экспериментальной и контрольной группах статистически существенно, то вероятность стандартного отклонения случайно достигать таких экстремальных и более экстремальных значений как в экспериментальной выборке очень высока. Таким образом, отличие стандартного отклонения в экспериментальной и контрольной группах несущественно при уровне значимости $\alpha = 0,05$.

ГЛАВА 3. ПОСТРОЕНИЕ QSPR - МОДЕЛЕЙ ДЛЯ ЦИКЛОАЛКАНОВ.

3.1. СТРУКТУРА И СВОЙСТВА ЦИКЛОАЛКАНОВ.

Циклоалканы, подобно алканам, также содержат только углерод - водородные связи и углерод - углеродные простые связи, но атомы углерода соединены в кольцо. Самым маленьким циклоалканом является циклопропан. Общая формула для циклоалканов является C_nH_{2n} . В нефтехимической промышленности они являются источником получения ароматических углеводородов путем каталитического риформинга. Наибольшее практическое значение приобрёл циклогексан, применяемый для синтеза капролактама, адипиновой кислоты и других соединений, используемых в производстве синтетического волокна. Названия циклоалканов строятся из названий соответствующих алканов с добавлением приставки «цикло», например циклопропан, 1,3-диметилциклогексан. Для циклоалканов характерны следующие виды изомерии:

- Изомерия углеродного скелета
- Пространственная (цис-транс-изомерия)
- Межклассовая изомерия с алкенами

Все атомы углерода в молекулах циклоалканов имеют sp^3 -гибридизацию. Не следует думать, что это все плоские молекулы. Кольца в циклоалканах (за исключением циклопропана) - неплоские. Так, циклобутан имеет слегка вспученную форму - один из углеродных атомов располагается выше или ниже плоскости, в которой находятся три остальных атома, циклопентан - конформацию конверта или твист-конформацию, циклогексан может существовать в двух конформациях кресла, при переходе между которыми (через конформацию ванны) все аксиальные заместители становятся экваториальными и наоборот. Для циклов больших размеров число конформаций возрастает, поэтому такие соединения

существуют в виде нескольких взаимопревращающихся конформеров. Так, для циклогептана возможны 4 устойчивые конформации: искаженное кресло, кресло, ванна, искаженная ванна, для циклооктана - 11 конформаций.

Циклоалканы очень похожи на алканы в реакционной способности, за исключением самых маленьких представителей - особенно циклопропана. Циклопропан гораздо более активный, чем можно было бы ожидать. Причина связана с величиной валентных углов в кольце. Обычно, когда углерод образует четыре одинарные связи, величина валентных углов около $109,5^\circ$. В циклопропане они 60° . Электронные пары расположены очень близко друг к другу, что вызывает много отталкивания между связями, соединяющих атомы углерода. Это позволяет легко разрушить соединение. Циклоалканы менее устойчивы, чем их ациклические аналоги. Мерой относительной устойчивости циклоалканов служит энергия напряжения. Устойчивость циклоалканов увеличивается от циклопропана к циклогексану (последний - полностью ненапряженное соединение), затем падает к циклононану и снова возрастает от циклодекана и далее в область больших циклов.

По химическим свойствам циклоалканы, начиная от C_5 , подобны предельным алифатическим углеводородам; циклопропан по склонности к электрофильному присоединению напоминает непредельные углеводороды, но пассивнее их. Циклоалканы вступают в реакции с изменением величины цикла, раскрытием цикла и трансаннулярной циклизацией (между атомами средних циклов C_8-C_{12}). Получают циклоалканы циклизацией дигалогенидов (в основном 3- и 4 - членные циклы); гидрированием циклоалкенов или ароматических соединений и используя функционально замещенные циклоалканы.

Таблица 3.1.1. Данные для циклоалканов: χ – индекс Рандича, RW/SZ – модифицированный индекс Винера для нечетных колец и индекса Гутмана для четных колец, D_{ADTI} – индекс среднего расстояния, $T_{кип}$ – температура кипения, $\Delta_f H^\circ_{(г,298К)}$ – энтальпия образования в газовой фазе.

Название соединения	X	RW/SZ	D_{ADTI}	$T_{кип}, ^\circ C$	$\Delta_f H^\circ_{(г,298К)},$ кДж/моль
1, 1 - диметилциклопентан	3,707	69,25	11,14		-138,2
1, 1 - диметилциклопропан	2,207	24,75	6,00		-8,2
цис-1,1-Диметилциклогексан	3,71	104	14,75	119,6	
Метилциклобутан	2,39	28	6,40	36,8	-44,5
Метилциклогексан	3,393	78	12,00	100,93	-154,7
Метилциклопентан	2,894	49	8,67	71,8	-105,73
Метилциклопропан	1,89	14,5	4,00	0,7	
н – бутилциклогексан	4,932	196	26,60	180,9	-213
н – гептилциклогексан	6,432	404	48,31		-288,1
н – пропилциклогексан	4,432	148	20,89	156,75	-193,2
н-пропилциклопентан	3,932	105	31,00		-147,9
транс-1, 2 - диметилциклогексан	3,305	106	15,00	123,5	-180,9
Циклобутан	2	16	4,00	12,6	28,37
Циклогексан	3	54	9,00	80,73	-123,4
Циклогептан	3,5	85,75	12,00	118,5	-118,03
Циклодекан	5	250	25,00	201	-154,3
Циклононан	4,5	182,25	20,00	179,44	-132,8
Циклооктан	4	128	16,00	149,0	-124,4
Циклопентан	2,5	31,25	6,00	49,3	-77,2
Циклопропан	1,5	6,75	2,00	-32,81	53,26

Таблица 3.1.1.

Название соединения	X	RW/SZ	DADTI	T _{кип} , °C	$\Delta_f H^\circ(\Gamma, 298K)$, кДж/моль
цис-1,3-Диметилциклопентан	3,29	66,75	11,71	90,8	
Этилциклобутан	2,932	45	9,67	70,8	-26,3
этилциклогексан	3,932	109	14,22	131,9	-171,7
этилциклопентан	3,432	73,25	12,29	103,5	-126,99

3.2. ПОСТРОЕНИЕ QSPR - МОДЕЛЕЙ ДЛЯ РАСЧЕТА ТЕМПЕРАТУРЫ КИПЕНИЯ ЦИКЛОАЛКАНОВ С УЧЕТОМ ТОПОЛОГИЧЕСКИХ ИНДЕКСОВ.

Исследуем зависимость между температурой кипения циклоалканов ($T_{кип}$) и индексом Рандича (χ) (рис. 3.1.) и построим линейно-регрессионную модель «структура-свойство».

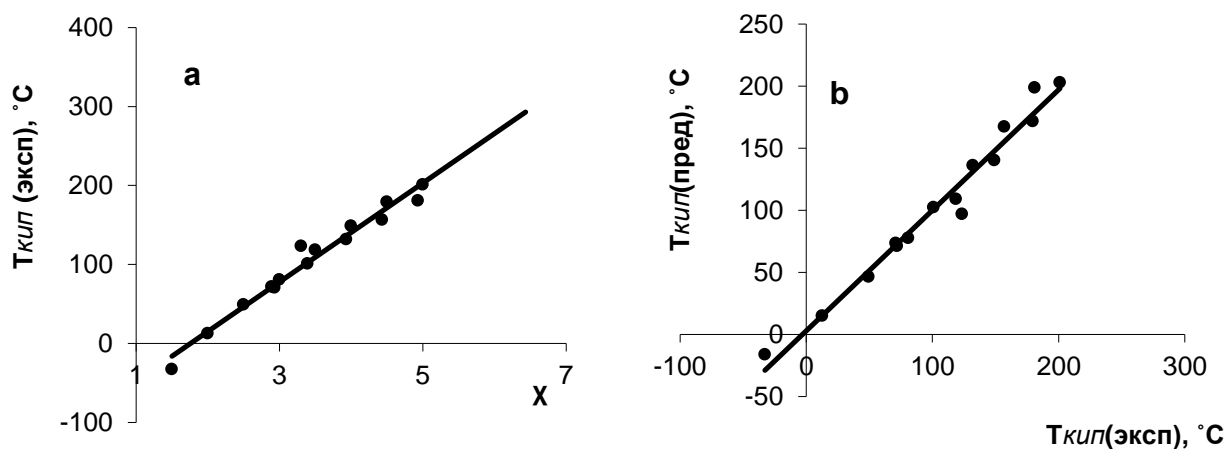


Рисунок 3.1. QSPR – зависимость, построенная для температуры кипения циклоалканов ($T_{кип}$) с учетом индекса Рандича (χ): а) Соотношение между экспериментальной температурой кипения циклоалканов ($T_{кип}$) и индексом Рандича (χ) для тренировочного набора; б) Соотношение между экспериментальной и предсказанной температурой кипения циклоалканов для тренировочного набора.

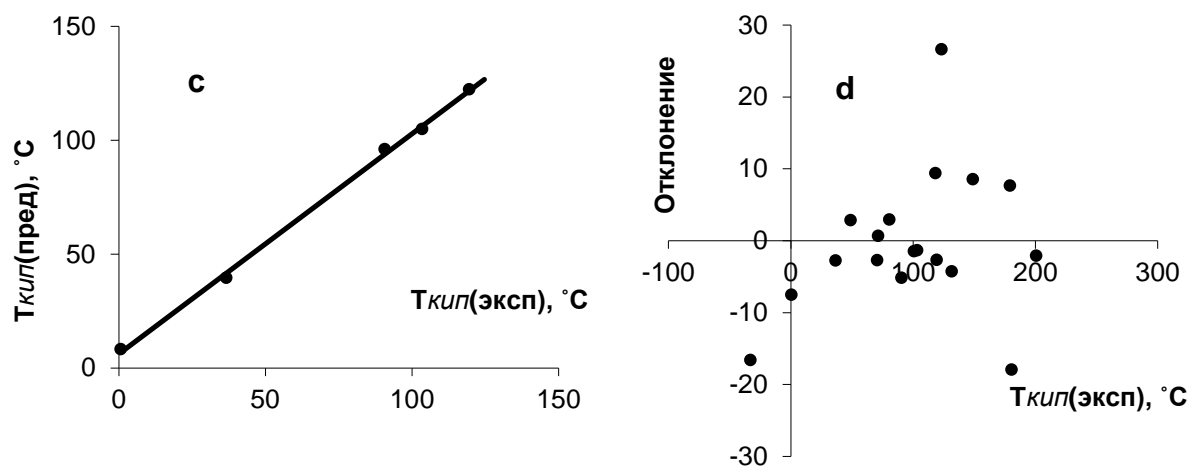


Рисунок 3.1.

QSPR – зависимость, построенная для температуры кипения циклоалканов ($T_{кип}$) с учетом индекса Рандича (χ):. c) Соотношение между экспериментальной и предсказанной температурой кипения циклоалканов для тестового набора; d) Анализ отклонений между экспериментальными и предсказанными значениями. (продолжение)

Анализ продемонстрировал сильную зависимость между температурой кипения циклоалканов и индексом Рандича. Это подтверждает коэффициент корреляции $r = 0,99$, среднеквадратичное стандартное отклонение $s = 9,4^\circ\text{C}$. График отклонений не имеет очевидной тенденции противоречащей линейной зависимости.

Протестируем зависимость между температурой кипения циклоалканов ($T_{кип}$) и модифицированным индексом Винера для нечетных колец и индексом Гутмана для четных колец (RW/Sz) (рис. 3.2.) и построим линейно-регрессионную модель “структура-свойство”.

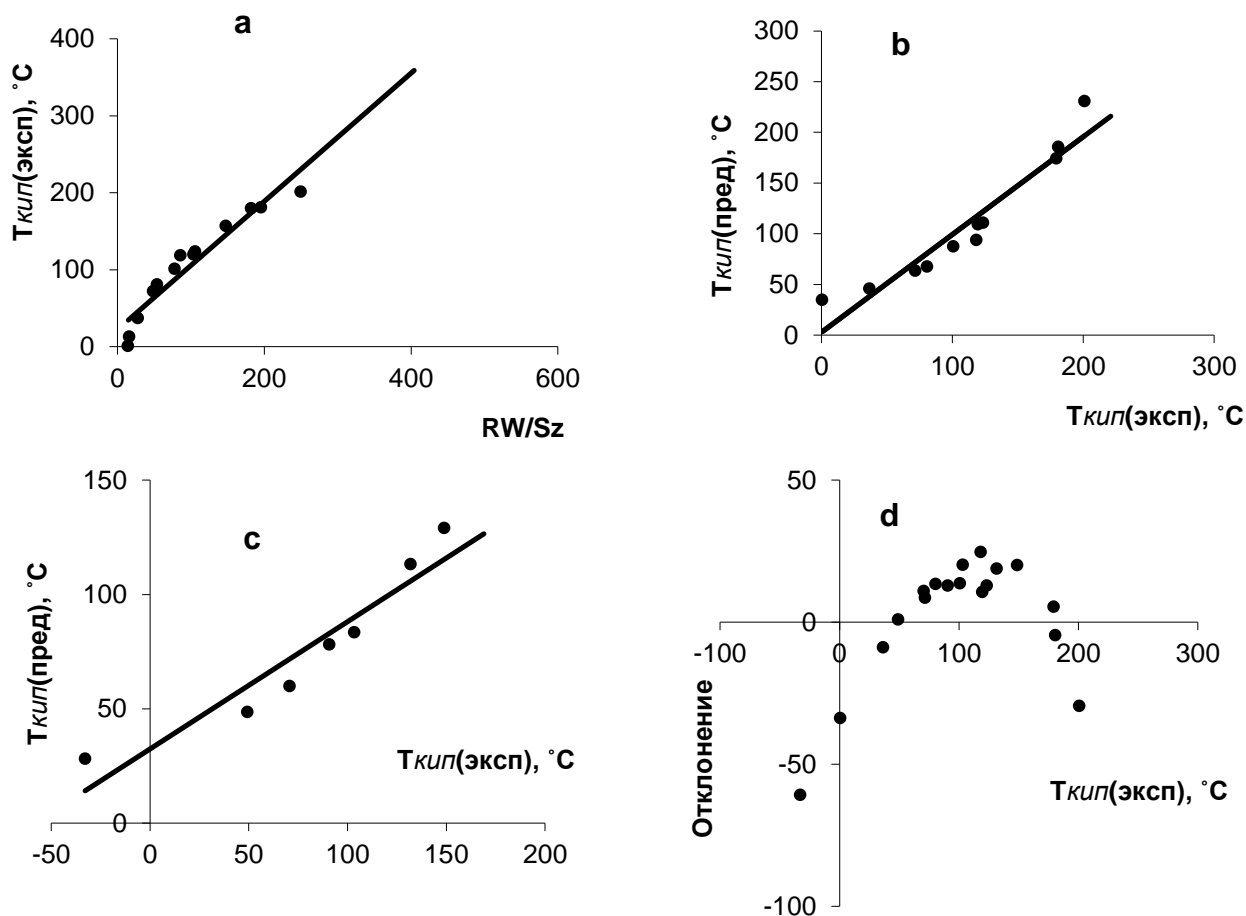


Рисунок 3.2.

QSPR – зависимость, построенная для температуры кипения циклоалканов ($T_{кип}$) с учетом модифицированного индекса Винера (RW) для нечетных колец и индекса Гутмана для четных колец (Sz): а) Соотношение между экспериментальной температурой кипения циклоалканов ($T_{кип}$) и модифицированным индексом Винера (RW) для нечетных колец и индексом Гутмана для четных колец (Sz) для тренировочного набора; б) Соотношение между экспериментальной и предсказанной температурой кипения циклоалканов для тренировочного набора; в) Соотношение между экспериментальной и предсказанной температурой кипения циклоалканов для тестового набора; г) Анализ отклонений между экспериментальными и предсказанными значениями.

Данная модель “структура-свойство” выражающая зависимость между температурой кипения циклоалканов и модифицированным индексом Винера для нечетных колец и индексом Гутмана для четных колец, имеет высокий коэффициент корреляции $r = 0,96$. Но среднеквадратичное стандартное отклонение сравнительно высоко $s = 85,8$ °С. График отклонений выявляет аномальность линейной функциональной зависимости. Можно заключить, что данная модель будет систематически предсказывать ложные значения температуры кипения.

Исходя из этого, построим модель, опираясь на логарифмическую зависимость (рис. 3.3.).

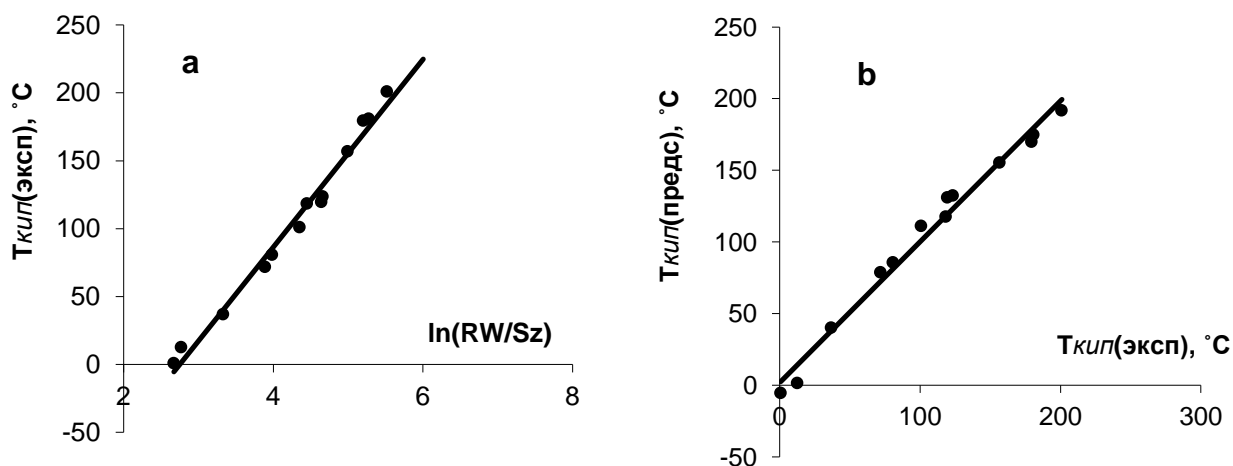


Рисунок 3.3.

QSPR – зависимость, построенная для температуры кипения циклоалканов ($T_{кип}$) с учетом логарифма модифицированного индекса Винера ($\ln(RW)$) для нечетных колец и индекса Гутмана для четных колец (Sz): а) Соотношение между экспериментальной температурой кипения циклоалканов ($T_{кип}$) и логарифмом модифицированного индекса Винера (RW) для нечетных колец и индекса Гутмана для четных колец (Sz) для тренировочного набора; б) Соотношение между экспериментальной и предсказанной температурой кипения циклоалканов для тренировочного набора.

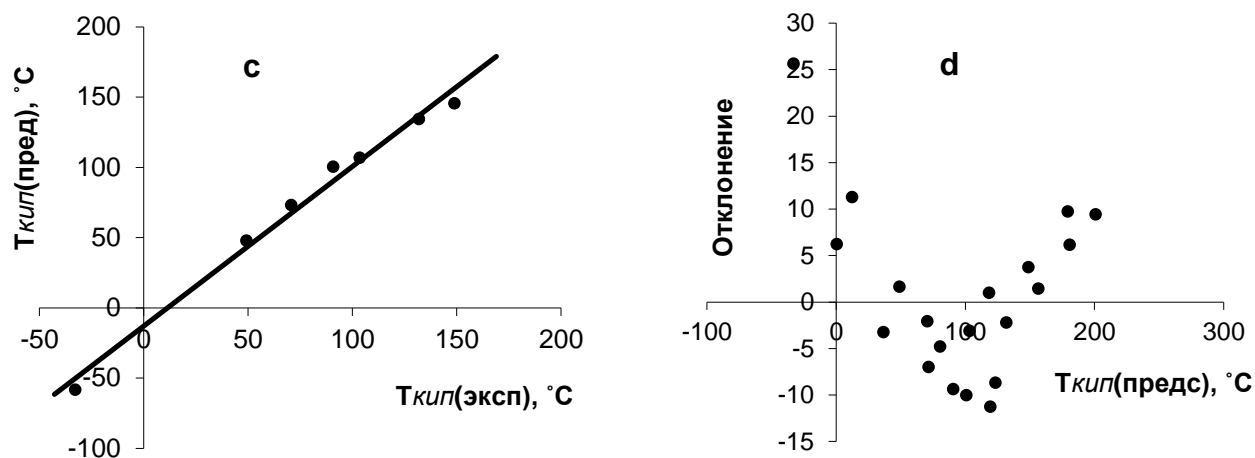


Рисунок 3.3.

QSPR – зависимость, построенная для температуры кипения циклоалканов ($T_{кип}$) с учетом логарифма модифицированного индекса Винера ($\ln(RW)$) для нечетных колец и индекса Гутмана для четных колец (Sz): c) Соотношение между экспериментальной и предсказанной температурой кипения циклоалканов для тестового набора ;d) Анализ отклонений между экспериментальными и предсказанными значениями. (продолжение)

Построенная модель «структура-свойство» улучшила статистические показатели коэффициент корреляции $r = 0,99$ и стандартное отклонение $s = 9^\circ\text{C}$. График отклонений доказывает лучшую прогностическую возможность модели.

3.3. ПОСТРОЕНИЕ QSPR МОДЕЛЕЙ ДЛЯ РАСЧЕТА ТЕМПЕРАТУРЫ КИПЕНИЯ ЦИКЛОАЛКАНОВ С УЧЕТОМ ИНДЕКСА СРЕДНЕГО РАССТОЯНИЯ.

Проанализируем двухмерный график зависимости между индексом среднего расстояния и температурой кипения алканов (рис. 3.4.).

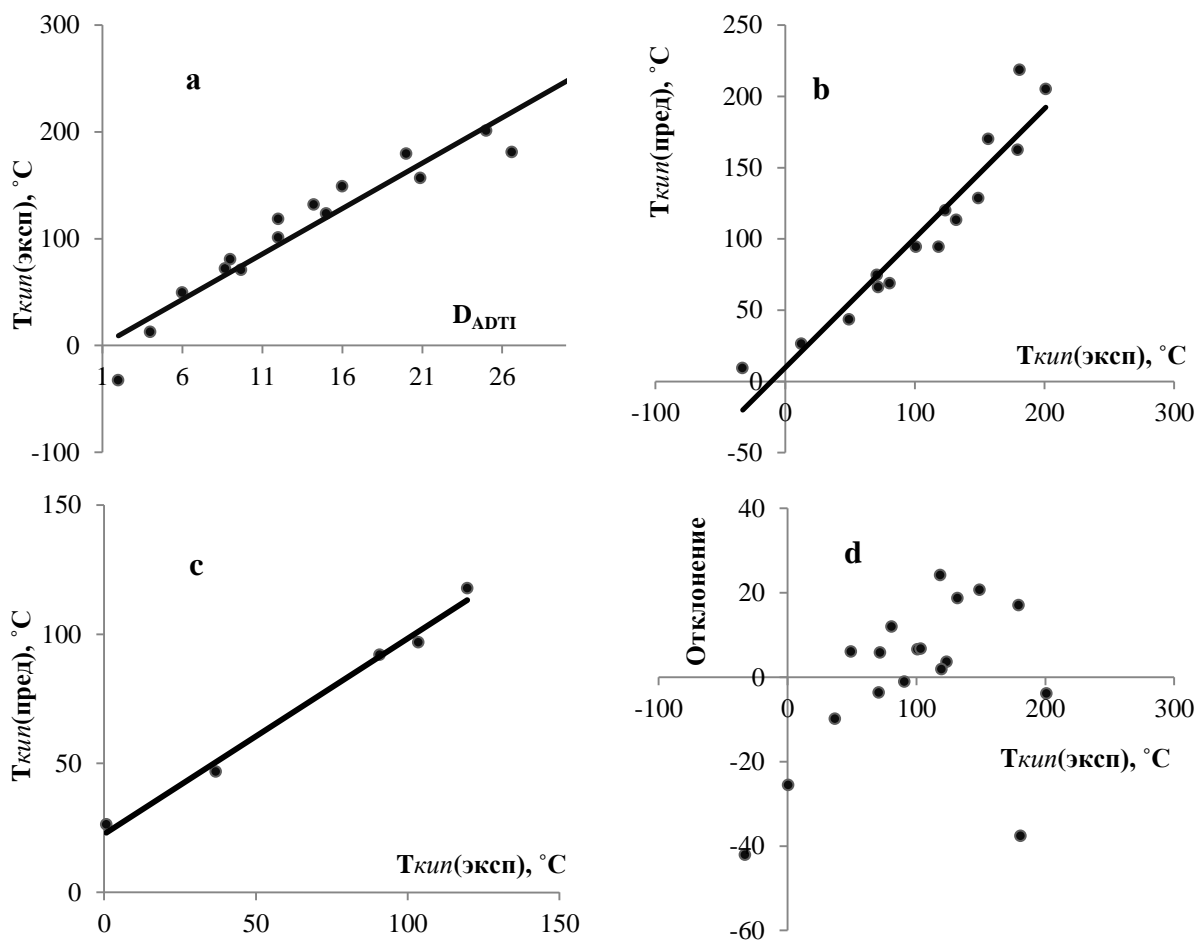


Рисунок 3.4.

QSPR – зависимость, построенная для температуры кипения циклоалканов ($T_{кип}$) с учетом индекса среднего расстояния (D_{ADTI}): а) Соотношение между экспериментальной температурой кипения циклоалканов ($T_{кип}$) и индекса среднего расстояния (D_{ADTI}) для тренировочного набора, б) Соотношение между экспериментальной и предсказанной температурой кипения циклоалканов для тренировочного набора; в) Соотношение между экспериментальной и предсказанной температурой кипения циклоалканов для тестового набора; д) Анализ отклонений между экспериментальными и предсказанными значениями.

Полученная QSPR-модель имеет стандартное отклонение $s = 17,8^{\circ}\text{C}$ и коэффициент корреляции $r = 0,95$. Учитывая график отклонений, можно сделать вывод об уместности выбранной функциональной зависимости. Сравнивая стандартное отклонение с предыдущей моделью, следует поставить цель на улучшение прогностической способности модели.

Попробуем улучшить качество QSPR модели путем использования индекса среднего расстояния в качестве параметра в полиномиальной функции (рис. 3.5.).

Очевиден факт, что корреляция построенной QSPR-модели улучшилась. Подтверждением тому являются статистические характеристики стандартное отклонение $s = 9,96^{\circ}\text{C}$ и коэффициент корреляции $r = 0,987$.

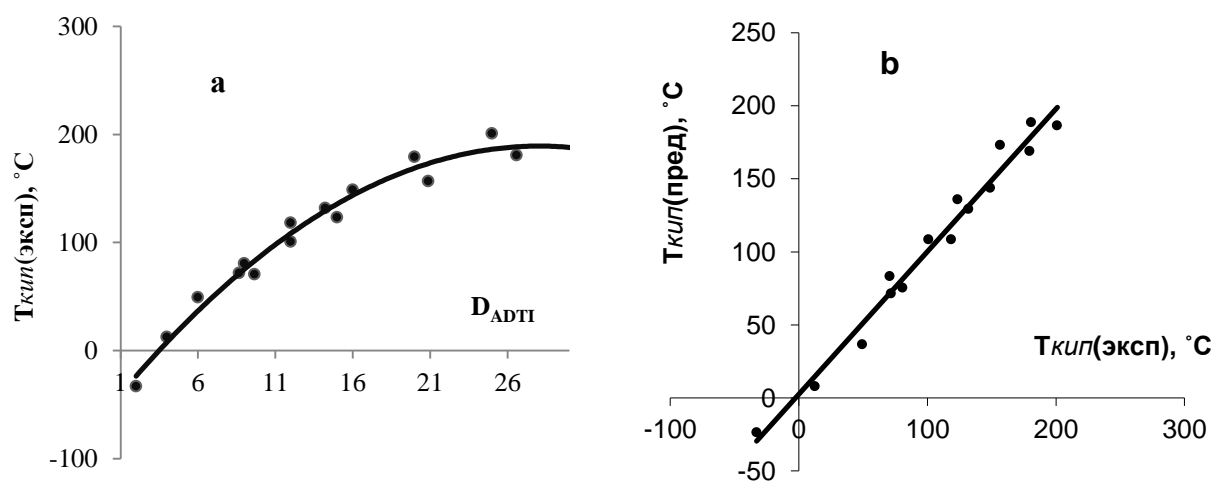


Рисунок 3.5.

QSPR – зависимость, построенная для температуры кипения циклоалканов ($T_{кип}$) с учетом индекса среднего расстояния (D_{ADTI}) посредством полиномиальной функции: а) Соотношение между экспериментальной $T_{кип}$ циклоалканов и индекса среднего расстояния (D_{ADTI}) для тренировочного набора, б) Соотношение между экспериментальной и предсказанной температурой кипения циклоалканов для тренировочного набора.

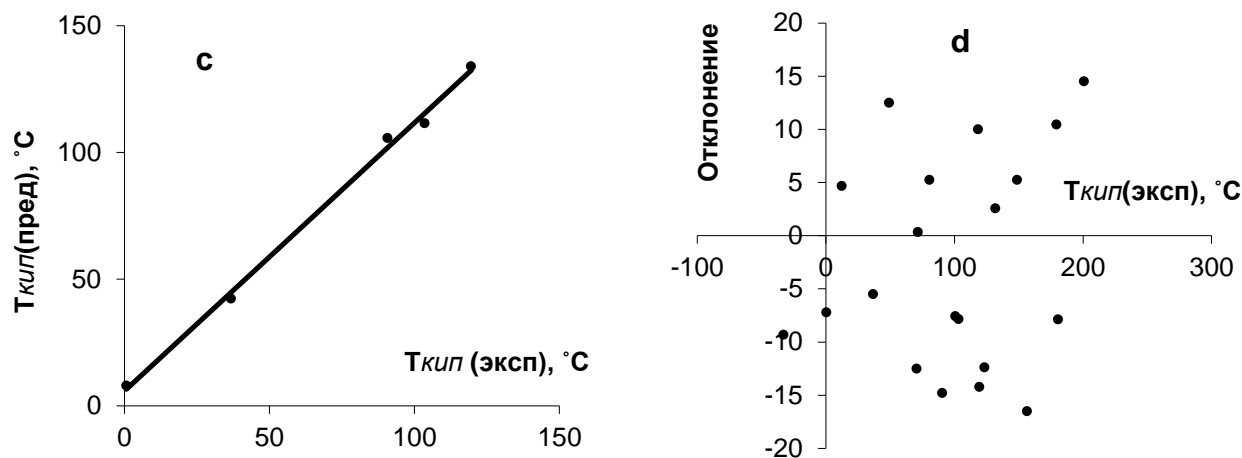


Рисунок 3.5.

QSPR – зависимость, построенная для температуры кипения циклоалканов ($T_{кип}$) с учетом индекса среднего расстояния (D_{ADTI}) посредством полиномиальной функции: с) Соотношение между экспериментальной и предсказанной температурой кипения циклоалканов для тестового набора; d) Анализ отклонений между экспериментальными и предсказанными значениями.

(продолжение)

3.4. ПОСТРОЕНИЕ QSPR - МОДЕЛЕЙ ДЛЯ РАСЧЕТА ЭНТАЛЬПИИ ОБРАЗОВАНИЯ ЦИКЛОАЛКАНОВ В ГАЗОВОЙ ФАЗЕ С УЧЕТОМ ТОПОЛОГИЧЕСКИХ ИНДЕКСОВ.

Исследуем зависимость между энтальпией образования в газовой фазе ($\Delta_f H^\circ_{(г,298К)}$) и индексом Рандича (χ) для циклоалканов (рис. 3.6.) и построим линейно-регрессионную модель “структура-свойство”.

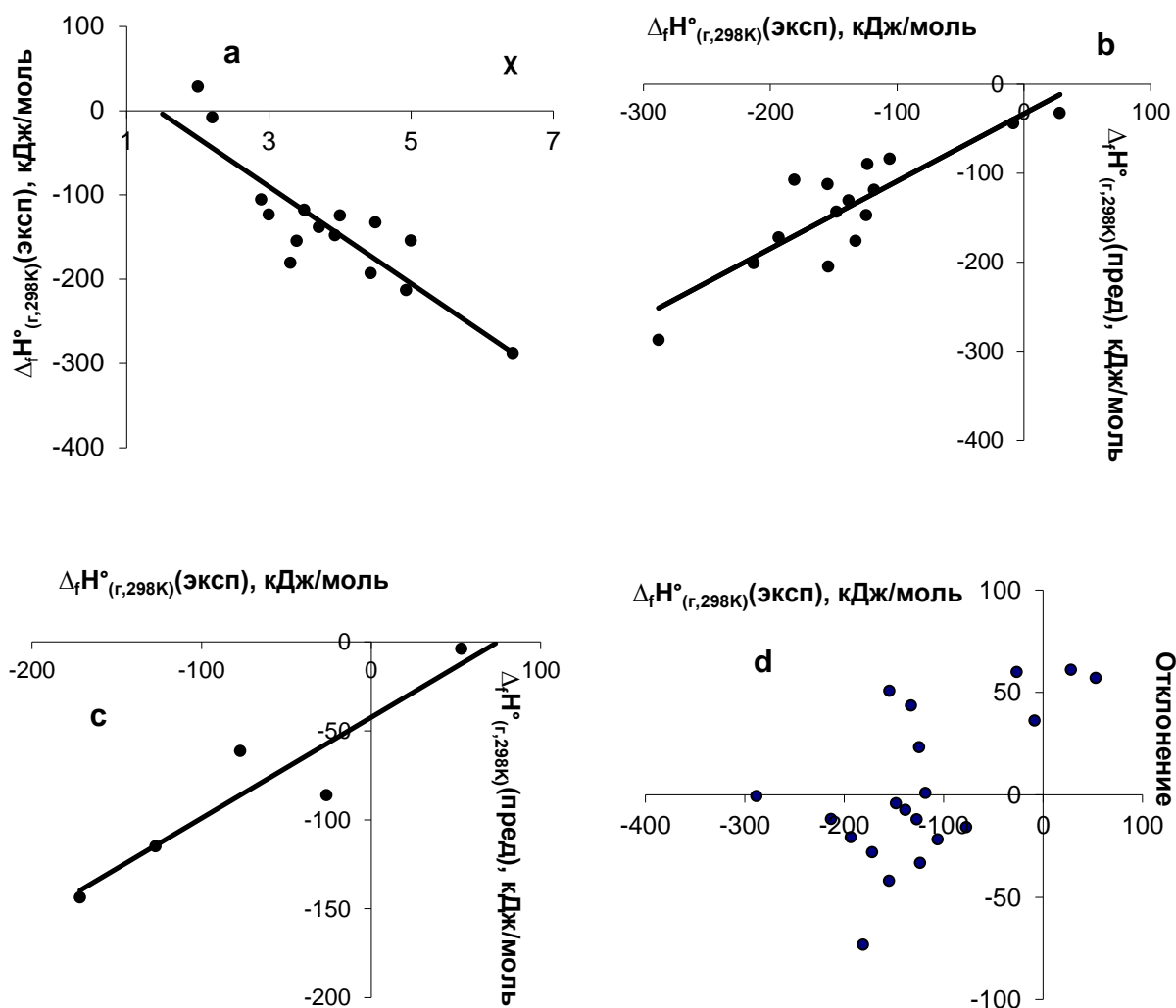


Рисунок. 3.6.

QSPR – зависимость, построенная для энтальпии образования циклоалканов в газовой фазе ($\Delta_f H^\circ_{(r,298K)}$) с учетом индекса Рандича (χ): а) Соотношение между экспериментальной $\Delta_f H^\circ_{(r,298K)}$ циклоалканов и индексом Рандича (χ) для тренировочного набора; б) Соотношение между экспериментальной и предсказанной энтальпией образования циклоалканов для тренировочного набора; в) Соотношение между экспериментальной и предсказанной энтальпией образования алканов для тестового набора; д) Анализ отклонений между экспериментальными и предсказанными значениями.

В результате исследования была построена модель “структура-свойство”, позволяющая вычислять энтальпию образования циклоалканов в газовой фазе с использованием индекса Рандича. Соотношение между значениями энтальпии образования экспериментальной и предсказанной имеет высокую корреляцию $r = 0,87$ и стандартное отклонение $s = 37,9$ кДж/моль. Модель демонстрирует умеренную прогностическую способность.

Построим модель “структура-свойство”, позволяющую рассчитать энтальпию образования циклоалканов в газовой фазе ($\Delta_f H^\circ_{(г,298К)}$) с учетом модифицированного индекса Винера (RW) для нечетных колец и индекса Гутмана для четных колец (Sz) (рис. 3.7.).

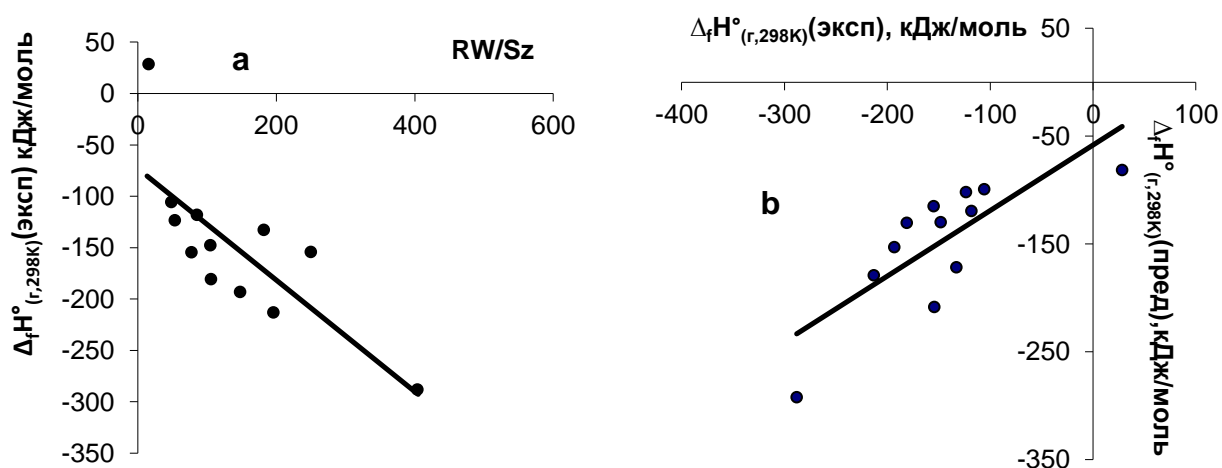


Рисунок 3.7.

QSPR – зависимость, построенная для энтальпии образования циклоалканов в газовой фазе ($\Delta_f H^\circ_{(г,298К)}$) с учетом модифицированного индекса Винера (RW) для нечетных колец и индекса Гутмана для четных колец (Sz): а) Соотношение между экспериментальной энтальпией образования циклоалканов ($\Delta_f H^\circ_{(г,298К)}$) и индексом Винера/Гутмана (RW/Sz) для тренировочного набора; б) Соотношение между экспериментальной и предсказанной энтальпией образования циклоалканов для тренировочного набора.

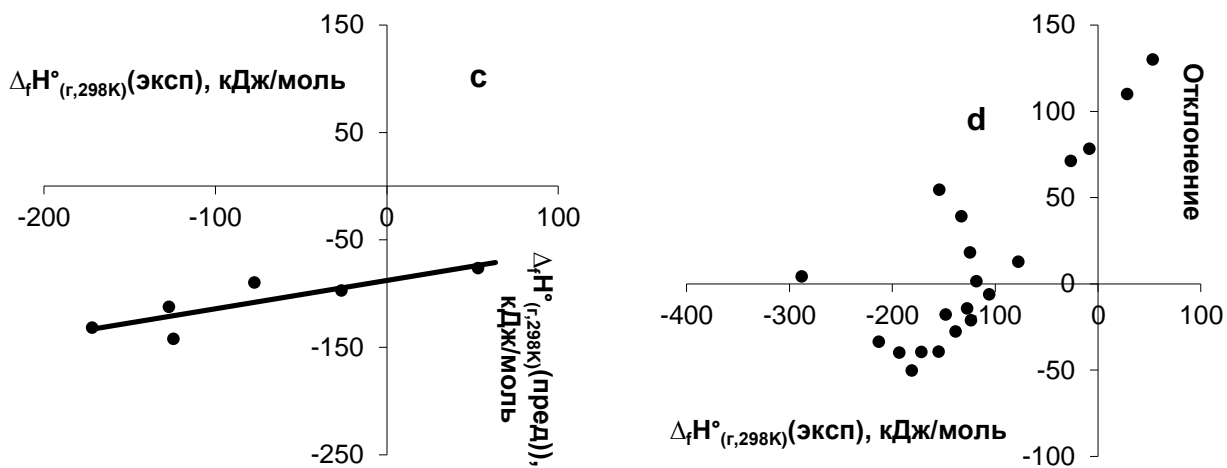


Рисунок 3.7.

QSPR – зависимость, построенная для энтальпии образования циклоалканов в газовой фазе ($\Delta_f H^\circ_{(г,298К)}$) с учетом модифицированного индекса Винера (RW) для нечетных колец и индекса Гутмана для четных колец (Sz): c) Соотношение между экспериментальной и предсказанной энтальпией образования алканов для тестового набора; d) Анализ отклонений между экспериментальными и предсказанными значениями. (продолжение)

Получена модель «структура-свойство» с коэффициентом корреляции $r = 0,78$ и среднеквадратичным стандартным отклонением $s = 59,5$ кДж/моль, статистические параметры которой уступают предыдущей модели.

3.5. ПОСТРОЕНИЕ QSPR - МОДЕЛЕЙ ДЛЯ РАСЧЕТА ЭНТАЛЬПИИ ОБРАЗОВАНИЯ ЦИКЛОАЛКАНОВ В ГАЗОВОЙ ФАЗЕ С УЧЕТОМ ТОПОЛОГИЧЕСКОГО ИНДЕКСА СРЕДНЕГО РАССТОЯНИЯ.

Проанализируем взаимосвязь между энтальпией образования циклоалканов в газовой фазе и топологическим индексом среднего расстояния. Построим двухмерный график (рис. 3.8.).

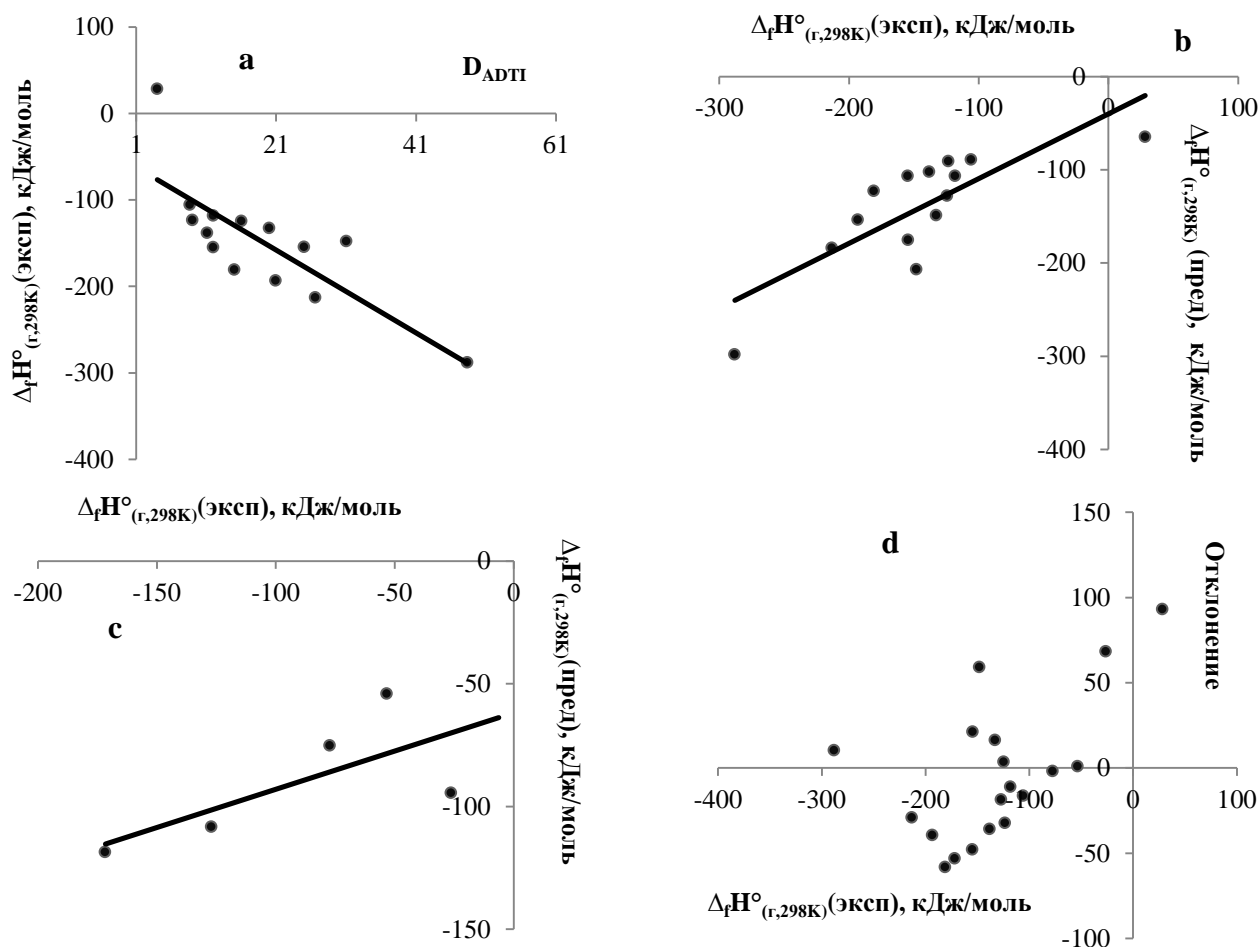


Рисунок 3.8.

QSPR – зависимость энтальпии образования циклоалканов в газовой фазе ($\Delta_f H^\circ_{(г,298К)}$) с учетом индекса среднего расстояния (D_{ADTI}): а) Соотношение между $\Delta_f H^\circ_{(г,298К)}$ циклоалканов и индексом среднего расстояния (D_{ADTI}) для тренировочного набора, б) Соотношение между экспериментальной и предсказанной энтальпией образования циклоалканов для тренировочного набора; в) Соотношение между экспериментальной и предсказанной энтальпией образования алканов для тестового набора; д) Анализ отклонений между экспериментальными и предсказанными значениями.

Статистические характеристики полученной QSPR-модели стандартное отклонение $s=41,7^\circ\text{кДж/моль}$ и коэффициент корреляции $r = 0,79$ указывают, что

модель нужно использовать с осторожностью вследствие значительного расхождения расчетных и экспериментальных данных.

В стремлении улучшить качество модели произведем трансформацию данных и проверим результаты. Построим двухмерный график между энтальпией образования циклоалканов в газовой фазе с учетом логарифма индекса среднего расстояния (рис. 3.9.).

Построенная QSPR-модель имеет следующие статистические показатели коэффициент корреляции $r = 0,89$ и среднеквадратичное стандартное отклонение $s = 33,4$ кДж/моль. Просматривается улучшение прогностического качества QSPR – модели. Следует подчеркнуть, что расхождения между экспериментальными данными уменьшились, но все еще значительны.

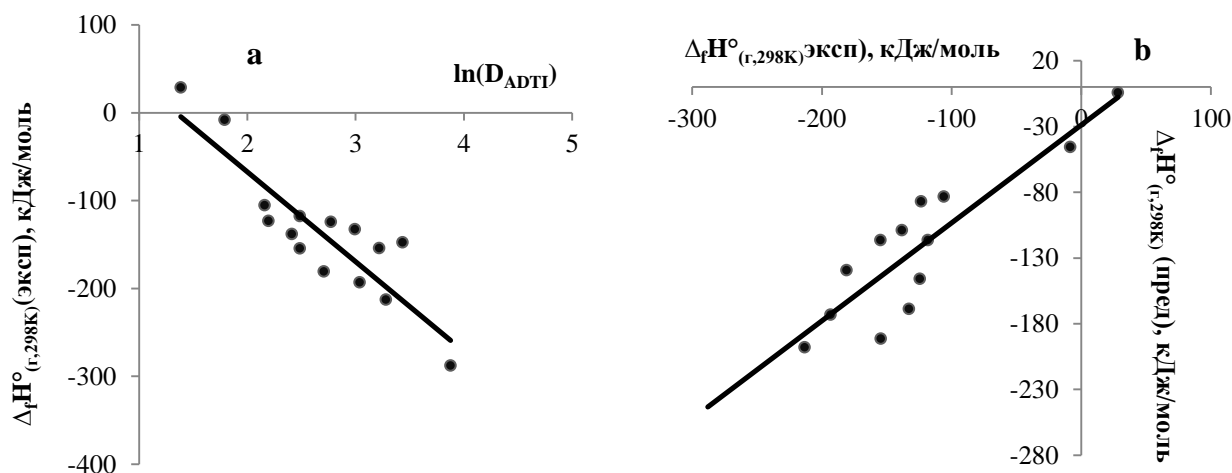


Рисунок 3.9.

QSPR – зависимость, построенная для энтальпии образования циклоалканов в газовой фазе ($\Delta_f H^\circ_{(r,298K)}$) с учетом логарифма индекса среднего расстояния (D_{ADTI}): а) Соотношение между экспериментальной $\Delta_f H^\circ_{(r,298K)}$ циклоалканов и логарифмом индекса среднего расстояния ($\ln(D_{ADTI})$) для тренировочного набора, б) Соотношение между экспериментальной и предсказанной энтальпией образования циклоалканов для тренировочного набора.

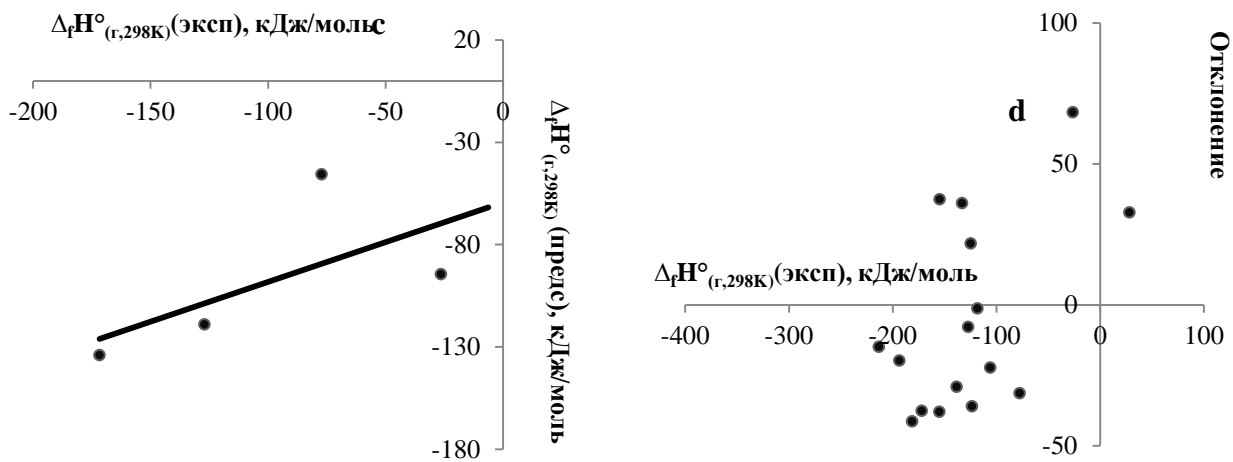


Рисунок 3.9.

QSPR – зависимость, построенная для энтальпии образования циклоалканов в газовой фазе ($\Delta_f H^\circ_{(r,298K)}$) с учетом логарифма индекса среднего расстояния (D_{ADTI}): c) Соотношение между экспериментальной и предсказанной энтальпией образования алканов для тестового набора; d) Анализ отклонений между экспериментальными и предсказанными значениями. (продолжение)

Таблица 3.5.1. QSPR-Уравнения полученные для циклоалканов, статистические характеристики: R^2 - коэффициент детерминации, r - коэффициент корреляции, s - стандартное отклонение.

№	QSPR-модель	R^2	r	s
1	$T_{кин} = 62,661\chi - 110,18$	0,9716	0,99	9,36 °C
2	$T_{кин} = 0,8322RW + 22,448$	0,9169	0,96	85,8 °C
3	$T_{кин} = 69,223\ln(RW) - 190,6$	0,9844	0,99	9,02 °C
4	$T_{кин} = 8,507D_{ADTI} - 7,740$	0,908,	0,95.	17,8°C
5	$T_{кин} = -0,313 D_{ADTI}^2 + 17,58 D_{ADTI} - 57,38$	0,975	0,99	9,96°C
6	$\Delta_f H^\circ_{(r,298K)} = -57,508\chi + 82,462$	0,7576	0,87	66,27 кДж/моль
7	$\Delta_f H^\circ_{(r,298K)} = -0,5431RW - 72,882$	0,6072	0,78	59,48 кДж/моль
8	$\Delta_f H^\circ_{(r,298K)} = -4,783D_{ADTI} - 57,46$	0,629	0,79	41,7°кДж/моль
9	$\Delta_f H^\circ_{(r,298K)} = -102,2 \ln(D_{ADTI}) - 57,46$	0,784	0,89	33,4кДж/моль

3.6. ПРОВЕРКА СТАТИСТИЧЕСКОЙ ГИПОТЕЗЫ ДИСКРИМИНИРУЮЩЕЙ СПОСОБНОСТИ ИНДЕКСА СРЕДНЕГО РАССТОЯНИЯ.

Для оценки дискриминирующей способности топологического индекса среднего расстояния при расчете температуры кипения циклоалканов выдвинем нулевую гипотезу при уровне значимости $\alpha = 0,05$: Стандартное отклонение температуры кипения циклоалканов экспериментальной группы не имеет существенного различия с контрольной группой. Альтернативная гипотеза утверждает существенное различие между экспериментальными и контрольными стандартными отклонениями.

$$H_0: \mu_{\text{эсп.}} = \mu_{\text{контр.}} \quad H_A: \mu_{\text{эсп.}} \neq \mu_{\text{контр.}}$$

При проведении t теста двух выборок получаем величину p -значимости 0,5012. Если альтернативная гипотеза верна и расхождение экспериментальных и контрольных значений стандартного отклонения статистически значимо, то вероятность наблюдаемых значений в двух выборках по чистой случайности очень большая. Таким образом, отличие стандартного отклонения в экспериментальной и контрольной группах несущественно при уровне надежности $\alpha = 0,05$. Статистически достоверные различия не обнаружены и нулевая гипотеза не может быть отвергнута.

Для оценки дискриминирующей способности топологического индекса среднего расстояния при расчете энтальпии образования циклоалканов выдвинем нулевую гипотезу при уровне значимости $\alpha = 0,05$: Стандартное отклонение энтальпии образования циклоалканов экспериментальной группы значительно больше или равно стандартному отклонению контрольной группы. Альтернативная гипотеза указывает, что экспериментальное стандартное отклонение существенно меньше, чем контрольное стандартное отклонение.

$$H_0: \mu_{\text{эсп.}} \geq \mu_{\text{контр.}} \quad H_A: \mu_{\text{эсп.}} < \mu_{\text{контр.}}$$

При проведении одностороннего t теста двух выборок получаем величину p -значимости 0,0226, что позволяет отклонить нулевую гипотезу. Если альтернативная гипотеза верна и экспериментальное стандартное отклонение значительно меньше контрольного стандартного отклонения, то вероятность наблюдаемых значений в двух выборках по чистой случайности чрезвычайно мало. Заключаем, что стандартное отклонение экспериментальной группы существенно ниже контрольной группы при уровне достоверности $\alpha = 0,05$. Это хороший результат, так как цель любого эксперимента уменьшать расхождения с экспериментальными данными.

ГЛАВА 4. ПОСТРОЕНИЕ QSPR - МОДЕЛЕЙ ДЛЯ ПОЛИЦИКЛИЧЕСКИХ АРОМАТИЧЕСКИХ УГЛЕВОДОРОДОВ.

4.1. СТРОЕНИЕ И СВОЙСТВА ПОЛИЦИКЛИЧЕСКИХ АРОМАТИЧЕСКИХ УГЛЕВОДОРОДОВ.

Полициклические ароматические углеводороды (ПАУ) представляют собой высокомолекулярные органические соединения, основным элементом структуры которых является бензольное кольцо. Известны несколько сотен индивидуальных полициклических ароматических углеводородов, различающихся по числу бензольных колец и особенностям их присоединения друг к другу. Это кристаллические соединения (за исключением ряда производных нафталина) с высокой температурой плавления и кипения. Растворимость полициклических ароматических углеводородов в воде сравнительно невелика. Знание физических характеристик ПАУ позволяет определить риск этих соединений для здоровья.

Гидрофобность (коэффициент распределения октанол-вода) – важный параметр органических соединений с точки зрения химии, химической индустрии, биологии, экологии и других научных отраслей. Использование параметров гидрофобности основано на предположении, что полярный растворитель может служить эталонной системой при исследовании взаимодействия активных молекул с липидной фазой. Данная характеристика наиболее часто используется в исследованиях «структура-активность». Коэффициент распределения определяется при помощи стандартной системы 1-октанол – вода. Логарифм коэффициента распределения незаряженных форм субстрата стандартно обозначается $\log P$ и вычисляется по следующей формуле:

$$\log P_{oct/wat} = \log \left(\frac{[solute]_{octanol}}{[solute]_{water}} \right).$$

Таблица 4.1.1. Данные для ПАУ: $T_{пл}$ – температура плавления, $T_{кип}$ – температура кипения, $LogP$ – коэффициент распределения в системе октанол – вода, W – индекс Винера, χ – индекс Рандича, D_{ADTI} – индекс среднего расстояния.

Название вещества	$T_{пл}, ^\circ C$	$T_{кип}, ^\circ C$	$LogP$	LEL	W	X	D_{ADTI}
Нафталин	81,00	218	3,35	13,341	109	4,966	21,80
1-метилнафталин	-22,00	245	3,87	14,572	140	5,377	25,45
2-метилнафталин	35,00	241	4	14,575	144	5,36	26,18
1-этилнафталин	-14,00	259	4,39	15,837	182	5,915	30,33
2-этилнафталин	-7,00	258	4,38	15,841	190	5,898	31,67
2-6-диметилнафталин	110,00	262	4,31	15,808	186	5,754	31,00
2-7-диметилнафталин	97,00	262		15,808	185	5,754	30,83
1-7-диметилнафталин	-14,00	263	4,44	15,805	180	5,771	30,00
1-5-диметилнафталин	80,00	269	4,31	15,802	176	5,788	29,33
1-2-диметилнафталин	-4,00	271	4,31	15,803	178	5,788	29,67
1-3-7-триметилнафталин	14,00	280		17,037	226	6,165	34,77
2-3-5-триметилнафталин	25,00	285		17,035	224	6,182	34,46
2-3-6-триметилнафталин	101,00	286	4,73	17,038	230	6,165	35,38
Фенален	85,00			17,919	210	6,449	32,31
1-фенилнафталин	45,00	334		21,739	412	7,949	51,50
2-фенилнафталин	104,00	360		21,744	436	7,933	54,50
Антрацен	216,00	340	4,5	19,197	279	6,933	39,86
1-метилантрацен	86,00	363		20,426	334	7,343	35,16
2-метилантрацен	209,00			20,429	342	7,327	45,60
2-7-диметилантрацен	241,00	370		21,66	413	7,72	51,63
2-6-диметилантрацен	250,00	370		21,66	414	7,72	51,75
2-3-диметилантрацен	252,00			21,658	408	7,737	51,00

Таблица 4.1.1.

Название вещества	Тпл, °С	Ткип, °С	LogP	LEL	W	X	DADTI
9-10-диметилантрацен	183,00		5,69	21,646	378	7,788	47,25
Фенантрен	101,00	338	4,52	19,194	271	6,949	38,71
1-метилфенантрен	123,00	359	5,08	20,422	326	7,36	43,47
2-метилфенантрен	56,00	355	5,24	20,425	334	7,343	44,53
3-метилфенантрен	65,00	352	5,15	20,425	330	7,343	44,00
4-метилфенантрен	50,00			20,422	322	7,36	42,93
9-метилфенантрен	91,00	355		20,421	322	7,36	42,93
3-6-диметилфенантрен	141,00	363		21,656	396	7,737	49,50
4-5-метилефенантрен	116,00	359		21,195	300	7,433	40,00
Тетрацен	257,00		5,76	25,047	569	8,899	63,22
Бенз[а]антрацен	162,00		5,91	25,043	553	8,916	61,44
Кризен	256,00	441	5,86	25,039	545	8,933	60,56
Бенз[с]фенантрен	68,00			25,038	529	8,933	58,78
Трифенилен	199,00	439	5,49	25,032	513	8,949	57,00
Пирен	156,00	393	5	22,49	362	7,933	45,25
1-метилпирен	70,00	410		23,717	428	8,343	50,35
2-метилпирен	144,00	410		23,72	434	8,327	51,06
4-метилпирен	148,00	410		23,717	424	8,343	49,88
Пентацен	271,00			30,894	1011	10,865	91,91
Дибенз[ai]антрацен	264,00		6,81	30,889	987	10,882	89,73
Дибенз[ah]антрацен	270,00		5,8	30,885	971	10,899	88,27
Дибенз[aj]антрацен	198,00			30,885	955	10,899	86,82
Бенз[b]кризен	294,00			30,885	971	10,899	88,27
Дибенз[ac]антрацен	205,00			30,877	907	10,916	82,45
Бенз[a]пирен	177,00	496	5,97	28,331	680	9,916	68,00

Таблица 4.1.1.

Название вещества	Тпл, °С	Ткип, °С	LogP	LEL	W	X	DADTI
Бенз[е]пирен	179,00	493		28,325	652	9,933	65,20
Перилен	278,00		6,25	28,326	654	9,933	65,40
Коронен	360,00		6,5	34,906	1002	11,899	83,50
Антрактрен	261,00			31,621	839	10,899	76,27
бенз[ghi]перилен	283,00		6,9	31,617	815	10,916	74,09
дибенз[ae]пирен	234,00			34,163	1082	11,916	90,17
1-метилкризен	161,00			26,265	620	9,343	65,26
6-метилкризен	257,00			26,267	632	9,343	66,53
3-метилколантрен	180,00		6,75	29,54	804	10,327	76,57
indeno[1-2-3-cd]пирен	163,00			31,599	845	10,916	76,82
Пентафен	263,00			30,889	979	10,882	89,00
Гексафен	308,00			36,734	1589	12,848	122,23
Индан	-51,00	178		12,043	79	4,466	17,56
Инден	-2,00	183	2,92	12,043	79	4,466	17,56
Азулен	100,00	270	3,22	13,335	107	4,966	21,40
Аценафтен	96,00	279	3,92	16,624	166	5,949	27,67
Аценафтилен	93,00	270		16,624	166	5,949	27,67
Флуорен	117,00	294	4,18	17,899	219	6,449	33,69
1-метилфлуорен	87,00	318	4,97	19,128	267	6,86	38,14
2-метилфлуорен	104,00	318		19,131	274	6,843	39,14
3-метилфлуорен	88,00	316		19,131	272	6,843	38,86
4-метилфлуорен	71,00			19,128	265	6,86	37,86
9-метилфлуорен	47,00			19,125	262	6,877	37,43
1-2-бензфлуорен	190,00	407	5,4	23,746	461	8,433	54,24
Флуорантен	111,00	383	5,2	22,466	364	7,949	45,50

Таблица 4.1.1.

Название вещества	Tпл, °С	Tкип, °С	LogP	LEL	W	X	DADTI
2-3-бензфлуорен	209,00	402	5,75	23,75	471	8,416	55,41
3-4-бензфлуорен	125,00	406		23,745	453	8,433	53,29
бенз[ghi]флуорантен	149,00	432	5,78	25,759	478	8,933	47,80
бенз[k]флуорантен	217,00	481		28,313	698	9,916	69,80
бенз[b]флуорантен	168,00	481		28,307	676	9,933	67,60
бенз[j]флуорантен	166,00	480		28,309	678	9,933	67,80
Овален	473,00			47,307	2106	15,865	131,63
Кватерилен	483,00			58,242	4544	19,865	227,20

4.2. ПОСТРОЕНИЕ QSPR МОДЕЛЕЙ ДЛЯ РАСЧЕТА ТЕМПЕРАТУРЫ ПЛАВЛЕНИЯ ПАУ И С УЧЕТОМ ТОПОЛОГИЧЕСКИХ ИНДЕКСОВ.

Исследуем корреляцию между температурой плавления ПАУ ($T_{пл}$) и индексом Рандича (χ). Для этого построим двухмерный график (рис. 4.1.) и определим уравнение зависимости между величинами температуры плавления и индексом Рандича и ее статистические параметры.

Полученная модель “структура-свойство” представляет линейную зависимость между температурой плавления ПАУ и индексом Рандича. Руководствуясь ее статистическими данными $r = 0,8$ и $s = 55,01^\circ\text{C}$, можно заключить об умеренной прогностической возможности.

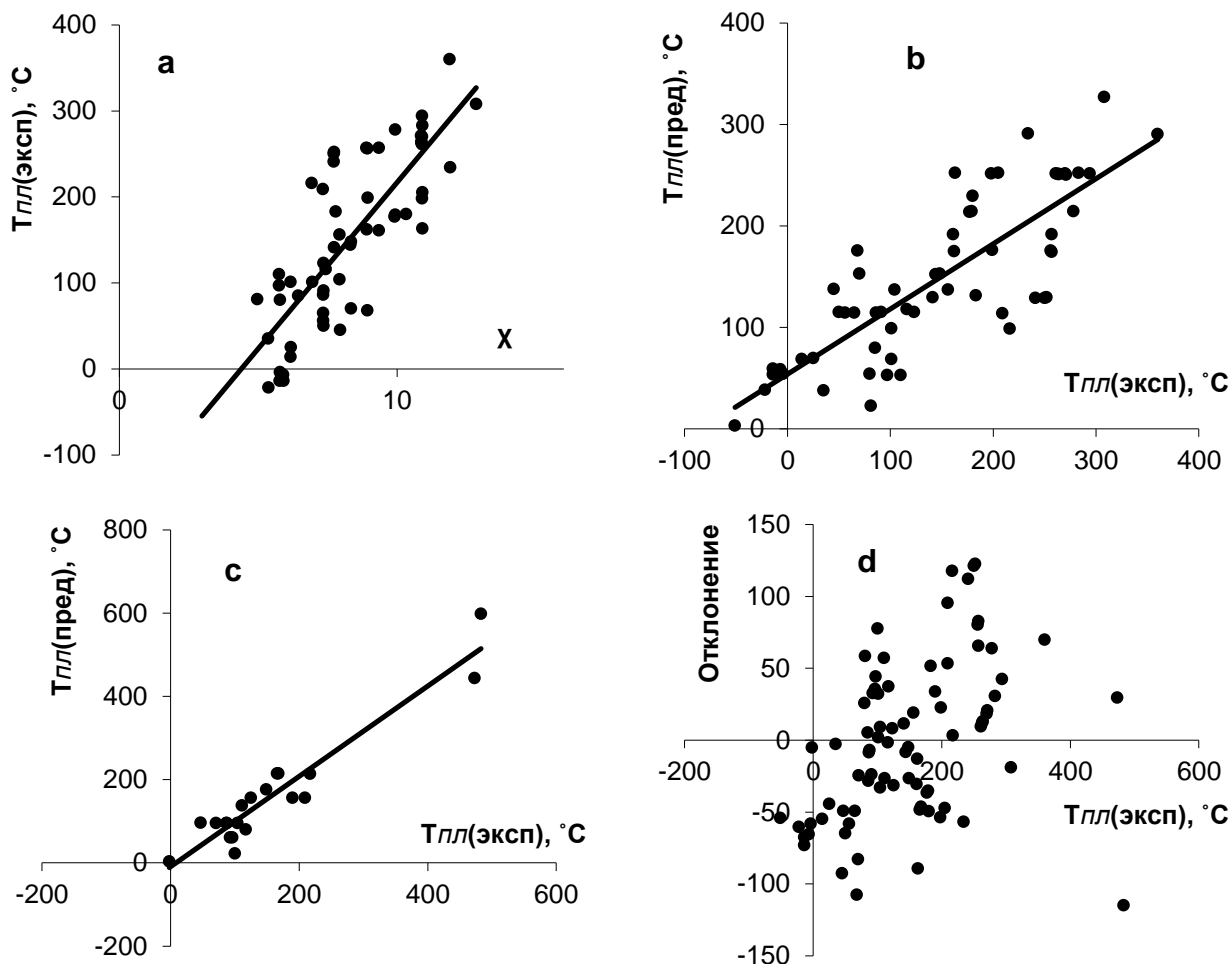


Рисунок 4.1.

QSPR – зависимость, построенная для температуры плавления ($T_{пл}$) полициклических ароматических углеводородов с учетом индекса Рандича (χ):

а) Соотношение между экспериментальной температурой плавления ПАУ ($T_{пл}$) и индексом Рандича (χ) для тренировочного набора; б) Соотношение между экспериментальной и предсказанной температурой плавления ($T_{пл}$) ПАУ для тренировочного набора; в) Соотношение между экспериментальной и предсказанной температурой плавления ($T_{пл}$) ПАУ для тестового набора; г) Анализ отклонений между экспериментальными и предсказанными значениями температуры плавления ($T_{пл}$) ПАУ.

Изучим зависимость между температурой плавления ПАУ ($T_{пл}$) и индексом Винера (W) (рис. 4.2.).

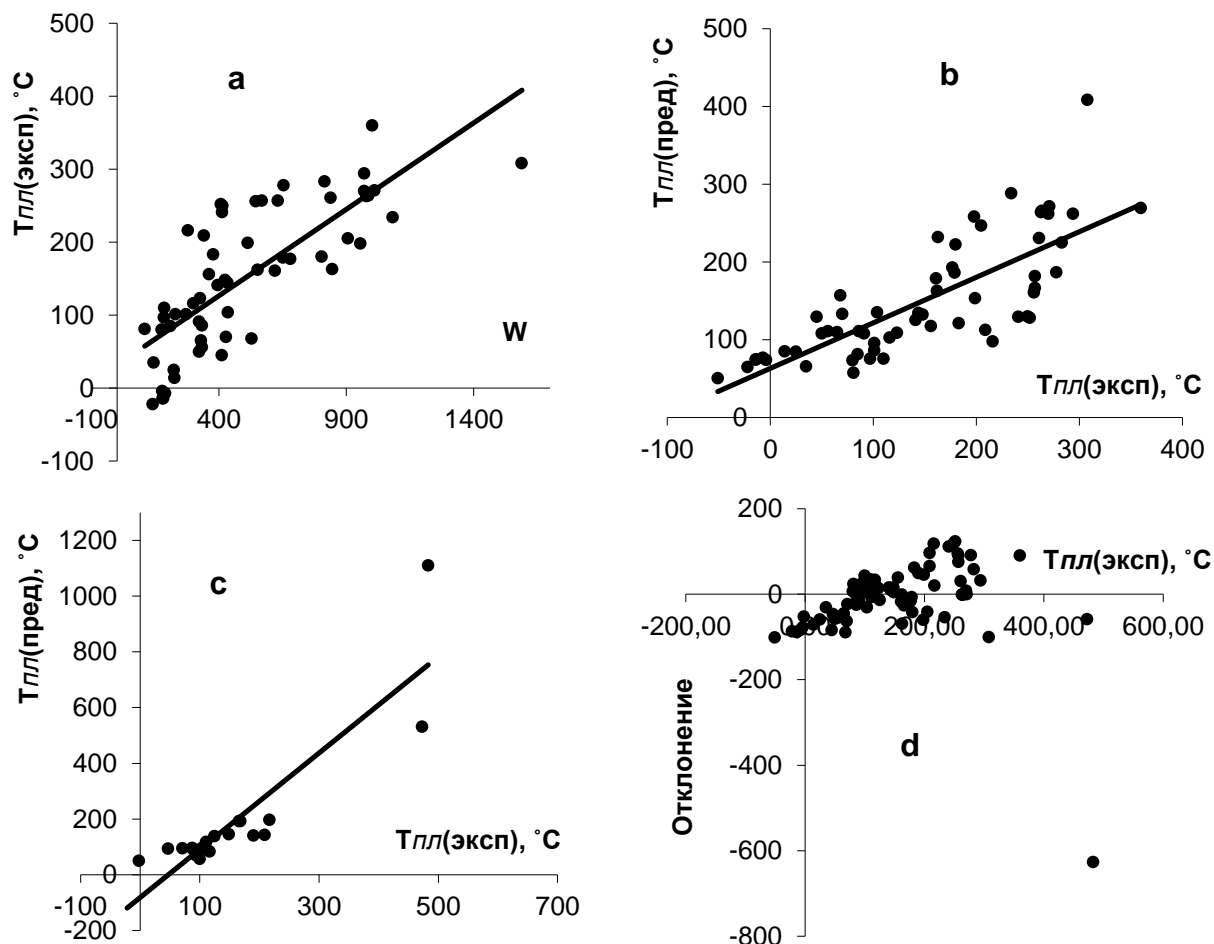


Рисунок 4.2.

QSPR – зависимость, построенная для температуры плавления ($T_{пл}$) полициклических ароматических углеводородов с учетом индекса Винера (W):
а) Соотношение между экспериментальной температурой плавления ПАУ ($T_{пл}$) и индексом Винера (W) для тренировочного набора; б) Соотношение между экспериментальной и предсказанной температурой плавления ($T_{пл}$) ПАУ для тренировочного набора; в) Соотношение между экспериментальной и предсказанной температурой плавления ($T_{пл}$) ПАУ для тестового набора; г) Анализ отклонений между экспериментальными и предсказанными значениями температуры плавления ($T_{пл}$) ПАУ.

Построенная модель “структура-свойство” имеет корреляцию $r = 0,77$ и среднеквадратичное стандартное отклонение $s = 90,02^\circ\text{C}$. Она уступает в прогностической возможности модели, полученной с применением индекса Рандича.

Проанализируем влияние величины инварианта графа на основе матрицы Лапласа $LEL(G)$ на температуру плавления ПАУ ($T_{пл}$) (рис. 4.3.).

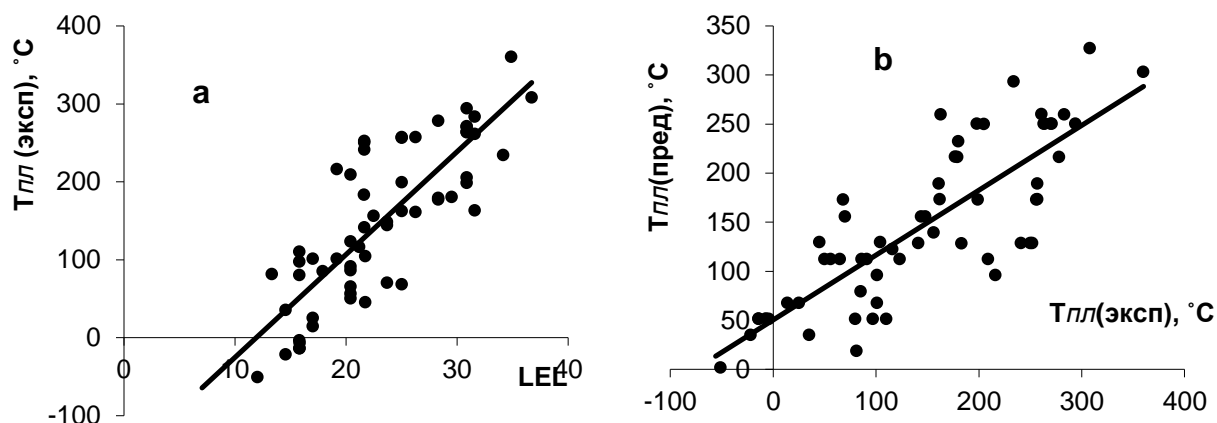


Рисунок 4.3.

QSPR – зависимость, построенная для температуры плавления ($T_{пл}$) полициклических ароматических углеводородов с учетом величины инварианта графа на основе матрицы Лапласа $LEL(G)$: а) Соотношение между экспериментальной температурой плавления ПАУ ($T_{пл}$) и величиной инварианта графа на основе матрицы Лапласа $LEL(G)$ для тренировочного набора; б) Соотношение между экспериментальной и предсказанной температурой плавления ($T_{пл}$) ПАУ для тренировочного набора;

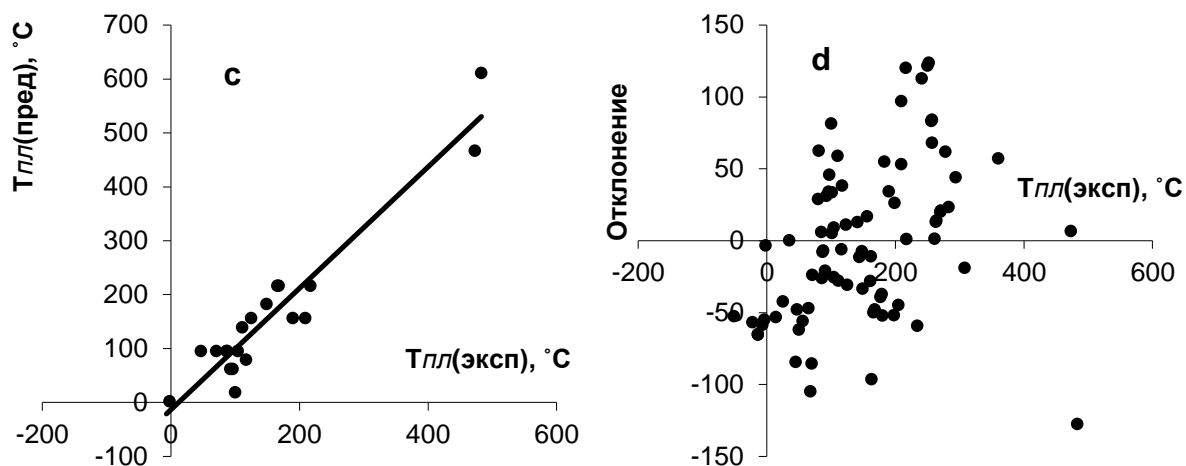


Рисунок 4.3.

QSPR – зависимость, построенная для температуры плавления ($T_{пл}$) полициклических ароматических углеводородов с учетом величины инварианта графа на основе матрицы Лапласа $LEL(G)$: c) Соотношение между экспериментальной и предсказанной температурой плавления ($T_{пл}$) ПАУ для тестового набора; d) Анализ отклонений между экспериментальными и предсказанными значениями температуры плавления ($T_{пл}$) ПАУ.

(продолжение)

Выведенное уравнение имеет следующие статистические параметры $r = 0,95$ и среднеквадратичное статистическое отклонение $s = 55,06^{\circ}\text{C}$. Эта модель имеет более высокую предсказательную возможность, чем зависимость от индекса Рандича.

Проанализируем корреляцию между температурой плавления ПАУ ($T_{пл}$) и комбинацией индексов Рандича (χ) и Винера (W) (рис. 4.4.).

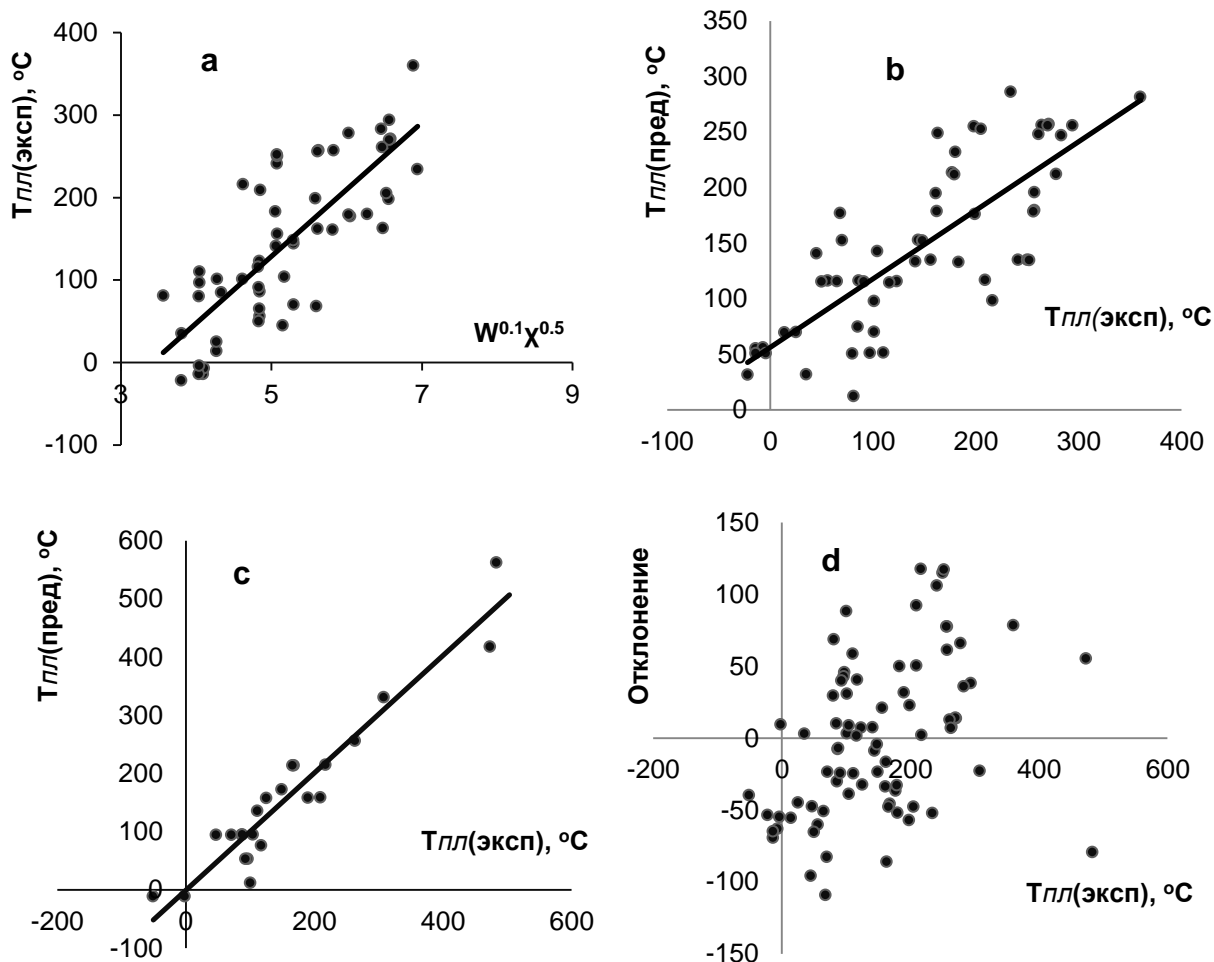


Рисунок 4.4.

QSPR – зависимость температуры плавления ($T_{пл}$) полициклических ароматических углеводородов с учетом индексов Рандича (χ) и Винера (W):
 а) Соотношение между экспериментальной температурой плавления ПАУ ($T_{пл}$) и индексом Рандича (χ) для тренировочного набора; б) Соотношение между экспериментальной и предсказанной температурой плавления ($T_{пл}$) ПАУ для тренировочного набора; в) Соотношение между экспериментальной и предсказанной температурой плавления ($T_{пл}$) ПАУ для тестового набора; г) Анализ отклонений между экспериментальными и предсказанными значениями температуры плавления ($T_{пл}$) ПАУ.

Статистические параметры этой модели корреляция $r = 0,95$ и среднее стандартное отклонение $54,10^{\circ}\text{C}$ улучшились и свидетельствуют о повышении прогностической возможности.

4.3. ПОСТРОЕНИЕ QSPR - МОДЕЛЕЙ ДЛЯ РАСЧЕТА ТЕМПЕРАТУРЫ ПЛАВЛЕНИЯ ПАУ С УЧЕТОМ ИНДЕКСА СРЕДНЕГО РАССТОЯНИЯ.

Воспользуемся данными температуры плавления полициклических ароматических углеводородов и значениями индекса среднего расстояния для графического представления зависимости обучающей выборки объектов исследования (рис. 4.5.).

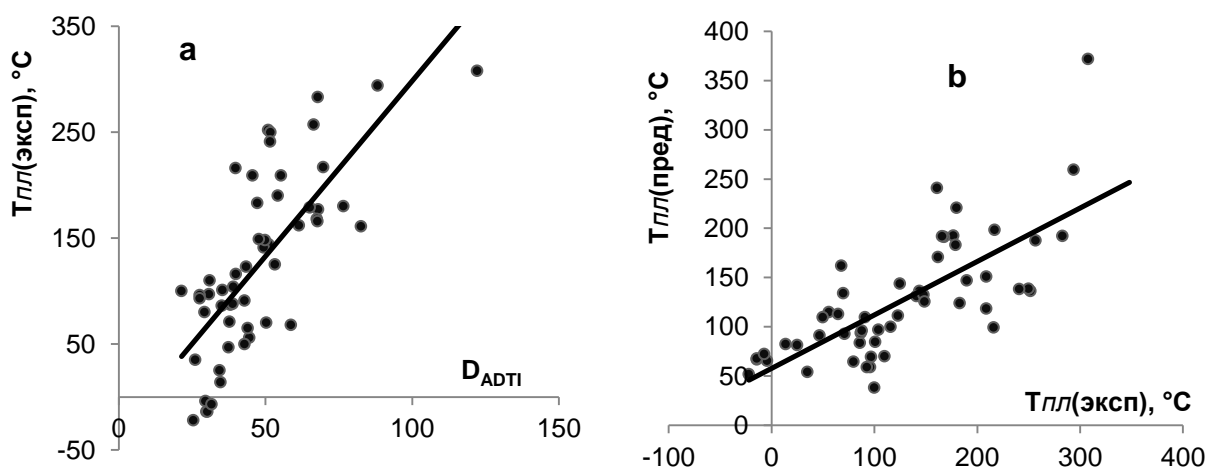


Рисунок 4.5.

QSPR – зависимость, построенная для температуры плавления ($T_{пл}$) полициклических ароматических углеводородов с учетом индекса среднего расстояния D_{ADTI} : а) Соотношение между экспериментальной температурой плавления ПАУ ($T_{пл}$) и индексом среднего расстояния D_{ADTI} для обучающего набора; б) Соотношение между экспериментальной и предсказанной температурой плавления ($T_{пл}$) ПАУ для обучающего набора;

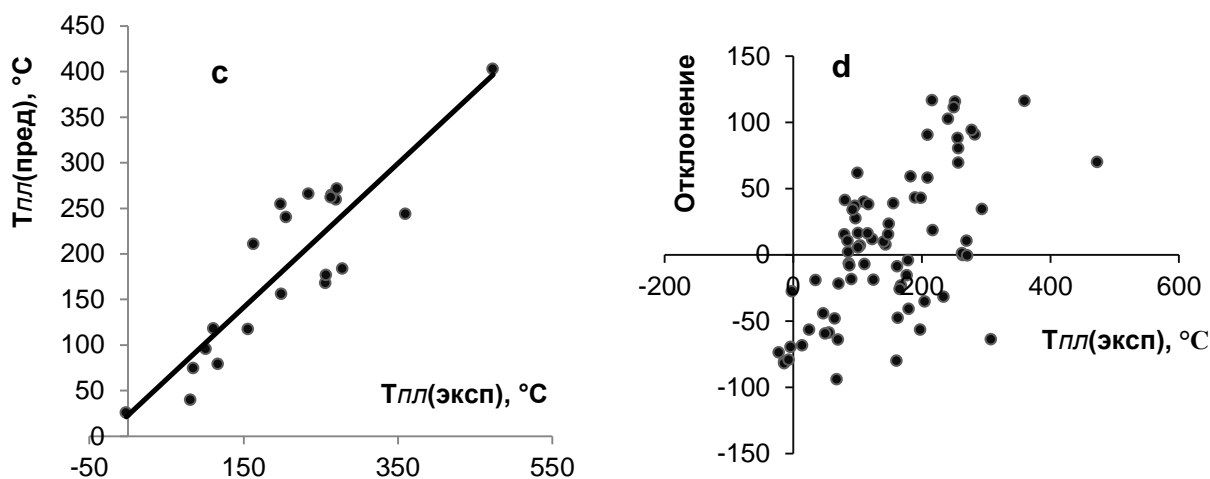


Рисунок 4.5.

QSPR – зависимость, построенная для температуры плавления ($T_{пл}$) полициклических ароматических углеводородов с учетом индекса среднего расстояния D_{ADTI} : c) Соотношение между экспериментальной и предсказанной температурой плавления ПАУ для контрольного набора; d) Анализ отклонений между экспериментальными и предсказанными значениями температуры плавления ($T_{пл}$) ПАУ. (продолжение)

Полученная модель «структура-свойство» имеет более низкие статистические показатели $r = 0,74$ и $s = 54,6^{\circ}\text{C}$. Значение коэффициента корреляции свидетельствует об умеренной взаимосвязи между индексом среднего расстояния и температурой плавления полициклических ароматических углеводородов.

Исследуем зависимость на предмет полиномиальной зависимости между температурой плавления ПАУ и топологическим индексом среднего расстояния. Обратимся к двумерному графику (рис. 4.6.).

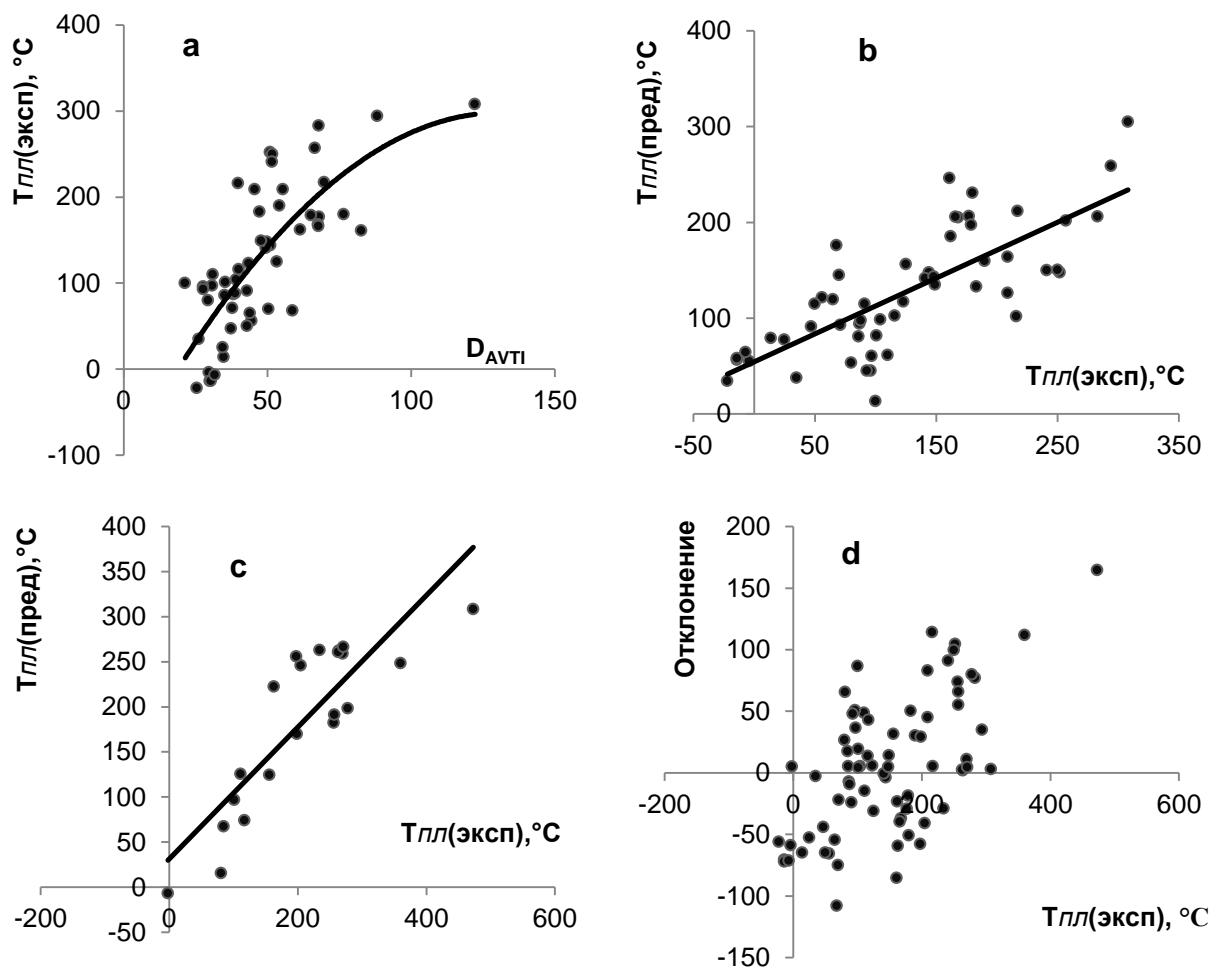


Рисунок 4.6.

QSPR – зависимость, построенная для температуры плавления ($T_{пл}$) полициклических ароматических углеводородов с учетом индекса среднего расстояния $D_{АВТИ}$ в виде полиномиальной функции: а) Соотношение между экспериментальной температурой плавления ПАУ ($T_{пл}$) и индексом среднего расстояния $D_{АВТИ}$ для обучающего набора; б) Соотношение между экспериментальной и предсказанной температурой плавления ($T_{пл}$) ПАУ для обучающего набора; в) Соотношение между экспериментальной и предсказанной температурой плавления ПАУ для контрольного набора; г) Анализ отклонений между экспериментальными и предсказанными значениями температуры плавления ($T_{пл}$) ПАУ.

Полученная количественная модель “структура-свойство” не привела к значительному повышению предсказательной способности для расчета температуры плавления ПАУ. Статистические параметры $r = 0,76$ и $s = 55,4^{\circ}\text{C}$.

4.4. ПОСТРОЕНИЕ QSPR МОДЕЛЕЙ ДЛЯ РАСЧЕТА ТЕМПЕРАТУРЫ КИПЕНИЯ ПАУ С УЧЕТОМ ТОПОЛОГИЧЕСКИХ ИНДЕКСОВ.

Исследуем корреляцию между температурой кипения ПАУ ($T_{кип}$) и индексом Рандича (χ) (рис. 4.7.).

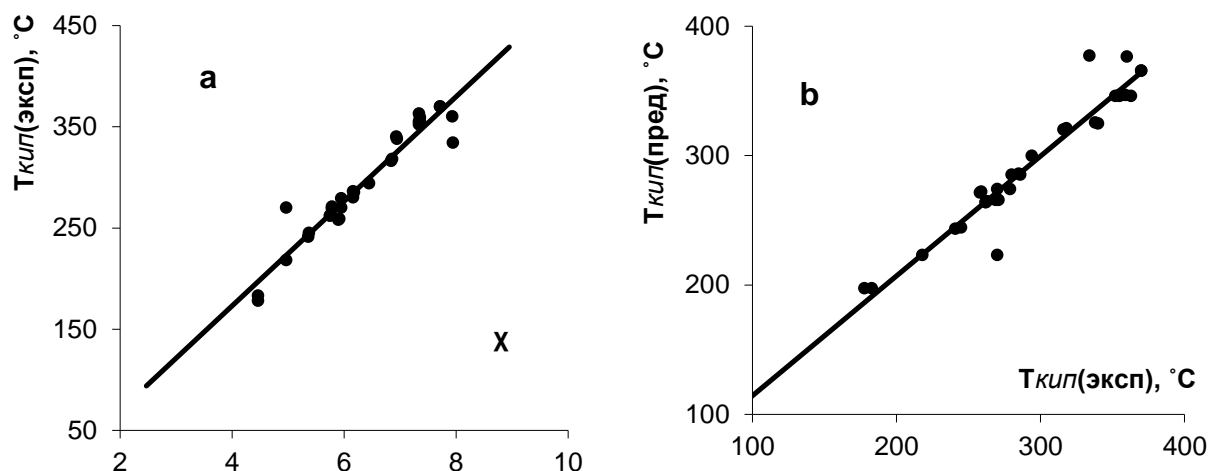


Рисунок 4.7.

QSPR – зависимость, построенная для температуры кипения ($T_{кип}$) полициклических ароматических углеводородов с учетом индекса Рандича (χ):
а) Соотношение между экспериментальной температурой кипения ПАУ ($T_{кип}$) и индексом Рандича (χ) для тренировочного набора; б) Соотношение между экспериментальной и предсказанной температурой кипения ($T_{кип}$) ПАУ для тренировочного набора.

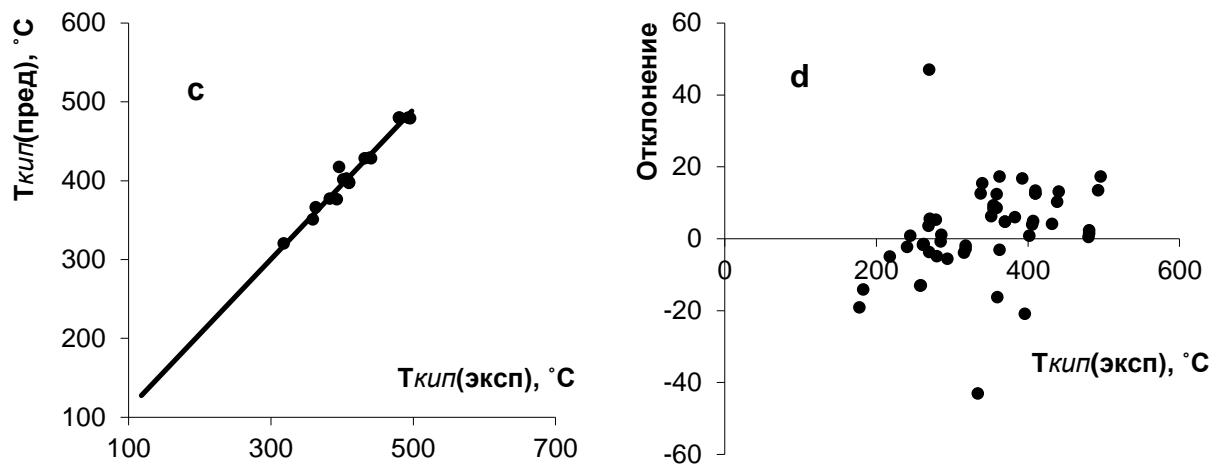


Рисунок 4.7.

QSPR – зависимость, построенная для температуры кипения ($T_{кип}$) полициклических ароматических углеводородов с учетом индекса Рандича (χ):
 с) Соотношение между экспериментальной и предсказанной температурой кипения ($T_{кип}$) ПАУ для тестового набора; d) Анализ отклонений между экспериментальными и предсказанными значениями температуры кипения ($T_{кип}$) ПАУ. (продолжение)

Статистические параметры полученной модели «структура-свойство» коэффициент корреляции $r = 0,98$ и среднеквадратичное стандартное отклонение $s = 12,86^\circ\text{C}$ свидетельствуют о высокой предсказательной способности.

Исследуем корреляцию между температурой кипения ПАУ ($T_{кип}$) и индексом Винера (W) (рис. 4.8.).

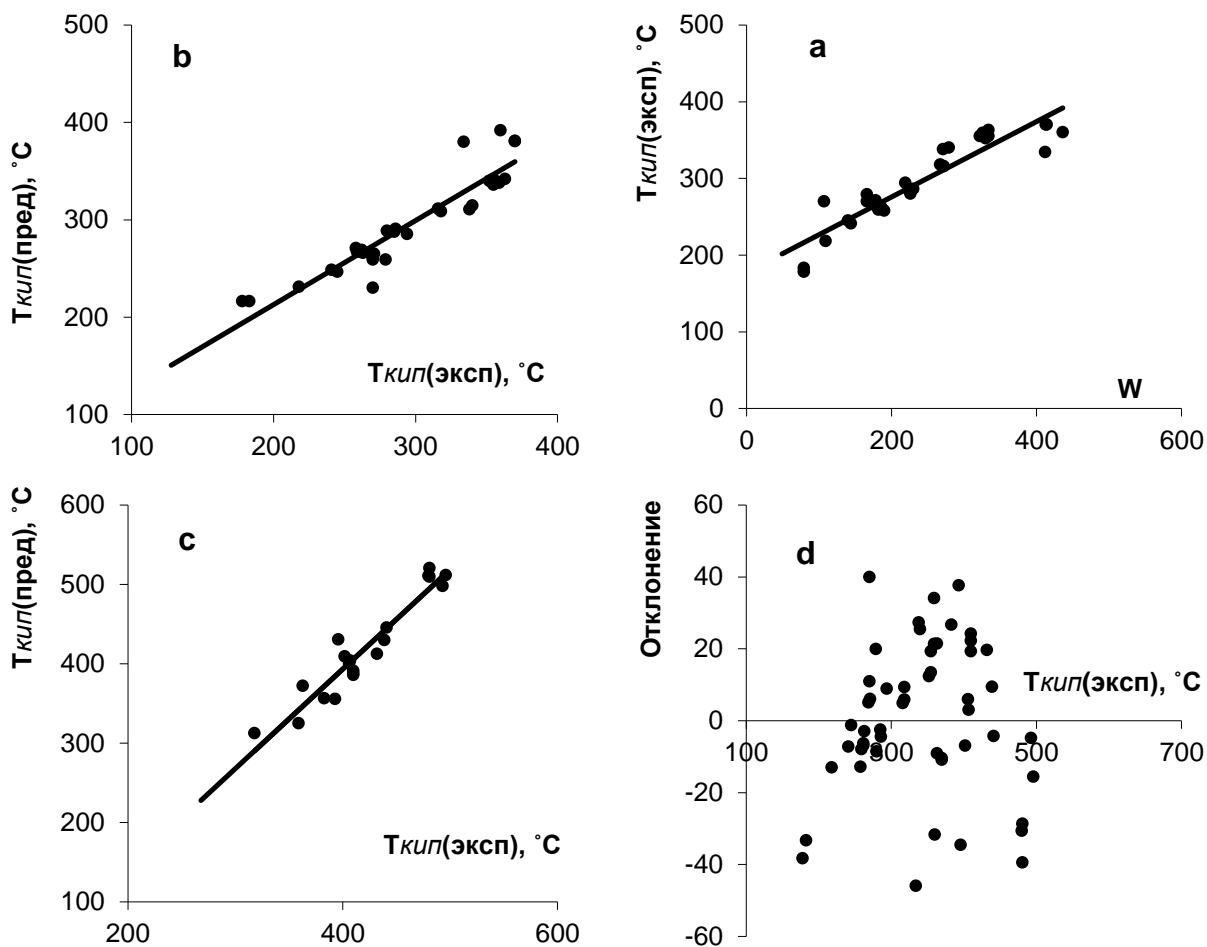


Рисунок 4.8.

QSPR – зависимость, построенная для температуры кипения ($T_{кип}$) полициклических ароматических углеводородов с учетом индекса Винера (W):

a) Соотношение между экспериментальной температурой кипения ПАУ ($T_{кип}$) и индексом Винера (W) для тренировочного набора; b) Соотношение между экспериментальной и предсказанной температурой кипения ($T_{кип}$) ПАУ для тренировочного набора; c) Соотношение между экспериментальной и предсказанной температурой кипения ($T_{кип}$) ПАУ для тестового набора; d) Анализ отклонений между экспериментальными и предсказанными значениями температуры кипения ($T_{кип}$) ПАУ.

Полученная модель «структура-свойство» имеет высокий коэффициент корреляции $r = 0,95$ и среднеквадратичную среднюю величину $s = 20,84^\circ\text{C}$. Но QSPR - уравнение с учетом индекса Рандича превосходит в этих показателях.

Проанализируем результат изменения величины инварианта графа на основе матрицы Лапласа $LEL(G)$ на температуру кипения ПАУ ($T_{кип}$). Построим зависимость в координатной плоскости на произведем поиск уравнения регрессии (рис. 4.9.).

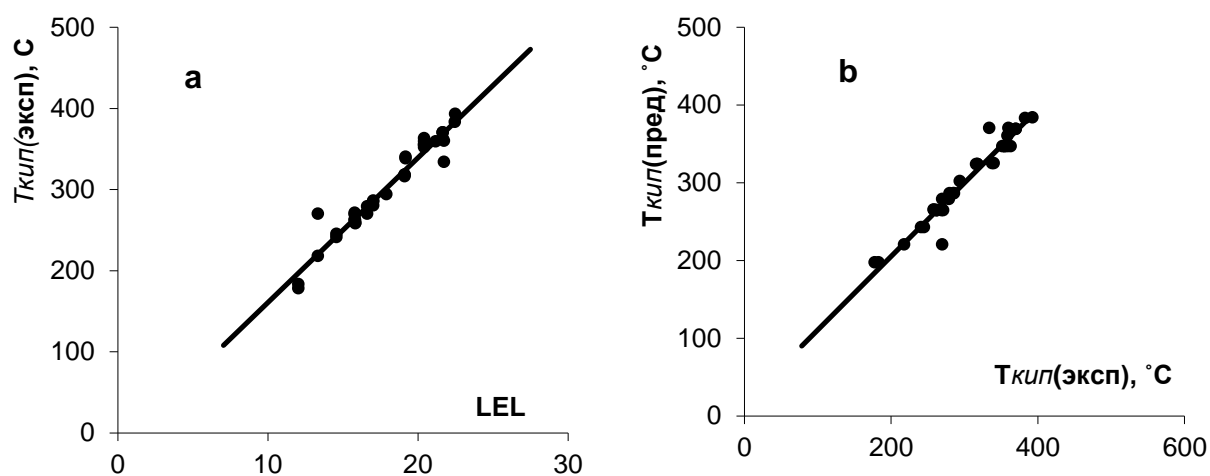


Рисунок 4.9.

QSPR – зависимость, построенная для температуры кипения ($T_{кип}$) полициклических ароматических углеводородов с учетом величины инварианта графа на основе матрицы Лапласа $LEL(G)$: а) Соотношение между экспериментальной температурой плавления ПАУ ($T_{кип}$) и величиной инварианта графа на основе матрицы Лапласа $LEL(G)$ для тренировочного набора; б) Соотношение между экспериментальной и предсказанной температурой плавления ($T_{кип}$) ПАУ для тренировочного набора;

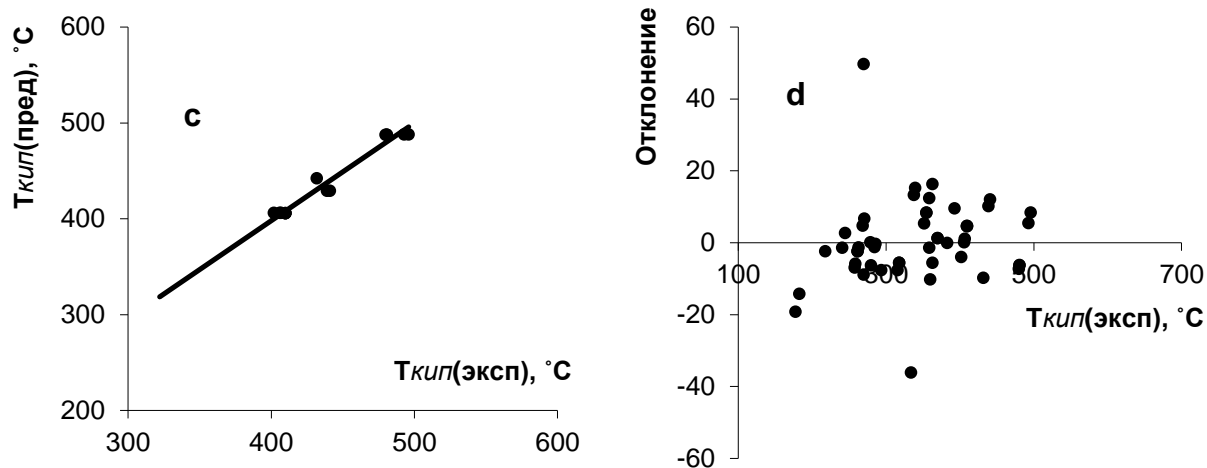


Рисунок 4.9.

QSPR – зависимость, построенная для температуры кипения ($T_{кип}$) полициклических ароматических углеводородов с учетом величины инварианта графа на основе матрицы Лапласа $LEL(G)$: c) Соотношение между экспериментальной и предсказанной температурой плавления ($T_{кип}$) ПАУ для тестового набора; d) Анализ отклонений между экспериментальными и предсказанными значениями температуры плавления ($T_{кип}$) ПАУ.

(продолжение)

Полученная модель «структура-свойство» со статистическими характеристиками $r = 0,98$ и $s = 11,58^\circ\text{C}$ немного улучшила качество рассчитываемых величин температуры кипения ПАУ по сравнению с моделью на основе индекса Рандича.

4.5. ПОСТРОЕНИЕ QSPR МОДЕЛЕЙ ДЛЯ РАСЧЕТА ТЕМПЕРАТУРЫ КИПЕНИЯ ПАУ С ПРИМЕНЕНИЕМ ИНДЕКСА СРЕДНЕГО РАССТОЯНИЯ.

Произведём анализ двухмерного графика для поиска зависимости (рис. 4.10.).

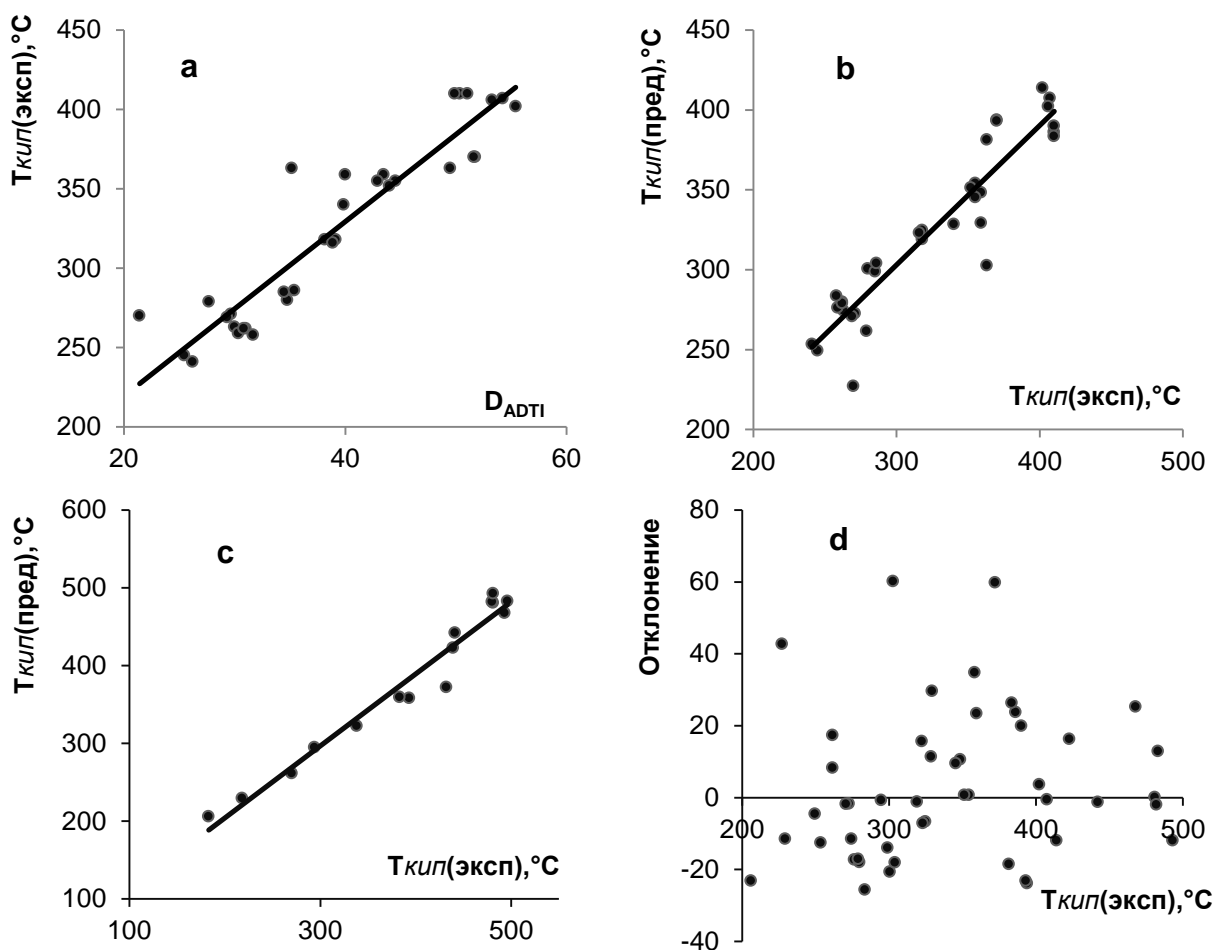


Рисунок 4.10.

QSPR – зависимость, построенная для температуры кипения ($T_{кип}$) полициклических ароматических углеводородов с учетом индекса среднего расстояния D_{ADTI} : а) Соотношение между экспериментальной температурой кипения ПАУ ($T_{кип}$) и величиной индекса среднего расстояния D_{ADTI} для обучающего набора; б) Соотношение между экспериментальной и предсказанной температурой кипения ($T_{кип}$) ПАУ для обучающего набора; в) Соотношение между экспериментальной и предсказанной температурой кипения ($T_{кип}$) ПАУ для контрольного набора ; д) Анализ отклонений между экспериментальными и предсказанными значениями температуры плавления ($T_{кип}$) ПАУ.

Разработанная количественная модель “структура-свойство” для определения температуры кипения полициклических ароматических углеводородов с применением индекса среднего расстояния D_{ADPI} имеет относительно высокий коэффициент корреляции $r = 0,93$, но среднеквадратичное стандартное отклонение высоко $s = 20,5^{\circ}\text{C}$. Это говорит о вероятности ошибочных расчетов при использовании данной модели.

4.6. ПОСТРОЕНИЕ QSPR - МОДЕЛЕЙ ДЛЯ РАСЧЕТА КОЭФФИЦИЕНТА РАСПРЕДЕЛЕНИЯ ПАУ В СИСТЕМЕ ОКТАНОЛ-ВОДА С УЧЕТОМ ТОПОЛОГИЧЕСКИХ ИНДЕКСОВ.

Проанализируем корреляцию между коэффициентом распределения в системе октанол – вода (LogP) и индексом Рандича (χ) (рис. 4.11.).

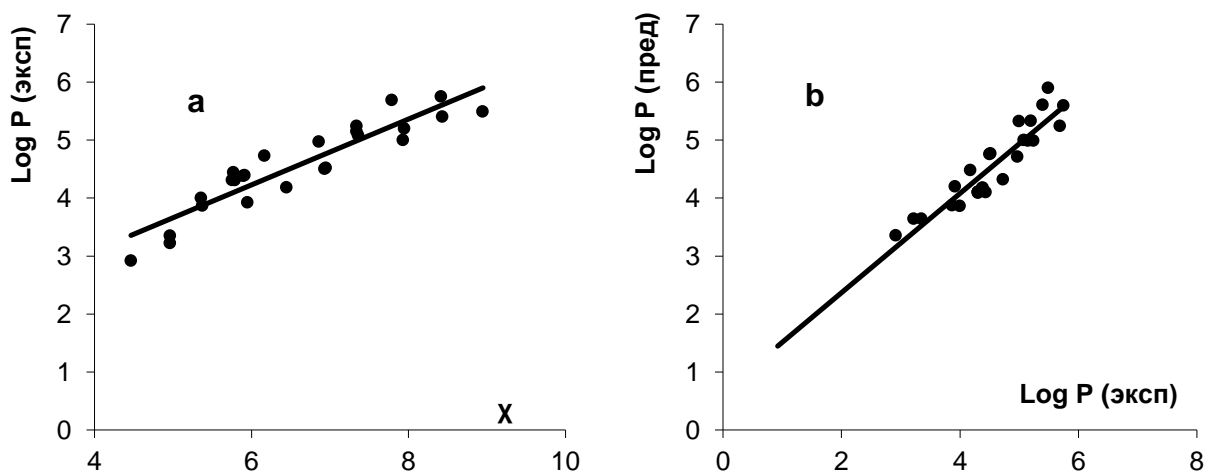


Рисунок 4.11.

QSPR – зависимость, построенная для коэффициента распределения (LogP) полициклических ароматических углеводородов с учетом индекса Рандича (χ):

- а) Соотношение между экспериментальной коэффициентом распределения ПАУ и индексом Рандича (χ) для тренировочного набора;
- б) Соотношение между экспериментальным и предсказанным коэффициентом распределения ПАУ для тренировочного набора.

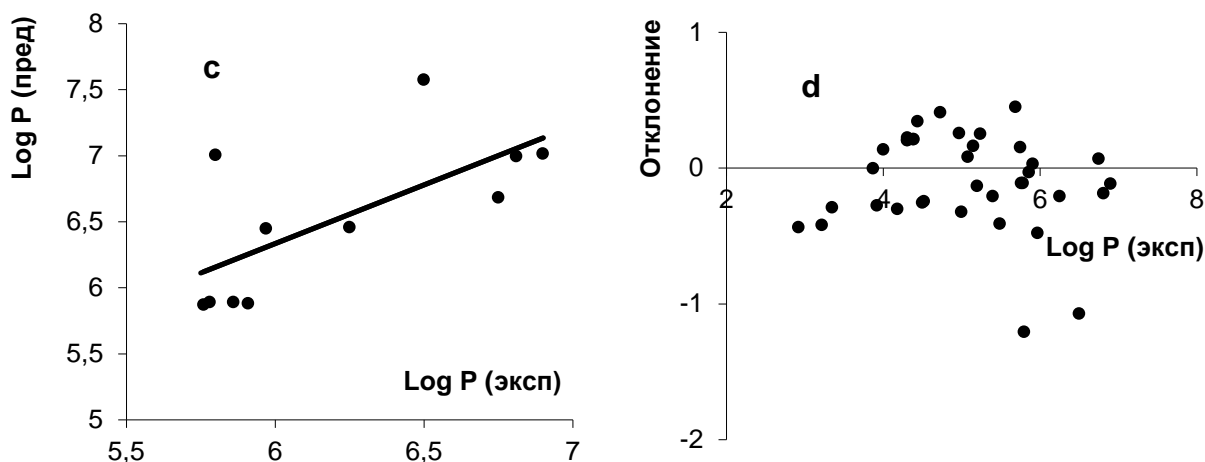


Рисунок 4.11.

QSPR – зависимость, построенная для коэффициента распределения (LogP) полициклических ароматических углеводородов с учетом индекса Рандича (χ):

- с) Соотношение между экспериментальным и предсказанным коэффициентом распределения ПАУ для тестового набора; d) Анализ отклонений между экспериментальными и предсказанными значениями коэффициента распределения (LogP) ПАУ.

Построенная модель “структура-свойство” имеет статистические показатели $r = 0,92$ и стандартное отклонение $s = 0,36$, которые указывают на высокую прогностическую способность. Но на графике отклонений вырисовывается увеличение негативных остатков.

Проверим качество логарифмической зависимости. Для этого построим двумерный график для логарифма величины индекса Рандича и коэффициента распределения полициклических ароматических углеводородов в системе октанол - вода (рис. 4.12.).

График отклонений имеет менее направленный характер. Логарифмическая модель улучшила статистические показатели $r = 0,94$ и $s = 0,27$.

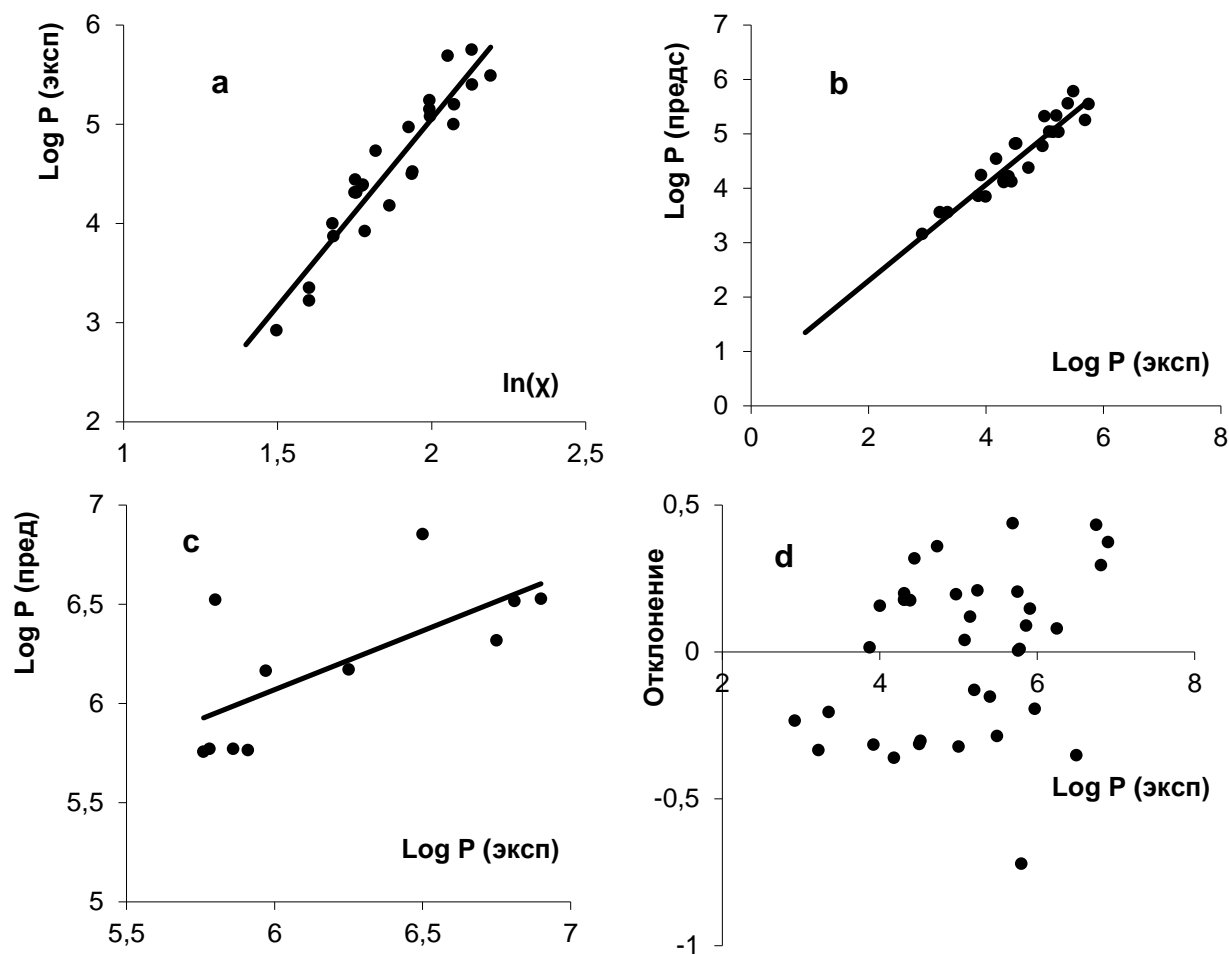


Рисунок 4.12.

QSPR – зависимость, построенная для коэффициента распределения ($\text{Log}P$) полициклических ароматических углеводородов с учетом логарифма индекса Рандича ($\ln(\chi)$): а) Соотношение между экспериментальной коэффициентом распределения ПАУ и логарифмом индекса Рандича ($\ln(\chi)$) для тренировочного набора; б) Соотношение между экспериментальным и предсказанным коэффициентом распределения ПАУ для тренировочного набора; в) Соотношение между экспериментальным и предсказанным коэффициентом распределения ПАУ для тестового набора; г) Анализ отклонений между экспериментальными и предсказанными значениями коэффициента распределения ($\text{Log}P$) ПАУ.

Протестируем зависимость между коэффициентом распределения ($\text{Log}P$) и индексом Винера (W) (рис. 4.13.).

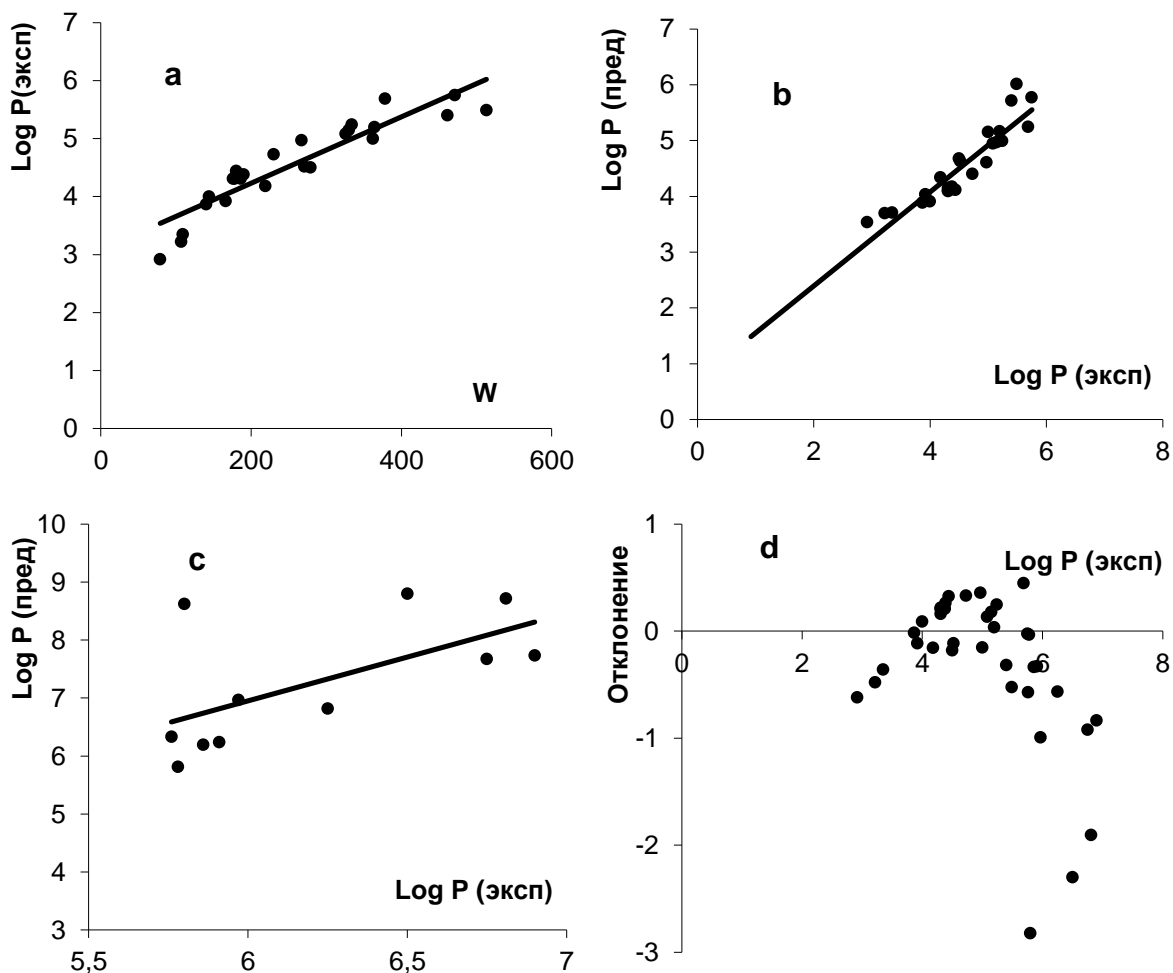


Рисунок 4.13.

QSPR – зависимость коэффициента распределения полициклических ароматических углеводородов ($\text{Log}P$) от величины индекса Винера (W): а)

Соотношение между экспериментальной величиной коэффициента распределения ПАУ и индексом Винера (W) для тренировочного набора; б) Соотношение между экспериментальным и предсказанным коэффициентом распределения ПАУ для тренировочного набора; в) Соотношение между экспериментальным и предсказанным коэффициентом распределения ПАУ для тестового набора; г) Анализ отклонений между экспериментальными и предсказанными значениями коэффициента распределения ($\text{Log}P$) ПАУ.

Данная модель “структура-свойство” имеет следующие статистические характеристики $r = 0,92$ и $s = 0,72$ и немного уступает уравнению с учетом индекса Рандича, так как рассчитанные величины больше расходятся от экспериментальных данных. На графике отклонений вырисовывается убывающая тенденция, что указывает на необходимость поиска нелинейной зависимости между переменными.

Проанализируем логарифмическую функциональную зависимость (рис. 4.14.).

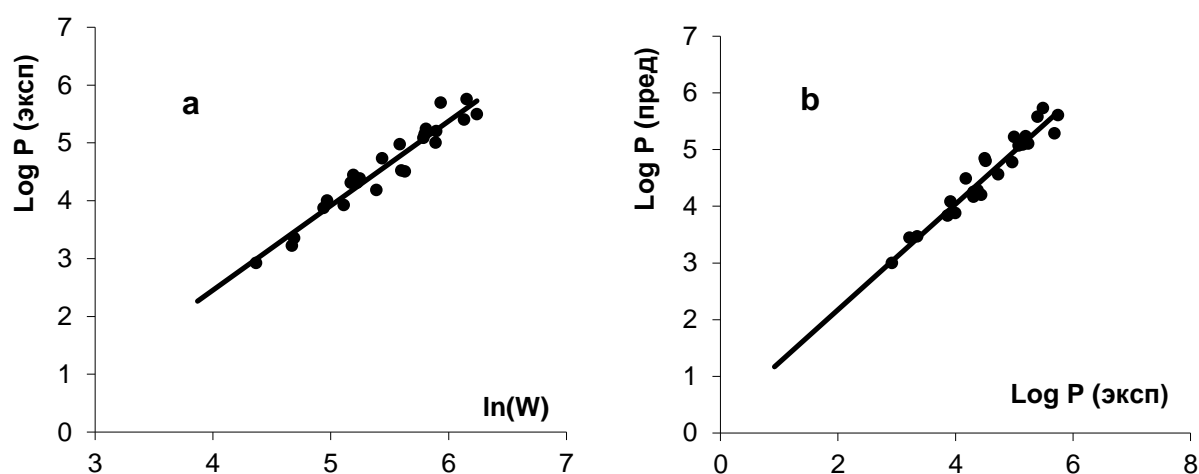


Рисунок 4.14.

QSPR – зависимость, построенная для коэффициента распределения полициклических ароматических углеводородов (LogP) с учетом логарифма индекса Винера (lnW): а) Соотношение между экспериментальным коэффициентом распределения ПАУ и логарифмом индекса Винера (lnW) для тренировочного набора; б) Соотношение между экспериментальным и предсказанным коэффициентом распределения ПАУ для тренировочного набора.

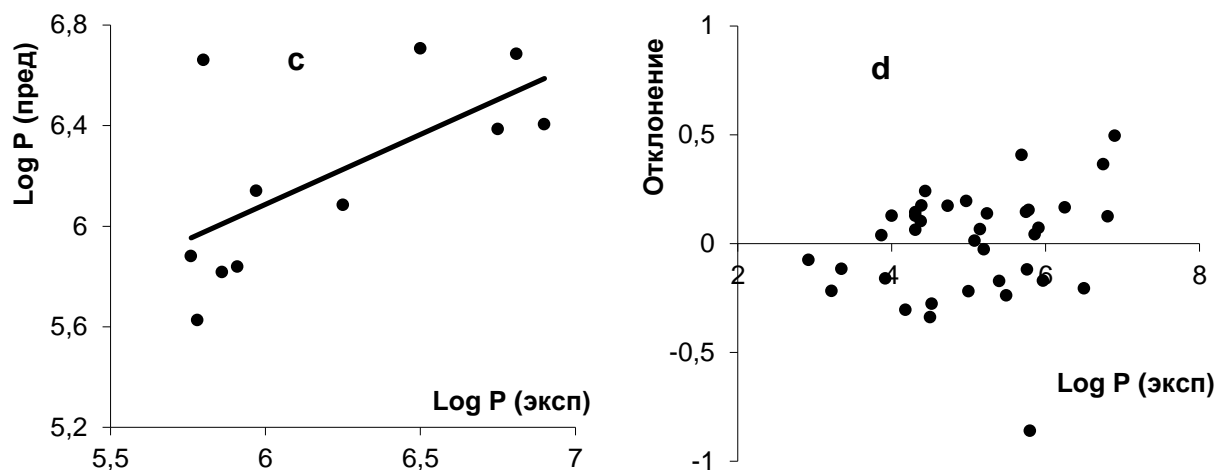


Рисунок 4.14.

QSPR – зависимость, построенная для коэффициента распределения полициклических ароматических углеводородов (LogP) с учетом логарифма индекса Винера (lnW): c) Соотношение между экспериментальным и предсказанным коэффициентом распределения ПАУ для тестового набора; d) Анализ отклонений между экспериментальными и предсказанными значениями коэффициента распределения (LogP) ПАУ. (продолжение)

Полученная модель “структура-свойство” улучшила статистические параметры $r = 0,96$ и $s = 0,25$ и предсказанные величины коэффициента распределения более близки к экспериментальным значениям.

Построим модель “структура-свойство”, представляющую зависимость между коэффициентом распределения полициклических ароматических углеводородов (LogP) с учетом величины инварианта графа на основе матрицы Лапласа $LEL(G)$ (рис. 4.15.).

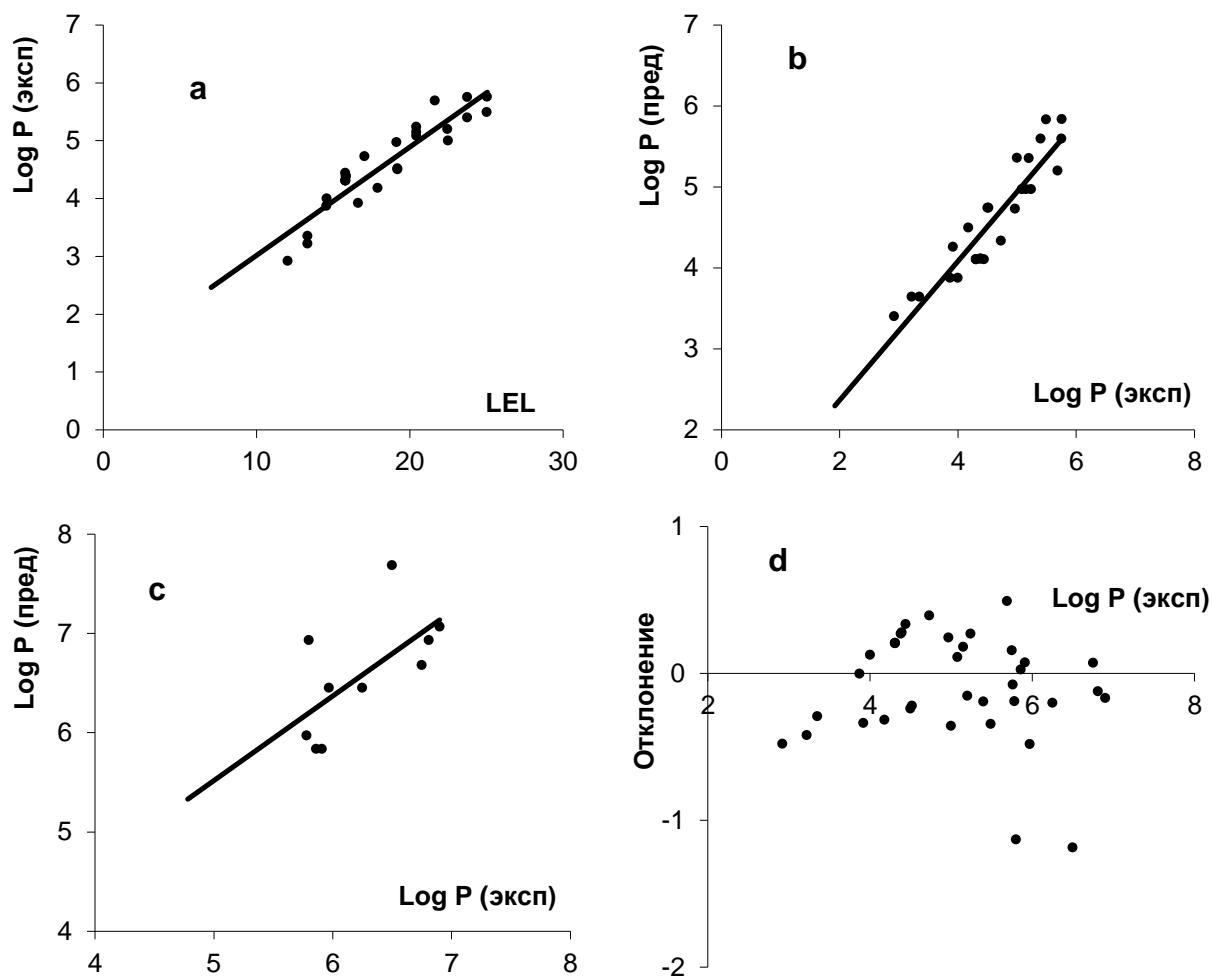


Рисунок. 4.15

QSPR – зависимость, построенная для коэффициента распределения полициклических ароматических углеводородов (LogP) с учетом величины инварианта графа на основе матрицы Лапласа (LEL): а) Соотношение между экспериментальной величиной коэффициента распределения ПАУ и индексом LEL для тренировочного набора; б) Соотношение между экспериментальным и предсказанным коэффициентом распределения ПАУ для тренировочного набора; в) Соотношение между экспериментальным и предсказанным коэффициентом распределения ПАУ для тестового набора; д) Анализ отклонений между экспериментальными и предсказанными значениями коэффициента распределения (LogP)ПАУ.

Построенная модель «структура-свойство» имеет следующие статистические параметры $r = 0,93$ и $s = 0,36$ и не является наиболее перспективной по предсказательной возможности.

Построим модель «структура-свойство» отражающую зависимость между коэффициентом распределения ($\text{Log}P$) и логарифмом $\ln(\text{LEL})$ (рис. 4.16.).

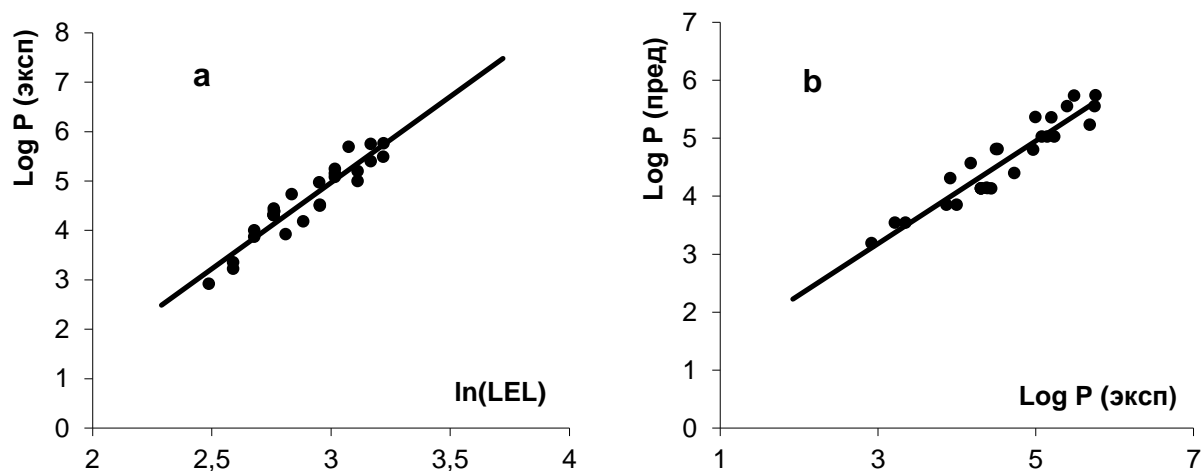


Рисунок 4.16.

QSPR – зависимость, построенная для коэффициента распределения полициклических ароматических углеводородов ($\text{Log}P$) с учетом логарифма величины инварианта графа на основе матрицы Лапласа ($\ln(\text{LEL})$): а)

Соотношение между экспериментальным коэффициентом распределения ПАУ и логарифмом величины инварианта графа ($\ln(\text{LEL})$) для тренировочного набора; б) Соотношение между экспериментальным и предсказанным коэффициентом распределения ПАУ для тренировочного набора.

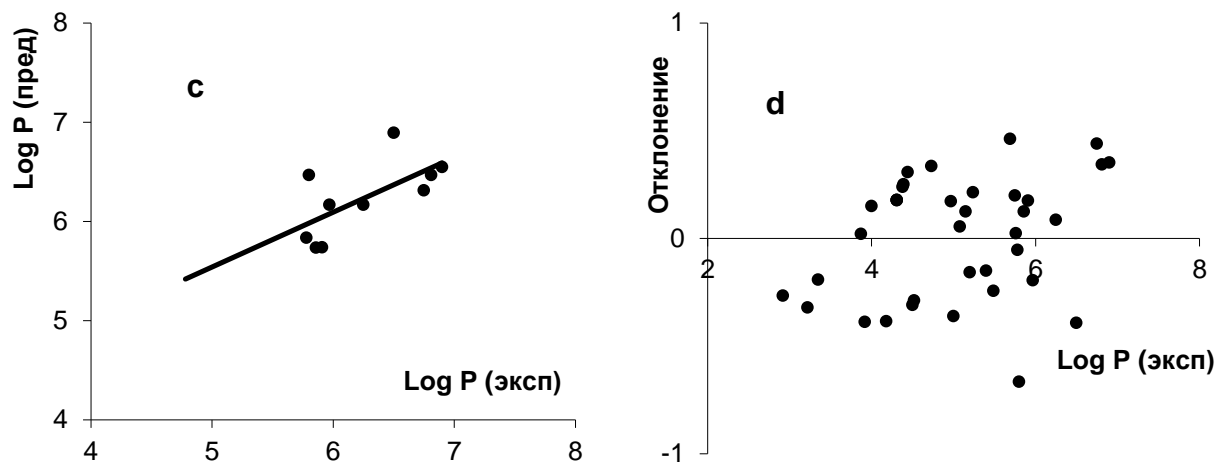


Рисунок 4.16.

QSPR – зависимость, построенная для коэффициента распределения полициклических ароматических углеводородов (LogP) с учетом логарифма величины инварианта графа на основе матрицы Лапласа ($\ln(\text{LEL})$): c) Соотношение между экспериментальным и предсказанным коэффициентом распределения ПАУ для тестового набора; d) Анализ отклонений между экспериментальными и предсказанными значениями коэффициента распределения (LogP) ПАУ. (продолжение)

Полученная модель “структура-свойство” имеет следующие статистические параметры $r = 0,93$ и $s = 0,28$, которые указывают на более качественные расчетные данные с учетом логарифмической функции. Но они не превысили показатели для расчета коэффициента распределения ПАУ модели с использованием логарифма индекса Винера.

4.7. ПОСТРОЕНИЕ QSPR МОДЕЛЕЙ ДЛЯ РАСЧЕТА КОЭФФИЦИЕНТА РАСПРЕДЕЛЕНИЯ ПАУ В СИСТЕМЕ ОКТАНОЛ-ВОДА С УЧЕТОМ ИНДЕКСА СРЕДНЕГО РАССТОЯНИЯ.

Построим и проанализируем двухмерный график взаимосвязи между коэффициентом распределения полициклических ароматических углеводородов в системе октанол-вода и логарифмом индекса среднего расстояния (рис. 4.17.).

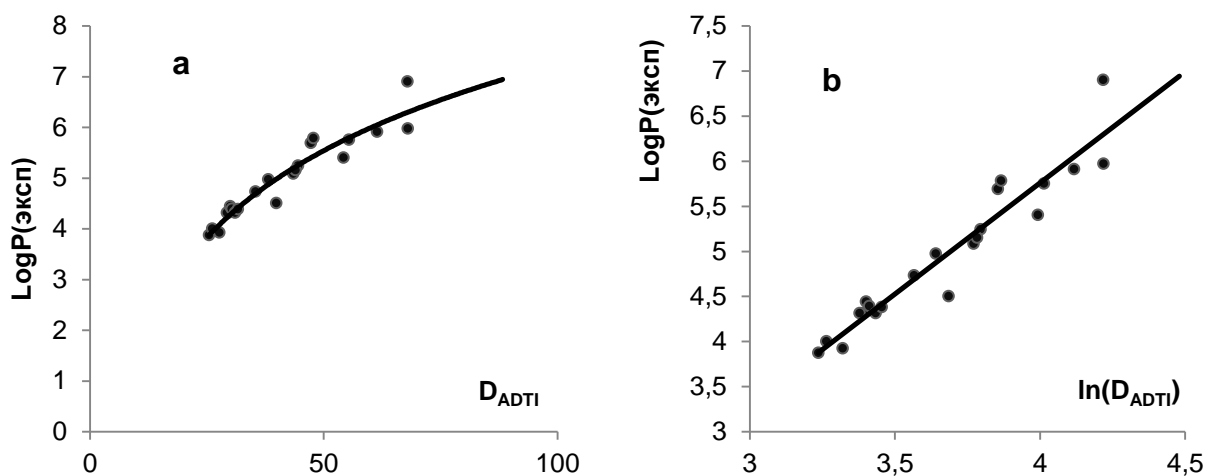


Рисунок 4.17.

QSPR – зависимость, построенная для коэффициента распределения полициклических ароматических углеводородов (LogP) с учетом логарифма индекса среднего расстояния ($\ln(D_{\text{ADTI}})$): а) Соотношение между экспериментальным коэффициентом распределения ПАУ и индексом среднего расстояния для тренировочного набора (D_{ADTI}), б) Соотношение между экспериментальным коэффициентом распределения ПАУ и логарифмом индекса среднего расстояния ($\ln(D_{\text{ADTI}})$) для тренировочного набора.

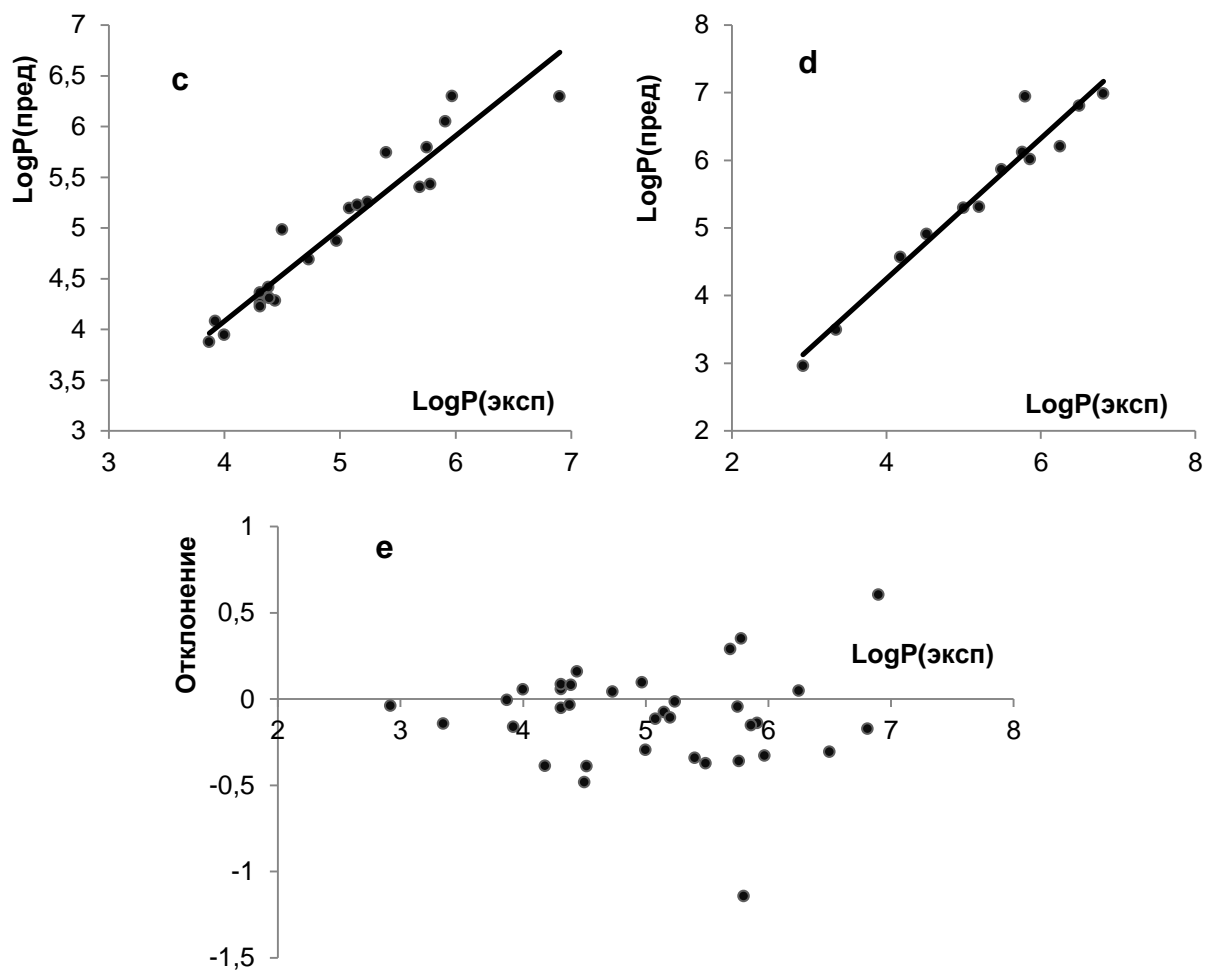


Рисунок 4.17.

QSPR – зависимость, построенная для коэффициента распределения полициклических ароматических углеводородов (LogP) с учетом логарифма индекса среднего расстояния ($\ln(D_{\text{ADTI}})$): c) Соотношение между экспериментальным и предсказанным коэффициентом распределения ПАУ для тренировочного набора; d) Соотношение между экспериментальным и предсказанным коэффициентом распределения ПАУ для тестового набора, e) Анализ отклонений между экспериментальными и предсказанными значениями коэффициента распределения (LogP) ПАУ. (продолжение)

Построенная QSPR – модель имеет следующие статистические характеристики стандартное отклонение $s = 0,287$ и коэффициент корреляции $r = 0,96$, которые ставят ее по предсказательной способности в аналогичную категорию с другими разработанными ранее моделями.

Таблица 4.7.1. QSPR-Уравнения полученные для полициклических ароматических углеводородов, статистические характеристики: R^2 - коэффициент детерминации, r - коэффициент корреляции, s - стандартное отклонение.

№	QSPR-модель	R^2	r	s
1	$T_{пл} = 38,628 \chi - 169,36$	0,6356	0,8	55,01°C
2	$T_{пл} = 0,2373 W + 31,523$	0,5925	0,77	8,96°C
3	$T_{пл} = 13,186LEL - 157,25$	0,6608	0,81	55,06°C
4	$T_{пл} = 81,083W^{0,1} \chi^{0,5} - 276,61$	0,6176	0,79	54,10°C
5	$T_{пл} = 3,308 D_{ADTI} - 32,48$	0,543	0,74	54,6 °C
6	$T_{пл} = -0,023 (D_{ADTI})^2 + 6,196D_{ADTI} - 108,7$	0,572	0,76	55,4°C
7	$T_{кип} = 51,662\chi - 33,539$	0,9255	0,96	12,86°C
8	$T_{кип} = 0,4914W + 177,5$	0,8636	0,93	20,84°C
9	$T_{кип} = 17,828LEL - 17,391$	0,9458	0,97	11,58°C
10	$T_{кип} = 5,489 D_{ADTI} + 109,8$	0,872	0,93	20,5°C
11	$LogP = 0,5674\chi + 0,8222$	0,8548	0,92	0,36
12	$LogP = 3,7736\ln(\chi) - 2,4927$	0,8828	0,94	0,27
13	$LogP = 0,0057W + 3,0885$	0,8457	0,92	0,72
14	$LogP = 1,461\ln(W) - 3,3881$	0,9307	0,96	0,25
15	$LogP = 0,1874LEL + 1,1431$	0,8591	0,93	0,36
16	$LogP = 3,4827\ln(LEL) - 5,48013$	0,8862	0,93	0,28
17	$LogP = 2,466\ln(D_{ADTI}) - 4,106$	0,914	0,96	0,287

4.8. ПОСТРОЕНИЕ QSPR - МОДЕЛЕЙ ДЛЯ ПРЕДСКАЗАНИЯ ЭНТАЛЬПИИ ОБРАЗОВАНИЯ ПАУ В ГАЗОВОЙ ФАЗЕ.

Энтальпия образования полициклических ароматических углеводородов изучена незначительно. Построим модели «структура - свойство» для предсказания энтальпии образования ПАУ с учетом топологических индексов (ТИ). Определим различные функциональные зависимости $\Delta H = f(ТИ)$, чтобы выбрать модель с высокой предсказательной способностью:

$$\Delta_f H^\circ_{(r,298K)} = a(ТИ) + b \quad (3)$$

$$\Delta_f H^\circ_{(r,298K)} = a(ТИ)^2 + b(ТИ) + c \quad (4)$$

$$\Delta_f H^\circ_{(r,298K)} = a(ТИ)^b \quad (5)$$

$$\Delta_f H^\circ_{(r,298K)} = a \ln(ТИ) + b \quad (6)$$

$$\Delta_f H^\circ_{(r,298K)} = a e^{(ТИ)} \quad (7)$$

$$\Delta_f H^\circ_{(r,298K)} = a(ТИ)_1^h b(ТИ)_2^k + c \quad (8)$$

$$\Delta_f H^\circ_{(r,298K)} = a(ТИ)_1 + b(ТИ)_2 + \dots + k(ТИ)_n + c \quad (9)$$

где a, b, c, h, k, n - некоторые величины, найденные при построении модели.

В связи с тем, что количество объектов для исследования небольшое, разбивку на обучающую и контрольную выборку производить не будем. Вся группа выступит как тренировочный набор.

Проанализируем двухмерный график на возможность линейной зависимости (рис. 4.18.).

Статистические показатели полученной количественной модели “структура-свойство” $r = 0,88$ и $s = 5,43$ ккал/моль, что указывает на вероятность качественной прогностической возможности.

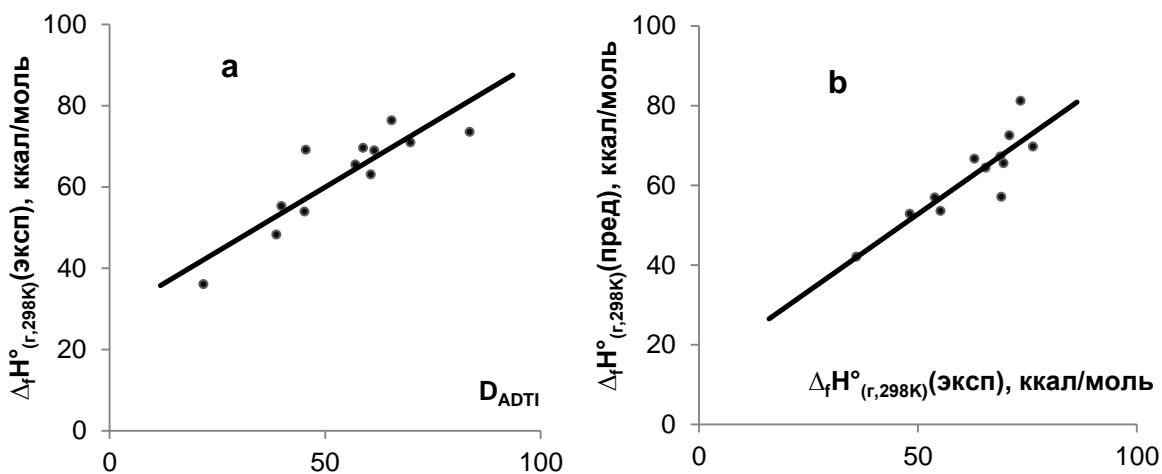


Рисунок 4.18.

QSPR – зависимость, построенная для энтальпии образования полициклических ароматических углеводородов с учетом индекса среднего расстояния: а) Соотношение между экспериментальными значениями энтальпии образования ПАУ для тренировочного набора; б) Соотношение между экспериментальными и предсказанными значениями энтальпии образования ПАУ для тренировочного набора.

Исследуем двухмерный график на предмет существования степенной функциональной зависимости. Для этого проведем трансформацию данных и изобразим зависимость между логарифмом индекса среднего расстояния и логарифма энтальпии образования полициклических ароматических углеводородов (рис. 4.19.).

После математических преобразований для выведения уравнения для расчета энтальпии образования ПАУ, получена степенная функциональная зависимость $\Delta_f H^\circ_{(r,298K)} (\text{эксп}) = 6,544(D_{ADTI})^{0,568}$. Данная количественная модель “структура-свойство” немного улучшила статистические характеристики $s = 4,99$ ккал/моль и $r = 0,93$, что позволяет предположить о повышении точности прогнозируемых результатов.

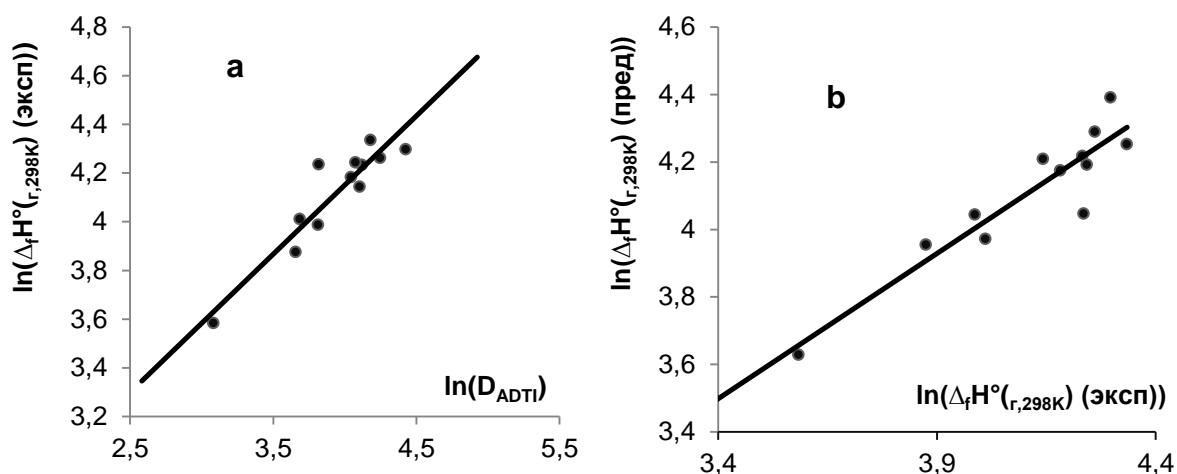


Рисунок 4.19.

QSPR – зависимость, построенная для энтальпии образования полициклических ароматических углеводородов с учетом индекса среднего расстояния используя логарифмическую трансформацию значений: а) Соотношение между экспериментальными значениями энтальпии образования ПАУ для тренировочного набора; б) Соотношение между логарифмом экспериментальных и логарифмом предсказанных значений энтальпии образования ПАУ для тренировочного набора..

Анализируя качественные показатели полученных моделей “структурасвойство” для расчета энтальпии образования ПАУ, напрашивается вывод о дискриминируемой ответной реакции точности величины энтальпии образования ПАУ на применяемый молекулярный дескриптор. Все полученные уравнения и статистические характеристики моделей перечислены в таблице 4.8.1. Сравним предсказательную способность моделей. Наилучшую прогностическую возможность показала многопараметровая модель на основании уравнения (9), построенная с учетом индекса Рандича, индекса Винера и инварианта графа на основании матрицы Лапласа. Это уравнение 17 в таблице 4.8.1. Результаты расчетов с применением уравнения 17 представлены в таблице 4.8.4. Корреляция

между экспериментальными и предсказанными значениями энтальпии образования полициклических ароматических углеводородов изображена на рисунке 4.20. Проиллюстрированы соотношения, задействовавшие уравнения 17 - 19 и методы Бенсона, Моисеевой и Армитажа. Графическое представление зависимости между экспериментальными величинами и предсказанными значениями энтальпии образования в газовой фазе посредством использования расчетных методов указывает на достаточно высокое качество. Сравнивая статистические характеристики обсуждаемых методов, можно упорядочить их в соответствии с точностью предсказательной возможности. Но можно также рассматривать погрешность вычислений, анализируя структуру молекулы. Дескрипторы не всегда в состоянии реагировать на индивидуальные особенности представителей класса объектов, когда они выпадают из общей тенденции. Существует мнение, что экспериментальное значение энтальпии образования в газовой фазе бензо[к]флюорантена содержит ошибку. [66] Исключение этой величины из анализа данных несколько меняет расстановку рассматриваемых методов и моделей по их предсказательной способности в следующем порядке: Армитаж, Бенсон, QSPR-модель 17, Моисеева, QSPR-модель 19, QSPR-модель 1.

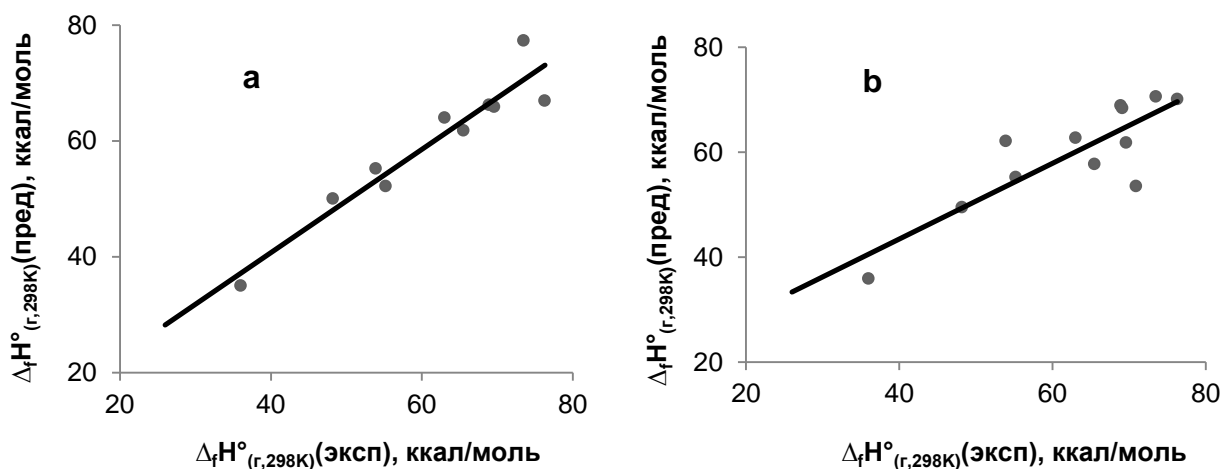


Рисунок 4.20. Соотношение экспериментальных и предсказанных значений энтальпии образования полициклических ароматических углеводородов в газовом состоянии, рассчитанных по а) методу Бенсона б) методу Моисеевой.

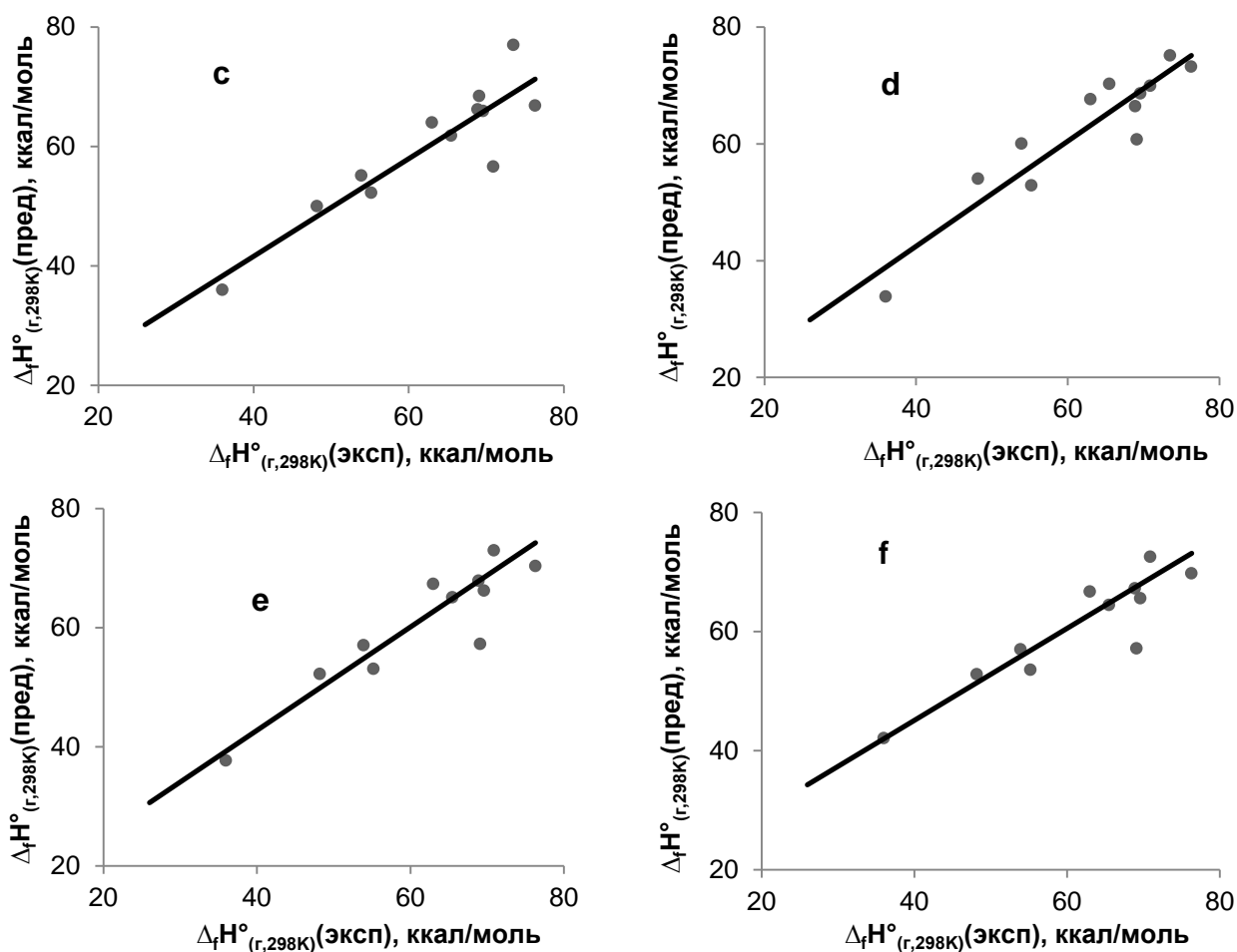


Рисунок 4.20. Соотношение экспериментальных и предсказанных значений энтальпии образования полициклических ароматических углеводородов в газовом состоянии, рассчитанных по с) методу Армитажа d) QSPR-модели 17 e) QSPR-модели 18 f) QSPR-модели 19. (продолжение)

Таблица 4.8.1. Уравнения «структура - свойство» для расчета энтальпии образования полициклических ароматических углеводородов и их статистические характеристики коэффициент корреляции r , стандартное отклонение s (ккал/моль).

№	QSPR-модель	s	r
1	$\Delta_f H^\circ_{(г,298К)} = 6,0786 \chi + 10,465$	6,07	0,88
2	$\Delta_f H^\circ_{(г,298К)} = 23.524e^{0.1114 \chi}$	7,49	0,88
3	$\Delta_f H^\circ_{(г,298К)} = 49.639 \ln(\chi) - 42.886$	5.25	0,91

Таблица 4.8.1.

№	QSPR-модель	s	r
4	$\Delta_f H^\circ_{(г,298К)} = -0.8022 \chi^2 + 19.535 \chi - 43.386$	4,74	0,93
5	$\Delta_f H^\circ_{(г,298К)} = 8.6332 \chi^{0.9211}$	6,11	0,92
6	$\Delta_f H^\circ_{(г,298К)} = 0.0422W + 41.456$	7,40	0,82
7	$\Delta_f H^\circ_{(г,298К)} = 41.739e^{0.0008W}$	7,32	0,81
8	$\Delta_f H^\circ_{(г,298К)} = 19.094 \ln(W) - 53.364$	5,20	0,91
9	$\Delta_f H^\circ_{(г,298К)} = -0,00007W^2 + 0.1225W + 23.607$	4,78	0,93
10	$\Delta_f H^\circ_{(г,298К)} = 7.063W^{0.3554}$	5,82	0,93
11	$\Delta_f H^\circ_{(г,298К)} = 1.948LEL(G) + 15.379$	6,33	0,87
12	$\Delta_f H^\circ_{(г,298К)} = 25.769e^{0.0356LEL(G)}$	7,82	0,87
13	$\Delta_f H^\circ_{(г,298К)} = 45,328 \ln(LEL(G)) - 80,622$	5,35	0,91
14	$\Delta_f H^\circ_{(г,298К)} = -0,0911LEL(G)^2 + 6,3253LEL(G) - 34,435$	4,73	0,93
15	$\Delta_f H^\circ_{(г,298К)} = 4,2855LEL(G)^{0.8412}$	6,27	0,92
16	$\Delta_f H^\circ_{(г,298К)} = 9,8329 \chi^{-0,8} W^{0,6} - 7,0822$	6,10	0,90
17	$\Delta_f H^\circ_{(г,298К)} = 40,9265R - 0,0605W - 8,7381LEL(G) - 46,2531$	4,23	0,93
18	$\Delta_f H^\circ_{(г,298К)} (\text{эксп}) = 0,634D_{ADTI} + 28,27$	5,43	0,88
19	$\Delta_f H^\circ_{(г,298К)} (\text{эксп}) = 6,544(D_{ADTI})^{0,568}$	4,99	0,93

Таблица 4.8.2. Источники используемых экспериментальных данных для энтальпии образования $\Delta_f H^\circ_{(г,298К)}$ ПАУ.

Соединение	Источник	Соединение	Источник
Бензол	30	Кризен	30
Нафталин	30	Трифенилен	30
Фенантрен	30	Бенз[а]антрацен	30
Пирен	176	Бенз[с]фенантрен	30
Антрацен	30	Перилен	72

Таблица 4.8.3. Статистические характеристики моделей, разработанных для расчета энтальпии образования полициклических ароматических углеводородов в газовой фазе при 298К.

Модель/Автор метода	R	s, ккал/моль
Бенсон	0,95	3,59
QSPR-модель 17	0,93	4,23
Армитаж	0,90	4,84
QSPR-модель 19	0,90	4,99
QSPR-модель 18	0,88	5,43
Моисеева	0,84	6,08

Таблица 4.8.4. Значения энтальпии образования полициклических ароматических углеводородов (ккал/моль) рассчитанные по QSPR-модели 17.

Название соединения	$\Delta_f H^\circ_{(г,298К)}$ (эксп)	$\Delta_f H^\circ_{(г,298К)}$ (пред)
Нафталин	35,9	33,8
1-метилнафталин		38,0
2-метилнафталин		37,0
1-этилнафталин		46,4
2-этилнафталин		45,2
2-6-диметилнафталин		39,9
2-7-диметилнафталин		39,9
1-7-диметилнафталин		40,9
1-5-диметилнафталин		41,9
1-2-диметилнафталин		41,8
1-3-7-триметилнафталин		43,5
2-3-5-триметилнафталин		44,3
2-3-6-триметилнафталин		43,3

Таблица 4.8.4

Название соединения	$\Delta_f H^\circ_{(г,298К)}$ (эксп)	$\Delta_f H^\circ_{(г,298К)}$ (пред)
Фенален		48,4
1-фенилнафталин		64,2
2-фенилнафталин		62,0
Антрацен		52,9
1-метилантрацен		55,6
2-метилантрацен		54,4
2-7-диметилантрацен		55,4
2-6-диметилантрацен		55,4
2-3-диметилантрацен		56,5
9-10-диметилантрацен		60,5
Фенантрен	49,6	54,0
1-метилфенантрен		56,8
2-метилфенантрен		55,6
3-метилфенантрен		55,8
4-метилфенантрен		57,0
9-метилфенантрен		57,0
3-6-диметилфенантрен		57,2
4-5-метилефенантрен		54,6
Тетрацен		64,7
Бенз[а]антрацен	70,0	66,4
Кризен	64,5	67,6
Бенз[с]фенантрен	69,6	68,6
Трифенилен		70,2
Пирен	53,9	60,0
1-метилпирен		62,1

Таблица 4.8.4

Название соединения	$\Delta_f H^\circ_{(г,298К)}$ (эксп)	$\Delta_f H^\circ_{(г,298К)}$ (пред)
2-метилпирен		61,0
4-метилпирен		62,3
2-7-диметилпирен		61,5
Пентацен		67,3
Дибенз[ai]антрацен		69,5
Дибенз[ah]антрацен		71,2
Дибенз[aj]антрацен		72,2
Бенз[b]кризен		71,2
Дибенз[ac]антрацен		75,8
Пицен		72,4
Бенз[a]пирен		70,9
Бенз[e]пирен		73,3
Перилен	76,3	73,2
Коронен	73,5	75,1
Антрактрен		72,7
бенз[ghi]перилен		74,9
дибенз[ae]пирен		77,4
1-метилкризен		69,1
6-метилкризен		68,4
3-метилколантрен		69,6
indeno[1-2-3-cd]пирен		73,3
Пентафен		70,0
Гексафен		62,5
Индан		26,5
Инден		26,5

Таблица 4.8.4

Название соединения	$\Delta_f H^\circ_{(г,298К)}$ (эксп)	$\Delta_f H^\circ_{(г,298К)}$ (пред)
Азулен		34,0
Аценафтен		41,9
Аценафтилен		41,9
Флуорен		48,0
1-метилфлуорен		51,2
2-метилфлуорен		50,1
3-метилфлуорен		50,2
4-метилфлуорен		51,3
9-метилфлуорен		52,2
1-2-бензфлуорен		63,5
Флуорантен	69,1	60,7
2-3-бензфлуорен		62,2
3-4-бензфлуорен		64,0
бенз[ghi]флуорантен		65,3
бенз[k]флуорантен	70,9	69,9
бенз[b]флуорантен		72,0
бенз[j]флуорантен		71,9
Овален		62,3
Кватерилен		-17,1

4.9. ПОСТРОЕНИЕ QSPR-МОДЕЛЕЙ ДЛЯ ОПРЕДЕЛЕНИЯ ЭНТАЛЬПИИ ОБРАЗОВАНИЯ ПАУ, СОСТОЯЩИХ ИЗ БЕНЗОЛЬНЫХ КОЛЕЦ.

Для построения QSPR-модели выберем объекты, структура которых включает только ароматические кольца. Расположим граф молекулы таким образом, чтобы вертикальная ось симметрии каждого бензольного кольца прошла через две противоположные вершины. Если граф содержит несколько рядов ароматических колец с общим ребром, то цепь с наименьшим количеством колец должна находиться на самом высоком уровне.

Например: нафто[2,1-а]пирен. Разные варианты ориентации графа этой молекулы изображены на рисунке 4.21.

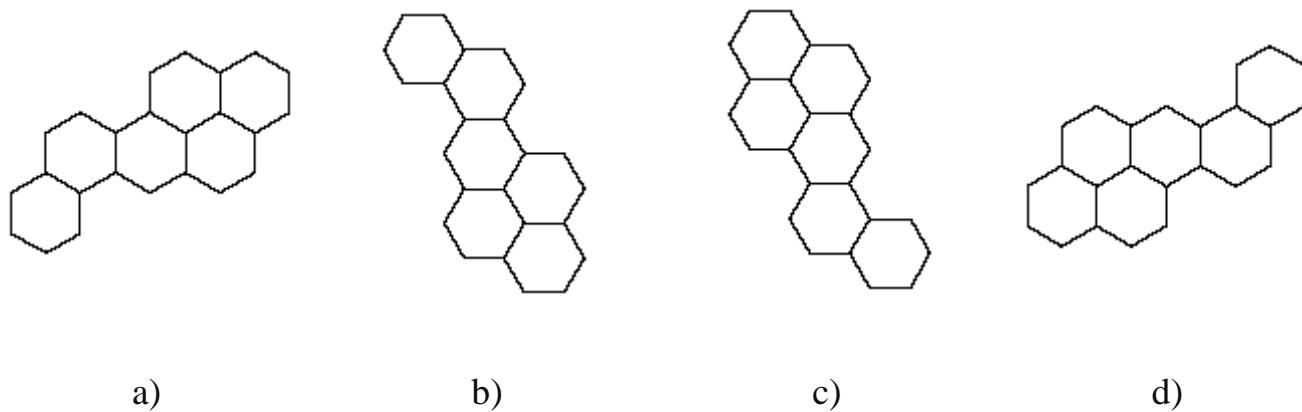


Рисунок 4.21.

Ориентация графа молекулы нафто[2,1-а]пирена.

Корректной версией является положение d). Далее, нужно перечислить количество колец в каждом ряду, начиная с самого нижнего в возрастающем порядке.

Таблица 4.9.1. Возможные конфигурации для шестичленных колец изомеров полициклических ароматических углеводородов. [167]

Молекулярный вес	Химическая Формула	Количество колец	Внутренние атомы углерода	Количество теоретических изомеров	Количество синтезированных изомеров	Количество перечисленных изомеров
78	C_6H_6	1	0	1	1	1
128	$C_{10}H_8$	2	0	1	1	1
178	$C_{14}H_{10}$	3	0	2	2	2
202	$C_{16}H_{10}$	4	2	1	1	1
228	$C_{18}H_{12}$	4	0	5	5	5
252	$C_{20}H_{12}$	5	2	3	3	3
276	$C_{22}H_{12}$	6	4	2	2	2
278	$C_{22}H_{14}$	5	0	12	12	12
300	$C_{24}H_{12}$	7	6	1	1	1
302	$C_{24}H_{14}$	6	2	13	13	13
326	$C_{26}H_{14}$	7	4	9	5	9
328	$C_{26}H_{16}$	6	0	37	27	37
350	$C_{28}H_{14}$	8	6	8	5	8
352	$C_{28}H_{16}$	7	2	62	24	45
374	$C_{30}H_{14}$	9	8	3	2	3
376	$C_{30}H_{16}$	8	4	58	11	50

Таблица 4.9.1.

Молекулярный вес	Химическая Формула	Количество колец	Внутренние атомы углерода	Количество теоретических изомеров	Количество синтезированных изомеров	Количество перечисленных изомеров
378	C ₃₀ H ₁₈	7	0	>123	23	93
398	C ₃₂ H ₁₄	10	10	1	1	1
400	C ₃₂ H ₁₆	9	6	46	7	40
402	C ₃₂ H ₁₈	8	2	>289	16	8
424	C ₃₄ H ₁₆	10	8	37	2	16
426	C ₃₄ H ₁₈	9	4	много	8	14
428	C ₃₄ H ₂₀	8	0	>411	8	11
448	C ₃₆ H ₁₆	11	10	20	2	20
450	C ₃₆ H ₁₈	10	6	много		9
452	C ₃₆ H ₂₀	9	2	много		13
478	C ₃₈ H ₂₂	9	0	>1489	6	5

Предсказательная возможность модели распространяется на молекулы с количеством горизонтальных цепей в пределах от одной до трех. С использованием множественного линейно-регрессионного анализа вывели следующее уравнение:

$$\Delta_f H^\circ_{(г, 298K)} (\text{пред}) = 1,28 + 0,21D_{ADTI} + 14,64N_1 + 9,24N_2 + 17,91N_3 \quad (10),$$

где D_{ADTI} – топологический индекс среднего расстояния, N_1 , N_2 и N_3 - количество ароматических колец в первом, втором и третьем рядах графа соответственно.

Таблица 4.9.2. Исследуемые объекты для проверки прогностической способности QSPR-модели $\Delta_f H^\circ_{(г,298К)} = 1,28 + 0,21D_{ADTI} + 14,64N_1 + 9,24N_2 + 17,91N_3$ и значения индекса среднего расстояния D_{ADTI} .


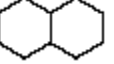
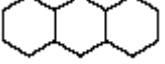
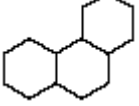
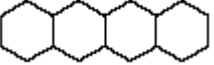
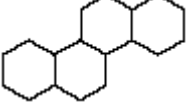
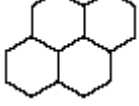
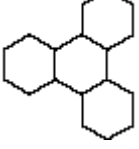
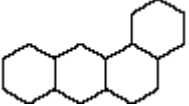
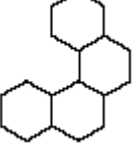
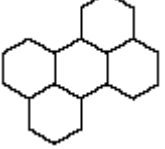
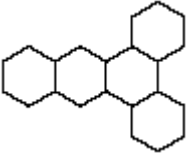
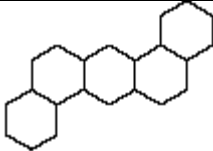
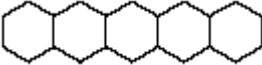
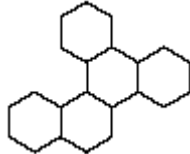
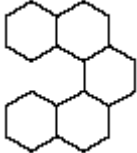

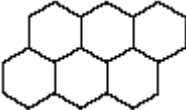
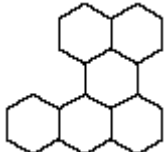
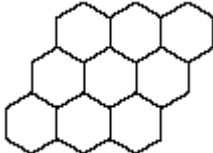
№	Название соединения	Формула	Структура	D_{ADTI}
1	Бензин	C_6H_6		9
2	Нафталин	$C_{10}H_8$		21,8
3	Антрацен	$C_{14}H_{10}$		39,9
4	Фенантрен	$C_{14}H_{10}$		38,7
5	Нафтацен	$C_{18}H_{12}$		63,2
6	Кризен	$C_{18}H_{12}$		60,6
7	Пирен	$C_{16}H_{10}$		45,2
8	Трифенилен	$C_{18}H_{12}$		57
9	Бенз[а]антрацен	$C_{18}H_{12}$		61,4
10	Бензо[с]фенантрен	$C_{18}H_{12}$		58,8
11	Перилен	$C_{20}H_{12}$		65,4

Таблица 4.9.2.

№	Название соединения	Формула	Структура	D _{ADTI}
12	Дибенз[а, с]антрацен	C ₂₂ H ₁₄		82,5
13	Дибенз[а, h]антрацен	C ₂₂ H ₁₄		86,8
14	Пентацен	C ₂₂ H ₁₄		91,9
15	Бензо[g]кризен	C ₂₂ H ₁₄		80,3
16	Дибензо[с, g]фенантрен	C ₂₂ H ₁₄		81,7
17	Бензо[g,h,i]перилен	C ₂₂ H ₁₂		74,1
18	Дибензо[def,m]кризен (антантрен)	C ₂₂ H ₁₂		76,3
19	Бензо[а]перилен	C ₂₄ H ₁₄		89
20	Дибензо[bc,kl]коронен	C ₃₀ H ₁₄		122,2

Проведен статистический анализ рассчитанных величин энтальпии образования исследуемых соединений, который указал на высокое качество прогностической способности QSPR-модели. Результаты изложены в таблице 4.9.4. Соотношение между экспериментальными и расчетными данными представлено графически на рисунке 4.22. Коэффициент корреляции $r = 0,99$ и стандартное отклонение $s = 2,17$ ккал/моль. В таблице 4.9.3 объединены предсказанные значения энтальпии образования посредством разработанной QSPR - модели и сравнение с их аналогами из расчетов по методам Бенсона, Моисеевой и Армитажа.

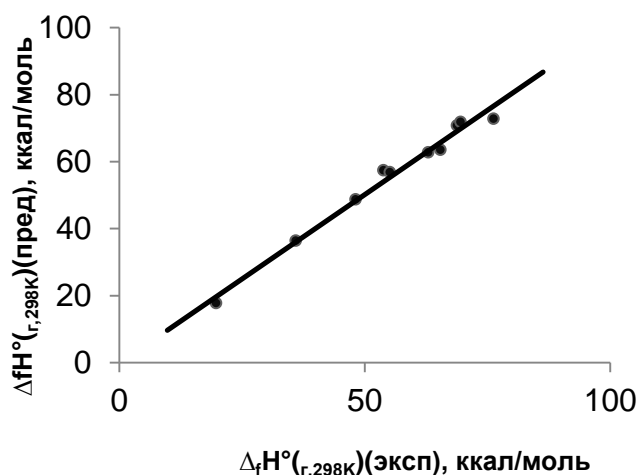


Рисунок 4.22.

Взаимозависимость между экспериментальными и предсказанными величинами энтальпии образования в газовой фазе исследуемых ароматических углеводородов.

Указанная QSPR-модель была использована для вычисления значений энтальпии образования в газовой фазе для исследуемой группы объектов. Результаты изложены в таблице 4.9.5.

Таблица 4.9.3. Результаты вычислений энтальпии образования ПАУ в газовой фазе с использованием QSPR-модели $\Delta_f H^\circ_{(г,298К)} = 1,28 + 0,21D_{ADTI} + 14,64N_1 + 9,24N_2 + 17,91N_3$ (10) и их соизмерение с другими методами. [66]

ПАУ	Молекулярная формула	$\Delta_f H^\circ_{(г,298К)}$ (эксп) ккал/моль	QSPR-модель ккал/моль	$\Delta_f H^\circ_{(г,298К)}$ (пред)- $\Delta_f H^\circ_{(г,298К)}$ (эксп)			
				Бенсон	Моисеева	Armitage	QSPR-модель
Бензол	C ₆ H ₆	19,7	17,8	0,1	-3,1	0,1	-1,9
Нафталин	C ₁₀ H ₈	35,9	35,1	0,1	0,0	0,1	-0,8
Фенантрен	C ₁₄ H ₁₀	49,6	47,9	0,4	0,0	0,4	-1,7
Пирен	C ₁₆ H ₁₀	53,9	58,5	1,3	8,2	1,2	4,6
Антрацен	C ₁₄ H ₁₀	55,2	53,6	-3,0	0,0	-3,0	-1,6
Кризен	C ₁₈ H ₁₂	64,5	61,8	-0,5	-1,8	-0,5	-2,7
Трифенилен	C ₁₈ H ₁₂	65,5	64,3	-3,7	-7,8	-3,7	-1,2
Бенз[а]антрацен	C ₁₈ H ₁₂	70,0	67,3	-4,5	-1,8	-4,5	2,7
Бенз[с]фенантрен	C ₁₈ H ₁₂	69,6	70,1	-3,4	-3,5	-4,5	0,5
Перилен	C ₂₀ H ₁₂	76,3	75,3	-9,4	-6,2	-9,5	-1,0

Таблица 4.9.4. Статистические характеристики расчетных моделей: r коэффициент корреляции между экспериментальными и предсказанными величинами и стандартное отклонение s (ккал/моль).

Модель	r	s	Модель	r	s
Бенсон	0,95	3,59	Армитаж	0,90	4,84
Моисеева	0,84	6,08	QSPR-модель	0,99	2,17

Таблица 4.9.5. Рассчитанные значения энтальпии образования ароматических тестируемых углеводородов с использованием QSPR-модели $\Delta_f H^\circ_{(г,298К)} = 1,28 + 0,21D_{ADП} + 14,64N_1 + 9,24N_2 + 17,91N_3$

№	$\Delta_f H^\circ_{(г,298К)}$ ккал/моль	№	$\Delta_f H^\circ_{(г,298К)}$ ккал/моль	№	$\Delta_f H^\circ_{(г,298К)}$ ккал/моль	№	$\Delta_f H^\circ_{(г,298К)}$ ккал/моль
1	17,8	6	61,8	11	75,3	16	92,8
2	35,1	7	58,5	12	78,9	17	91,7
3	53,6	8	64,3	13	79,8	18	88,9
4	47,9	9	67,3	14	93,8	19	109,0
5	73,1	10	70,1	15	83,8	20	152,3

Используем полученную модель (10) для экстраполяции энтальпии образования в газовой фазе соединений состоящих из трех горизонтальной цепей бензольных колец и проанализируем характер изменения значений (рис. 4.23.).

Графическая зависимость показывает, что изменение значений энтальпии образования приближается с величине близкой к 2.00 ккал/моль.

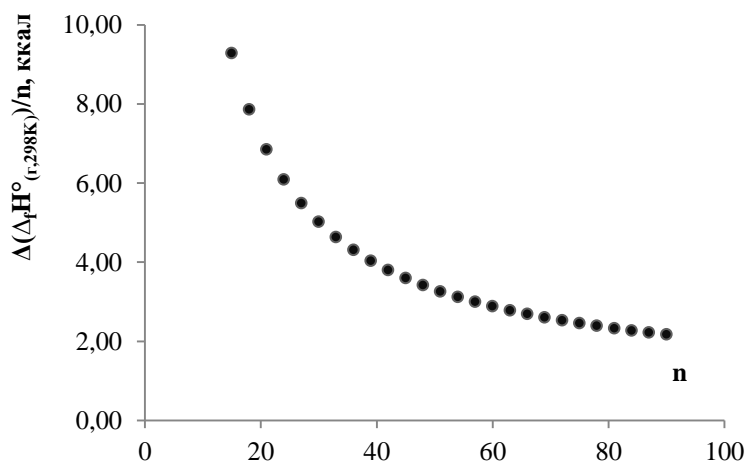


Рисунок 4.23. Зависимость между приращением энтальпии образования в газовой фазе исследуемых ароматических углеводородов ($\Delta(\Delta_f H^\circ_{(г,298К)})/n$) и количеством бензольных колец, (n).

Графическая зависимость показывает, что изменение значений энтальпии образования приближается с величине близкой к 2.00 ккал/моль.

Таблица 4.9.6. Предсказанные значения энтальпии образования $\Delta_f H^\circ_{(г,298К)}$ для молекул состоящих из трех горизонтальных цепей ароматических колец.

$$\Delta(\Delta_f H^\circ_{(г,298К)}) = (\Delta_f H^\circ_{(г,298К)})_n - (\Delta_f H^\circ_{(г,298К)})_{(n-3)}$$

Количество колец, n	$\Delta_f H^\circ_{(г,298К)}$, ккал/моль	$\Delta(\Delta_f H^\circ_{(г,298К)})$, ккал/моль	$\Delta(\Delta_f H^\circ_{(г,298К)})/n$, ккал/моль
12	539,53		
15	678,81	139,28	9,29
18	820,32	141,51	7,86
21	964,06	143,74	6,84
24	1110,03	145,97	6,08
27	1258,23	148,21	5,49
30	1408,68	150,44	5,01
33	1561,36	152,68	4,63
36	1716,28	154,92	4,30
39	1873,43	157,16	4,03
42	2032,83	159,40	3,80
45	2194,47	161,64	3,59
48	2358,34	163,88	3,41
51	2524,46	166,11	3,26
54	2692,81	168,35	3,12
57	2863,40	170,59	2,99
60	3036,24	172,83	2,88
63	3211,31	175,07	2,78
66	3388,62	177,31	2,69

Таблица 4.9.6.

Количество колец, n	$\Delta_f H^\circ_{(г,298К)}$, ккал/моль	$\Delta(\Delta_f H^\circ_{(г,298К)})$, ккал/моль	$\Delta(\Delta_f H^\circ_{(г,298К)})/n$, ккал/моль
69	3568,18	179,55	2,60
72	3749,97	181,79	2,52
75	3934,00	184,03	2,45
78	4120,27	186,27	2,39
81	4308,78	188,51	2,33
84	4499,54	190,75	2,27
87	4692,53	192,99	2,22
90	4887,76	195,23	2,17

4.10. ПРОВЕРКА СТАТИСТИЧЕСКОЙ ГИПОТЕЗЫ ДИСКРИМИНИРУЮЩЕЙ СПОСОБНОСТИ ИНДЕКСА СРЕДНЕГО РАССТОЯНИЯ.

Для оценки дискриминирующей способности топологического индекса среднего расстояния при расчете температуры плавления ПАУ выдвинем нулевую гипотезу при уровне значимости $\alpha = 0,05$: Стандартное отклонение температуры плавления ПАУ экспериментальной группы не имеет существенного различия с контрольной группой. Альтернативная гипотеза предполагает существенное различие между экспериментальными и контрольными стандартными отклонениями.

$$H_0: \mu_{\text{эксп.}} = \mu_{\text{контр.}} \quad H_A: \mu_{\text{эксп.}} \neq \mu_{\text{контр.}}$$

При проведении t теста двух выборок получаем величину p -значимости 0,3813, что делает отклонение нулевой гипотезы нецелесообразным. Если альтернативная гипотеза верна и расхождение экспериментальных и контрольных значений стандартного отклонения статистически значимо, то вероятность

наблюдаемых значений в двух выборках по чистой случайности очень велика. Следовательно, разница стандартного отклонения в экспериментальной и контрольной группах несущественна при доверительном уровне $\alpha = 0,05$.

Для оценки дискриминирующей способности топологического индекса среднего расстояния при расчете энтальпии образования ПАУ выдвинем нулевую гипотезу при вероятностном пороге $\alpha = 0,05$: Стандартное отклонение энтальпии образования ПАУ экспериментальной группы не имеет существенного различия с контрольной группой. Альтернативная гипотеза предполагает существенное различие между экспериментальным и контрольными стандартными отклонениями.

$$H_0: \mu_{\text{эксп.}} = \mu_{\text{контр.}} \quad H_A: \mu_{\text{эксп.}} \neq \mu_{\text{контр.}}$$

При проведении t теста двух выборок величина p -значимости равна 0,0835. Так как уровень надежности не достигнут и вероятность ошибки выше 0,05, то разница в величинах может быть вполне случайной и поэтому отклонить нулевую гипотезу нельзя. Можно заявить, что отличие стандартного отклонения в экспериментальной и контрольной группах несущественно при уровне достоверности $\alpha = 0,05$.

ВЫВОДЫ.

1. В работе обсуждены и использованы важные аспекты теории графов на примере систематизации и вывода изомеров бензола, вывода формул для расчета циклового индекса группы подстановок и производящей функции для определения числа изомеров замещения бензола.
2. С целью создания стандарта для сравнения экспериментальных результатов была построена контрольная группа из 43 QSPR-моделей с использованием известных топологических индексов для определения температуры плавления, температуры кипения, энтальпии образования в газовой фазе, коэффициента распределения в системе октанол-вода углеводородов.
3. Разработан новый топологический индекс среднего расстояния D_{ADTI} и использован для создания экспериментальной группы QSPR-моделей (15) для определения температуры плавления, температуры кипения, энтальпии образования в газовой фазе, коэффициента распределения в системе октанол-вода углеводородов. Проанализирована дискриминирующая способность индекса D_{ADTI} . В построенных QSPR – моделях экспериментальной группы статистические показатели следовали тренду контрольной группы, не выходя за границы доверительного интервала.
4. Проанализирована группа полициклических ароматических углеводородов, состоящая только из ароматических колец. Систематизирована структура этих объектов и разработана многопараметровая QSPR - модель с применением нового топологического индекса для расчета энтальпии образования в газовой фазе. Проведена качественная оценка этой модели посредством сравнения с широко применяемыми расчетными методами, которые составили контрольную группу. Статистические данные построенной QSPR – модели превзошли свои аналоги контрольной группы и подтвердили потенциальную дискриминирующую способность индекса среднего расстояния.

5. Руководствуясь тем фактом, что сравнительный анализ построенной модели с контрольной группой продемонстрировал высокое качество, данная QSPR – модель использована для расчетов неизученных аналогов отобранной группы полициклических углеводородов.
6. Проведена проверка статистических гипотез, которые были выдвинуты по результатам моделирования температуры кипения и энтальпии образования алканов и циклоалканов, температуры плавления и энтальпии образования полициклических ароматических углеводородов для проверки дискриминирующей способности индекса среднего расстояния.

ЛИТЕРАТУРА, ОПУБЛИКОВАННАЯ ПО ТЕМЕ ДИССЕРТАЦИИ.

СТАТЬИ В ЖУРНАЛАХ, РЕКОМЕНДУЕМЫХ ВАК РФ:

1. Vinogradova, M.G. Graph Theory in Structure–Property Correlations / M.G. Vinogradova., Yu.A. Fedina., Yu.G. Papulov // Russian Journal of Physical Chemistry A, 2016, Vol. 90, №2, PP. 411–416.
2. Папулов, Ю.Г. Шестивершинные графы и их применение / Ю.Г. Папулов, Ю.А. Федина, М.Г. Виноградова // Вестн. Твер. Гос. Ун-та. Сер. Химия. 2014. № 1. С. 89 – 96.
3. Федина, Ю.А. КОЛИЧЕСТВЕННЫЕ МОДЕЛИ В КОРРЕЛЯЦИЯХ «СТРУКТУРА-СВОЙСТВО» ОРГАНИЧЕСКИХ СОЕДИНЕНИЙ / Ю.А. Федина, Ю.Г. Папулов, М.Г. Виноградова // Вестн. Тверск. Гос. Ун-та Серия "Химия", №14, 2012 С. 160 – 172
4. Папулов, Ю.Г. Взаимосвязь между строением и свойствами замещенных бензола и родственных соединений: феноменологическое изучение / Ю.Г. Папулов, О.В. Артюхова, Ю.А. Федина, И.Ю. Сеницына // Вестн. Тверск. гос. ун-та. Сер. "Химия". 2009, 13 (вып. 8). С. 4-30.
5. Виноградова, М.Г. Графические зависимости в изучении корреляций структура-энтальпия образования хлорпроизводных алкилсиланов. / М.Г. Виноградова, Ю.Г. Папулов, Ю.А. Федина // Вестник Тверского государственного университета Серия "Химия", №1, 2014 С. 82 – 87.

РАБОТЫ, ОПУБЛИКОВАННЫЕ В ЦЕНТРАЛЬНОЙ ПЕЧАТИ, В
МАТЕРИАЛАХ ОБЩЕРОССИЙСКИХ И МЕЖДУНАРОДНЫХ
КОНФЕРЕНЦИЙ:

6. Papulov, Yu. G. Graph Theoretical Approach To Study In Correlation Of Structure-Property / Yu.G. Papulov, Yu.A. Fedina, M.G. Vinogradova // The Modeling of non-linear processes and systems (The Third International Scientific Symposium). Moscow, June 22 - 25, 2015, P. 147
7. Папулов, Ю.Г. Систематика изомеров замещения молекул / Ю.Г. Папулов, Ю.А. Федина, Р.Ю. Папулов, М.Г. Виноградова // Материалы конференции: Химическая наука: современные достижения и историческая перспектива (II ежегодная Всероссийская научная Интернет-конференция с международным участием). Казань, 9 апреля 2014, С. 115 – 117.
8. Папулов, Ю.Г. Расчет энтальпий образования X-замещенных фенола в атом - атомном представлении / Ю.Г. Папулов, О.В. Артюхова, М.Г. Виноградова, Ю.А. Федина // Актуальные проблемы химической науки, практики и образования: Сб. статей Междунар. научно-практич. конф.: в 2-х ч. Ч. I. Курск, гос. техн. ун-т. 2009, С. 200-202.
9. Федина, Ю.А. Систематика изомеров замещения С₆H₆ / Ю.А. Федина, О.В. Артюхова, Ю.Г. Папулов // Тез. докл. VI Региональных Менделеевских чтений. Тверь: ТвГУ. 2010. С. 36-37.(Удомля, 16-17 апр. 2010 г.)
10. Папулов, Ю.Г. Шестивершинные графы и их приложения / Ю.Г. Папулов, Ю.А. Федина, О.С. Фурялина // Сб. тезисов докл. Международной научной конференции "Моделирование нелинейных процессов и систем". М.: МГТУ "Станкин". 2008. С. 115.
11. Федина, Ю.А. Генерирование и систематизация изомеров замещения молекулярных полиэдров / Ю.А. Федина, Ю.Г. Папулов, О.В. Артюхова // Тез.

докл. VII Региональных Менделеевских чтений. Тверь: ТвГУ. 2011. С. 20-22 (Удомля, 11 февр. 2011 г.).

12. Папулов, Ю.Г. Расчет термодинамических свойств замещенных бензола / Ю.Г. Папулов, О.В. Артюхова, М.Г. Виноградова, Ю.А. Федина // Тез. докл. II Междунар. конф. “Моделирование нелинейных процессов и систем”. Москва, Станкин. 2011. С. 194.

13. Papulov, Yu.G. Calculation of the enthalpies of formation of substituted benzenes in the atom-atomic representation. / Yu.G. Papulov, Yu.A. Fedina, M.G. Vinogradova, O.V. Artyukhova // Absr. XVIII Intern. Conference on Chemical Thermodynamics in Russia, Samara, Oct. 3-7, 2011. Vol. 2. P. 46-47.

14. Папулов, Ю.Г. Теоретико - графовый подход в изучении корреляций “Структура - свойство” / Ю.Г. Папулов, М.Г. Виноградова, Ю.А. Федина // Тез. докл. VI Междунар. науч. конф. “Химическая термодинамика и кинетика”. Тверь, ТвГУ. 2016. С. 202 – 203

РАБОТЫ В ДРУГИХ ИЗДАНИЯХ:

15. Папулов, Ю.Г. Молекулярные графы с шестью вершинами: их характеристические полиномы и спектры. / Ю.Г. Папулов, Т.Г. Кеменова, Ю.А. Федина // Расчетные методы в физической химии. Калинин: КГУ, 1988. С. 4-15

16. Fedina, Y.A. Correlation analysis between boiling point and Randić index of alkanes / Y.A. Fedina, Yu.G. Papulov, M.G. Vinogradova // News of Science and Education, 2014, V. 20, PP. 95 – 98

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

- 17 Набивач, В.М. Сорбционно-структурные корреляции в ряду гетероциклических азотистых соединений / В.М. Набивач // Журн. физич. химии. 1993. Т. 67. №4. С. 821–826.
- 18 Randic, M. In search of structural invariants / M. Randic // J. Math. Chem. 1992. V. 9. P. 97–146.
- 19 Seybold, P.G. Molecular structure —property relationships / P.G. Seybold, M. May, U.A Bagal // J. Chem. Educ. 1987. V. 64. P. 575–581.
- 20 Hansen, P.J. Chemical applications of graph theory. Part I. Fundamentals and topological indices / P.J. Hansen, P.C. Jurs // J. Chem. Educ. 1988. V.65, № 7. P. 574–580.
- 21 Станкевич, М.И. Топологические индексы в органической химии / М.И. Станкевич, И.В. Станкевич, Н.С. Зефиоров // Успехи химии. 1988. Т. 57. №3. С. 337–366.
- 22 Galvez, J. Indices of differences of path lengths: novel topological descriptors derived from electronic interferences in graphs / J. Galvez, R. Garcia—Domenech, C. Gregorio—Alapont // J. Comput. — Aided Molecular Design. 2000. No. 14. P. 679–687.
- 23 Ivanciuc, O. Design of topological indices. Parameters based on electronegativity and covalent radius for the computation of molecular graph descriptors for heteroatom-containing molecules / O. Ivanciuc, T. Ivanciuc, A.T. Balaban // J. Chem. Inf. Comput. Sci. 1998. V. 38. P. 395–401.
- 24 Mandloi, M. A comparative QSAR study using Wiener, Szeged and molecular connectivity indices / M. Mandloi, A. Sikarwar, Sapre // J. Chem. Inf. Comput. Sci. 2000. V. 40. No. 1. P. 57–62.

- 25 Basac, S.C. Topological indices: their nature and mutual relatedness / S.C. Basac, A.T. Balaban, G.D. Grunwald // J. Chem. Inf. Comput. Sci. 2000. V. 40. No. 4. P. 891–898.
- 26 Быков, Г.В. История органической химии./ Г.В. Быков // М.: Химия, 1976.
- 27 Креман, Р. Зависимость между физическими свойствами и химическим строением. / Р. Креман, М. Пестимер //Л.-М.: ГОНТИ НКТП СССР, 1939.
- 28 Бутлеров, А.М. Избранные работы по органической химии./ А.М. Бутлеров //М.: АН СССР, 1951.
- 29 Pedley, J.B.. Thermochemical data of organic compounds. / J.B. Pedley, R.D. Naylor, S.P. Kirby // London, New York: Chapman and Hall, 1986.
- 30 Вилков, Л.В. Физические методы исследования в химии: структурные методы и оптическая спектроскопия. / Л.В. Вилков, Ю.А. Пятин // М.: Высшая школа, 1987.
- 31 Вшков, Л.В. Физические методы исследования в химии: резонансные и электрооптические методы. / Л.В. Вшков, Ю.А. Пятин // М.: Высшая школа, 1989.
- 32 Степанов, Н.Ф. Методы линейной алгебры в физической химии./ Н.Ф. Степанов, М.Е. Ерлыкина, Г.Г. Филиппов // М.: МГУ, 1976.
- 33 Киреев, В.А. Методы практических расчетов в термодинамике химических реакций. / В.А. Киреев // М.: Химия, 1975.
- 34 Татевский, В.М. Химическое строение углеводородов и закономерности в их физико-химических свойствах. / В.М. Татевский // М.: МГУ, 1953.
- 35 Татевский, В.М. Классическая теория строения молекул и квантовая механика. / В.М. Татевский // М.: МГУ, 1973.
- 36 Татевский, В.М. Связь энергии образования молекулы из свободных атомов с её строением. I / В.М. Татевский, Ю.Г. Папулов // Журн. физ. химии. 1960. Т. 34. № 2. С. 241-258.

- 37 Татевский, В.М. Связь энергии образования молекулы из свободных атомов с её строением. II / В.М. Татевский, Ю.Г. Папулов // Журн. физ. химии. 1960. Т. 34. № 3. С. 489-504.
- 38 Папулов, Ю.Г. Аддитивные схемы расчета энергий образования алканов в атом-атомном приближении / Ю.Г. Папулов, В.М. Смоляков, В.А. Филиппов, В.Т. Кондратьева, А.А. Ктикян // Свойства веществ и строение молекул. Калинин, 1975.
- 39 Виноградова, М.Г. Количественные корреляции «структура-свойство» алканов. Аддитивные схемы расчета. / М.Г. Виноградова, Ю.Г. Папулов, В.М. Смоляков // Тверь: ТвГУ, 1999.
- 40 Папулов, Ю.Г. Физические свойства и химическое строение. / Ю.Г. Папулов, В.М. Смоляков // Калинин: КГУ, 1981.
- 41 Татевский, В.М. Строение и физико-химические свойства молекул и веществ. / В.М. Татевский // М.: МГУ, 1993.
- 42 Кинг, Р. Химические приложения топологии и теории графов / Р. Кинг. // М. Мир, 1987.
- 43 Зефилов, Н.С. Применение теории графов в химии / Н.С.Зефилов, С.И. Кучанов. // Новосибирск: Наука, 1988.
- 44 Chirico, R.D. The Thermodynamic Properties to the Temperature 700 K of Naphthalene and of 2,7-Dimethylnaphthalene / R.D. Chirico; S.E. Knipmeyer; A. Nguyen, W.V. Steele, // J. Chem. Thermodyn. 1993, 25, 1461-1494.
- 45 Михайлов, И.В. Корреляция между размерами и топологическими индексами для ряда сверхразветвленных полимеров / И.В Михайлов, О.В. Шавыкин, А.А. Даринский, И.М. Неелов // Вестник ТвГУ, 2015, Серия: Химия (2). с. 68-87
- 46 Папулов, Ю.Г. Молекулярные графы в химических исследованиях / Ю.Г.Папулов //Калинин, 1990.

- 47 Раевский, О.А. Дескрипторы молекулярной структуры в компьютерном дизайне биологически активных молекул / О.А. Раевский // Успехи химии. 1999. Т.68, № 6. С. 555-575.
- 48 Стьюпер, Э. Машинный анализ связи химической структуры и биологической активности. / Э. Стьюпер, У. Брюгер, П.М. Джуре // Мир, 1982.
- 49 Balaban, A.T. Correlations between chemical structure and normal boiling points of halogenated alkanes C1-C4 / A.T. Balaban, N. Joshi, L.B. Kier, L.H. Hall // J. Chem. Inf. Comput. Sci. 1992. V.32, № 3. P. 233-237.
- 50 Katritzky, A.R. Normal boiling points for organic compounds / A.R. Katritzky // J. Chem. Inf. Comput. Sci. 1998. V.38, № 1. P. 28-38.
- 51 Zheng, O. An effective topological symmetry perception and unique numbering algorithm / O. Zheng, Y. Shengang, J Brandt, Z. Chongzhi // J. Chem. Inf. Comput. Sci. 1999. V.39, № 2. P. 299-303.
- 52 Xinhua, L. A novel quantum-topology index / L. Xinhua, Y. Qingsen, Z. Longguan // J. Chem. Inf. Comput. Sci. 2000. V.40, № 2. P. 399-402.
- 53 Sullivan, J. QSAR treatment of electronic substituent effects frontier orbital theory and topological parameters / J. Sullivan, A.D. Jones, K. Tanji Kenneth // J. Chem. Inf. Comput. Sci. 2000. V.40, №5. P. 1113-1127.
- 54 Oblak, M. QSAR of flavonoid analogues. 3 / M. Oblak, M. Randic, T. Solmajer // J. Chem. Inf. Comput. Sci. 2000. V.40, № 4. P. 994-1001.
- 55 Беек, В. QM/NN QSPR models with error estimation: vapor pressure and log P / В. Беек, А. Brendl, Т. Clark. // J. Chem. Inf. Comput. Sci. 2000. V.40, № 4. P. 1046-1051.
- 56 Randic, M. Molecular shape profiles / M. Randic // J. Chem. Inf. Comput. Sci. 1995. V.35, № 2. P. 373-382.
- 57 Randic, M. On characterization of chemical structure / M. Randic // J. Chem. Inf. Comput. Sci. 1997. V.37, № 4. P. 672-685.

- 58 Ghose, A.K. Distance geometry QSAR: current achievements and future extensions. In "QSAR and strategies in the design of bioactive compounds" / A.K. Ghose, G.M Crippen // Seydel J.K., Ed.; VCH: 1985. P. 116-119.
- 59 Estrada, E. Three-dimensional molecular descriptors based on charge density weighted graphs / E. Estrada // J. Chem. Inf. Comput. Sci. 1995. V.35. P. 708-713.
- 60 Talele, T.T. Three-dimensional quantitative structure-activity relationships and receptor mapping of cytochrome P-450 inhibiting azole antifungal agents / T.T. Talele, K.M. Kulkazni // J. Chem. Inf. Comput. Sci. 1999. V.39, № 2. P. 204-210.
- 61 Venkatarangan, P. Prediction of ligand-receptor binding free energy by 4D-QSAR analysis: application to a set of glucose analogue inhibitors of glycogen phosphorylase / P. Venkatarangan, A.J. Hopfinger // J. Chem. Inf. Comput. Sci. 1999. V.39, № 6. P. 1141-1150.
- 62 Thy-Hou, Lin Classification of some active compounds and their inactive analogues using two three-dimensional molecular descriptors / Lin Thy-Hou, YU Yi-S., Chen Hong-Jih. // J. Chem. Inf. Comput. Sci. 2000. V.40, № 5. P. 1210-1221.
- 63 Bursi, R. Comparative Spectra Analysis (CoSA): spectra as three-dimensional molecular descriptors for the prediction of biological activities / R. Bursi., T. Dao, T. van Wijk et a. // J. Chem. Inf. Comput. Sci. 1999. V.39, № 5. P. 861-867.
- 64 Livingstone, D.J. The characterization of chemical structure using molecular properties / D.J. Livingstone // J. Chem. Inf. Comput. Sci. 2000. V.40, № 2. P. 195-209.
- 65 Cajan, M. Stability of complexes of aromatic amides with bromide anion: quantitative structure-activity relationships / M. Cajan, J Dumborsky, I. Stibor, J. Koca // J. Chem. Inf. Comput. Sci. 2000. V.40, №5. P. 1151-1157.
- 66 Yu, J. Accurate and efficient method for predicting thermochemistry of polycyclic aromatic hydrocarbons – Bond-centered group additivity / J. Yu, R. Sumathi, W.H. Jr. Green // Journal of the American Chemical Society 2004 V. 126 Issue 39, PP. 12685-12700

- 67 McClelland, H.E. Quantitative structure-property relationships for the prediction of vapor pressures of organic compounds from molecular structures / H.E. McClelland, P.C. Jurs // *J. Chem. Inf. Comput. Sci.* 2000. V.40, № 4. P. 967-975.
- 68 Hiob, R. Quantitative relationship between rate constants of the gas-phase homolysis of C-X bonds and molecular descriptors / R. Hiob, M. Karelson // *J. Chem. Inf. Comput. Sci.* 2000. V.40, №4. P. 1062-1071.
- 69 Schweitzer, R.C. Improved quantitative structure-property relationships for the prediction of dielectric constants for a set of diverse compounds by subsetting of the data set / R.C. Schweitzer, J.B. Morris // *J. Chem. Inf. Comput. Sci.* 2000. V.40, № 5. p. 1253-1261.
- 70 Jain, D.V.S. Correlation between topological features and physicochemical properties of molecules / D.V.S Jain., S. Singh, V. Gombar // *Proc. Indian. Acad. Sci. (Chem. Sci.)*. 1984. V.93, № 6. P. 927-945.
- 71 Kumskov, M.I. Automatic formation method for structural descriptors of organic compounds for quantitative structure property relationships / M.I. Kumskov, L.A. Ponomareva E.A., Smolenskii, D.F. Mitushev, N.S. Zefirov // *Russian Chemical Bulletin*, 1994 V. 43, №8. pp. 1317-1319.
- 72 Smolenskii, E.A. Method of inverse functions for solution of structure-property relationship problems / E.A. Smolenskii., A.N. Ryzhov, V.M. Bavykin, I.V. Chuvaeva, A.L. Lapidus // *Doklady Chemistry.*, 2007. V. 417. №1. pp. 267-272
- 73 Mihalic, Z. Graph-theoretical approach to structure-property relationships / Z. Mihalic, N. Trinajstic // *J. Chem. Educ.* 1992. V.69, № 9. P. 701-712.
- 74 Randic, M. The nature of chemical structure / M. Randic // *J. Math. Chem.* 1990. V.4. P. 157-184.
- 75 Рувре, Д.Г. Химиюпрогнозируеттопология / Д.Г. Рувре // *В мире науки (Scientificamerican)*. 1986. №11. С. 14-22.
- 76 Balaban, A.T. Chemical graphs: looking back and glimpsing ahead / A.T. Balaban // *J. Chem. Inf. Comput. Sci.* 1995. V.35. P. 339-350.

- 77 Balaban, A.T. Application of graph theory in chemistry / A.T. Balaban // J. Chem. Inf. Comput. Sci. 1985. V.25, № 3. P. 334-343.
- 78 Wiener, H. Structural determination of paraffin boiling points / H. Wiener // J. Am. Chem. Soc. 1947. V.69, № 1.P. 17-20.
- 79 Wiener, H. Correlation of heats of vaporization isomers, among the paraffin hydrocarbons / H. Wiener // J. Am. Chem. Soc. 1947. V.69. P. 2636-2638.
- 80 Wiener, H. Influence of interatomic forces on paraffin properties / H. Wiener // J. Chem. Phys. 1947. V.15.P. 766.
- 81 Wiener, H. Vapor pressure-temperature relationships among the branched paraffin hydrocarbons / H. Wiener // J. Phys. Chem. 1948. V.52. P. 425-430.
- 82 Wiener, H. Relation of the physical properties of the isomeric alkanes to molecular structure / H. Wiener // J. Phys. Chem. 1948. V.52. P. 1082-1089.
- 83 Piatt, J.R. Prediction of isomeric differences in paraffin properties / J.R. Piatt // J. Phys. Chem. 1952. V.56. P. 328-336.
- 84 Balaban, A.T. Chemical applications of graph theory. / A.T. Balaban // London: Academic Press, 1976.
- 85 Randic, M. On characterization on molecular branching / M. Randic // J. Am. Chem. Soc.1975. V.97. P. 6609-6615.
- 86 Kier, L.B. Molecular connectivity in chemistry and drug research./ L.B. Kier, L.H. Hall // New York: Academic Press, 1976.
- 87 Kier, L.B. Molecular connectivity in structure-activity analysis../ L.B. Kier, L.H. Hall // New York: Wiley, 1986.
- 88 Milne, G.W. Mathematics as a basis for chemistry / G.W. Milne // J. Chem. Inf. Comput. Sci. 1997. V.37. P. 639-644.
- 89 Ни, С. On highly molecular topological index / С. Ни, L. Xu // J. Chem. Inf. Comput. Sci. 1996. V.36. P. 82-90.

- 90 Yao, Y. Study on structure-activity relationships of organic compounds: three new topological indices and their applications / Y. Yao , L. Xu, Y. Yang, X. Yuan // J. Chem. Inf. Comput. Sci. 1993. V.33. P. 590-594.
- 91 Bogdanov, B. On the three-dimensional Wiener number / B. Bogdanov, S. Nolic, N. Trinajstic // J. Math. Chem. 1989. V.3. P. 299-309.
- 92 Bogdanov, B. On the three-dimensional Wiener number. A comment / B. Bogdanov, S. Nolic, N. Trinajstic // J. Math. Chem. 1990. V.5. P. 305-306.
- 93 Mihalic, Z. Comparative study of molecular descriptors derived from the distance matrix / Z. Mihalic, S. Nolic, N. Trinajstic // J. Chem. Inf. Comput. Sci. 1992. V.32. P. 28-37.
- 94 Zhu, H. Graph-geometric invariants for molecular structure / H. Zhu, D.J. Klein // J. Chem. Inf. Comput. Sci. 1996. V.36. P. 1067-1075.
- 95 Randic, M. Development of 3-dimensional molecular descriptors / M. Randic, B. Jeman-Blazic, N. Trinajstic // J. Chem. Inf. Comput. Sci. 1992. V.32. P. 28-37.
- 96 Diudea, M.V. Molecular topology. 1. 3D-distance matrices and related topological indices / M.V. Diudea, D. Horvath, A. Graovac // J. Chem. Inf. Comput. Sci. 1995. V.35. P. 129-135.
- 97 Randic, M. Molecular topographic indices / M. Randic // J. Chem. Inf. Comput. Sci. 1995. V.35. P. 140-147.
- 98 Randic, M. Molecular profiles. Novel geometry-dependent molecular descriptors / M. Randic // New J. Chem. 1995. V. 19. P. 781-791.
- 99 Randic, M. Molecular bonding profiles / M. Randic // J. Math. Chem. 1996. V. 19. P. 375-392.
- 100 Hall, L.H. The electrotopological state: structure information at the atomic level for molecular graphs / L.H. Hall, B. Mohny, L.B. Kier // J. Chem. Inf. Comput. Sci. 1991. V.31. P. 76-82.

- 101 Hall, L.H. Electrotopological state indices for atom types: a novel combination of electronic, topological and valence shell information / L.H. Hall, L.B. Kier // J. Chem. Inf. Comput. Sci. 1995. V.35. P. 1039-1045.
- 102 Xu, L. New topological index and prediction of phase transfer energy for protonated amines and tetraalkylamines ions / L. Xu, Y. Yao, X.-M. Wang // J. Chem. Inf. Comput. Sci. 1995. V.35, №1. P. 45-49.
- 103 Estrada, E. Edge adjacency relationships and a novel topological index related to molecular volume / E. Estrada // J. Chem. Inf. Comput. Sci. 1995.V.35.P.31-33.
- 104 Gutman, I. Topological indices based on the line graphs of the molecular graph / I. Gutman I, E. Estrada //J. Chem. Inf. Comput. Sci. 1996. V.36. P. 541-543.
- 105 Tratch, S.S. Combinatorial models and algorithm in chemistry. The expanded Wiener number a novel topological index / S.S. Tratch, M.I. Stankevitch, N.S. Zefirov // J. Comput. Chem. 1990. V.30, № 11. P. 899-908.
- 106 Randic, M. Wiener matrix: source of novel graph invariants / M. Randic, G. Xiaofeng, T. Qxley, H. Krishnapriyan // J. Chem. Inf. Comput. Sci. 1993. V.33. P. 706-716.
- 107 Hall, L.H. Computational aspects of molecular connectivity and its role in structure-property modeling. In " Computational graph theory". / H.L. Hall // New York: Nova Publishing, 1990. P. 202-233.
- 108 Randic, M. Distance/distance matrices / M. Randic, A.E. Kleiner, L.M. De Alba // J. Chem. Inf. Comput. Sci. 1994. V.34. P. 277-286.
- 109 Trinajstic, N. The detour matrix in chemistry / N. Trinajstic, S. Nicolic, B. Lucic, D. Amic, Z. Mihalic // J. Chem. Inf. Comput. Sci. 1997. Y.37. P. 631-638.
- 110 Klein, D.J. Resistance distance / D.J. Klein, M. Randic // J. Math. Chem.1993. V.12. P. 81-95.
- 111 Trinajstic, N. The laplasian matrix in chemistry / N. Trinajstic, D. Babic, S. Nicolic, D. Plavsic, D. Amic, Z. Mihalic // J. Chem. Inf. Comput. Sci. 1994. V.34. P. 368-376.

- 112 Lucovits, I.A. A compact form of the adjacency matrix / I.A. Lucovits // *J. Chem. Inf. Comput. Sci.* 2000. V.40, № 5. P. 1147-1150.
- 113 Shushen, L. Molecular distance-edge vector: an extension from alkanes to alcohols / L. Shushen, L. Hailing, X. Zhining et al. // *J. Chem. Inf. Comput. Sci.* 1999. V.39. P. 951-957.
- 114 Schultz, H.P. Topological organic chemistry. 1 / H.P. Schultz // *J. Chem. Inf. Comput. Sci.* 1989. V.29. P. 227-228.
- 115 Schultz, H.P. Topological organic chemistry. 2 / H.P. Schultz, E.B. Schultz, T.P. Schultz // *J. Chem. Inf. Comput. Sci.* 1990. V.30. P. 27-29.
- 116 Schultz, H.P. Topological organic chemistry. 3 / H.P. Schultz, T.P. Schultz // *J. Chem. Inf. Comput. Sci.* 1991. V.31. P. 144-147.
- 117 Schultz, H.P. Topological organic chemistry. 4 / H.P. Schultz, E.B. Schultz, T.P. Schultz // *J. Chem. Inf. Comput. Sci.* 1992. V.32. P. 68-72.
- 118 Schultz, H.P. Topological organic chemistry. 5 / H.P. Schultz, E.B. Schultz // *J. Chem. Inf. Comput. Sci.* 1992. V.32. P. 364-368.
- 119 Schultz, H.P. Topological organic chemistry. 6 / H.P. Schultz, E.B. Schultz // *J. Chem. Inf. Comput. Sci.* 1993. V.33. P. 240-244.
- 120 Schultz, H.P. Topological organic chemistry. 7 / H.P. Schultz, E.B. Schultz, T.P. Schultz // *J. Chem. Inf. Comput. Sci.* 1993. V.33. P. 863-867.
- 121 Schultz, H.P. Topological organic chemistry. 12 / H.P. Schultz, T.P. Schultz // *J. Chem. Inf. Comput. Sci.* 2000. V.40. P. 107-112.
- 122 Schultz, H.P. Topological organic chemistry. 13 / H.P. Schultz // *J. Chem. Inf. Comput. Sci.* 2000. V.40. P. 1158-1159.
- 123 Bytautas, L. Mean Wiener numbers and other mean extensions for alkanes trees / L. Bytautas, D.J. Klein // *J. Chem. Inf. Comput. Sci.* 2000. V.40. P. 471-481.
- 124 Basak, S.C. QSPR modeling: graph connectivity indices versus line graph connectivity indices / S.C. Basak, S. Nicolic, N. Trinajstic // *J. Chem. Inf. Comput. Sci.* 2000. V.40. P. 920-933.

- 125 Bonchev, D. Overall connectivities/topological complexities: a new powerful tool for QSPR/QSAR /D. Bonchev // J. Chem. Inf. Comput. Sci. 2000. V.40. P. 934-941.
- 126 Gutman, I. Bounds for the Randic connectivity index /I. Gutman, O. Araujo, D.A. Morales // J. Chem. Inf. Comput. Sci. 2000. V.40. P. 593-598.
- 127 Mandloi, M. A comparative QSAR study using Wiener, Szeged and molecular connectivity indices / M. Mandloi, A. Sikarwar, N.S. Sapre et al. // J. Chem. Inf. Comput. Sci. 2000. V.40. P. 57-62.
- 128 Gutman, I. The multiplicative version of the Wiener index / I. Gutman, W. Linert, I. Lucovits // J. Chem. Inf. Comput. Sci. 2000. V.40. P. 113-116.
- 129 Estrada, E. Edge-connectivity indices in QSPR/QSAR studies. 1. Comparison to other topological indices in QSPR studies / E. Estrada, L. Rodriguez // J. Chem. Inf. Comput. Sci. 1999. V.39. P. 1037-1041.
- 130 Dobrynin, A.A. The average Wiener index of hexagonal chains / A.A. Dobrynin, I. Gutman // J. Chem. Inf. Comput. Sci. 1999. V.39. P. 571-576.
- 131 Balaban, A.T. Correlation between chemical structure and normal boiling points of acyclic ethers, peroxides, acetals, and their sulfur analogues / A.T. Balaban, L.B. Kier, N. Joshi // J. Chem. Inf. Comput. Sci. 1992. V.32. P. 237-244.
- 132 Amic, D. Calculation of retention times of anthocyanins with orthogonalized topological indices / D. Amic, D. Davidovic-Amic, N. Trinajstic // J. Chem. Inf. Comput. Sci. 1995. V.35. P. 136-139.
- 133 Amic, D. Structure-activity correlation of flavone derivatives for inhibition of cAMP phosphodiesterase / D. Amic, D. Davidovic-Amic, A. Juric, B. Lucic, N. Trinajstic // J. Chem. Inf. Comput. Sci. 1995. V.35. P. 1034-1038.
- 134 Pogliani, L. Molecular connectivity descriptors of the physicochemical properties of the α -aminoacids / L. Pogliani // J. Phys. Chem. 1994. V.98. P. 1494-1499.
- 135 Estrada, E. Graph theoretical invariants of Randic revisited /E. Estrada // J. Chem. Inf. Comput. Sci. 1995. V.35. P. 1022-1025.

- 136 Qian-Nan, H. The Matrix Expression, Topological Index and Atomic Attribute of Molecular Topological Structure / H. Qian-Nan, L. Yi-Zeng, F. Kai-Tai // J. of Data Science 2003 №1 p. 361-389
- 137 Diudea, M.V. Molecular topology. Hyper-Wiener index of dendrimers / M.V. Diudea, B. Parv // J. Chem. Inf. Comput. Sci. 1995. V.35. P. 1015-1018.
- 138 Lucovits, I. A novel definition of hyper-Wiener index for cycles / I. Lucovits, W. Linert // J. Chem. Inf. Comput. Sci. 1994. V.34. P. 899-902.
- 139 Diudea, M.V. Wiener and Hyper-Wiener numbers in a single matrix / M.V. Diudea // J. Chem. Inf. Comput. Sci. 1996. V.36. P. 833-836.
- 140 Mohar, B. A novel definition of the Wiener index for trees / B. Mohar, D. Babic, N. Trinajstic // J. Chem. Inf. Comput. Sci. 1993. V.33.P. 153-154.
- 141 Nandy, A. Simple numerical descriptors for quantifying effect of toxic substances on DNA sequences / A. Nandy, S.C. Basak // J. Chem. Inf. Comput. Sci. 2000. V.40. P. 915-919.
- 142 Ponce, A.M. Study of the action of flavonoids on xanthineoxidase by molecular topology / A.M. Ponce, S.E. Blanco, A.S. Molina et al. // J. Chem. Inf. Comput. Sci. 2000. V.40. P. 1039-1045.
- 143 Sung-Sau, S. Quantitative structure-activity relationships studies of progesterone receptor binding steroids / S. Sung-Sau, P. Van Helden Steven, V.J. van Geerestein // J. Chem. Inf. Comput. Sci. 2000. V.40. P. 762-772.
- 144 Katritzky, A.R. Structurally diverse quantitative structure-property relationship correlations of technologically relevant physical properties / A.R. Katritzky, U. Maran, V.S. Lobanov // J. Chem. Inf. Comput. Sci. 2000. V.40. P. 1-18.
- 145 Balaban, A.T. Normal boiling points of 1,co-alkanedinitriles: the highest increment in a homologous series / A.T. Balaban, S.C. Basac, D. Mills // J. Chem. Inf. Comput. Sci. 1999. V.39. P. 769-774.
- 146 Rucker, G. On topological indices, boiling points, and cycloalkanes / G. Rucker, C. Rucker // J. Chem. Inf. Comput. Sci. 1999. V.39. P. 788-802.

- 147 Balaban, A.T. Correlation between structure and normal boiling points of acyclic carbonyl compound / A.T. Balaban., D. Mills, S.C. Basac // J. Chem. Inf. Comput. Sci. 1999. V.39. P. 758764.
- 148 Banken, G.A Prediction of metyl radical addition rate constants from molecular structure / G.A. Banken, P.C. Jurs // J. Chem. Inf. Comput. Sci. 1999. V.39. P.508-514.
- 149 Ivanciuc, O. Estimation of the liquid viscosity of organic compounds with a QSPR model / O. Ivanciuc, T. Ivanciuc, P.A. Filip et al. // J. Chem. Inf. Comput. Sci. 1999. V.39. P. 515-524.
- 150 Pompe, M. Prediction of gas-chromatographic retention indices using topological descriptors / M. Pompe, M. Novic // J. Chem. Inf. Comput. Sci. 1999. V.39. P.59-67.
- 151 Zerovnik, J. Szeged index of symmetric graphs /Zerovnik J. //J. Chem. Inf. Comput. Sci. 1999. V.39. P. 77-80.
- 152 Gao, HuaComparative QSAR analysis of estrogen receptor ligands / Hua Gao, G. Rajni, H. Corwin // Chem. Rev. 1999. V.99, № 3. P. 723-744.
- 153 Estrada, E. Edge-connectivity indices in QSPR/QSAR studies. 2. Accounting for long-range bond contributions / E. Estrada // J. Chem. Inf. Comput. Sci. 1999. V.39. P. 1042-1048.
- 154 Скворцова, М.И. Методология построения общей модели связи «структура-свойство» на топологическом уровне / М.И. Скворцова, И.И. Баскин, О.Л. Словоохотова, Н.С. Зефирова // Докл. РАН. 1994. Т. 336. С. 496-499.
- 155 Randic, M. Resolution of ambiguities in structure-property studies by use of orthogonal descriptors / M. Randic // J. Chem. Inf. Comput. Sci. 1991. V.31. P. 311-320.
- 156 Randic, M. Orthogonal molecular descriptors / M. Randic // New J. Chem. 1991. V. 15. P.517-525.
- 157 Randic, M. Chemical structure what is "she"? / M. Randic // J. Chem. Educ.1992. V.69. P. 713-718.
- 158 Randic, M. ON GENERALIZATION OF WIENER INDEX FOR CYCLIC STRUCTURES / M. Randic //Acta Chim. Slov. 2002. V. 49 P. 483–496.

- 159 Randic, M. Construction of high-quality structure-property-activity regressions: the boiling points of sulfides / M. Randic, S.C. Basak // *J. Chem. Inf. Comput. Sci.* 2000. V.40. P. 899-905.
- 160 Dearden, J.C. The History and Development of QSARs / J.C. Dearden // *Int. J. of QSPR* 2016 V. 1, Issue 1, pp 1-44
- 161 Soskic, M. Link between orthogonal and standart multiple linear regression models / M. Soskic, D. Plavsic, N. Trinajstic // *J. Chem. Inf. Comput. Sci.* 1996. V.36. P. 829-832.
- 162 Vracko, M. A study of structure-carcinogenic potency relationships with artificial neural networks / M. Vracko // *J. Chem. Inf. Comput. Sci.* 1997. V.37. P. 1037-1043.
- 163 Tetko, U.V. Neural network studies / U.V Tetko, D. Livingstone // *J. Chem. Inf. Comput. Sci.* 1995. V.35. P. 826-833.
- 164 Kireev, D.V. A novel neural network based method for graph/property mapping / D.V. Kireev // *J. Chem. Inf. Comput. Sci.* 1995. V.35. P. 175-180.
- 165 Elrod, D. Application of neural networks in chemistry. 1. Prediction of electrophilic aromatic substitution reactions / D. Elrod, G.M. Maggiora, R.G. Trenary // *J. Chem. Inf. Comput. Sci.* 1990. V.30. P. 477-484.
- 166 Скоробогатов, В.А. Применение теории графов в химии полициклических бензоидных углеводородов. / В.А. Скоробогатов, Е.В. Мжельская // Новосибирск ИМ, 1987, 33 с.
- 167 Dias, J.R. Handbook of Polycyclic Hydrocarbons / J.R. Dias // Elsevier, Amsterdam, 1988
- 168 Hopfinger, A.J. Application of genetic algoritms to the general QSAR problem and guiding molecular diversity experiments. In In "Genetic algoritms in molecular modeling" / A.J. Hopfinger, H.C. Patel // (Ed. Devillers J.) London: Academic Press, 1996. P. 137-157.
- 169 Kier, L.B. A celler automata model of water / L.B. Kier, C.-K. Cheng // *J. Chem. Inf. Comput. Sci.* 1994.V.34. P. 647-652.

- 170 Kier, L.B. A celler automata model of aqueous solution / L.B. Kier, C.-K. Cheng // J. Chem. Inf. Comput. Sci. 1994. V.34. P. 1334-1337.
- 171 Cheng, C.-K. A celler automata model of oil-water partitioning / C.-K. Cheng, L.B. Kier // J. Chem. Inf. Comput. Sci. 1995. V.35. P. 1054-1059.
- 172 Lucic, B. Non linear multivariate regression out performs several concisely designed neural networks on three QSPR data sets / B. Lucic, D. Amic, N. Trinajstic // J. Chem. Inf. Comput. Sci. 2000. V.40. P. 403-413.
- 173 Meiler, J. Fast determination of C-NMR chemical shifts using artificial neural networks / J. Meiler, R. Meusinger // J. Chem. Inf. Comput. Sci. 2000. V.40. P.1 169-1176.
- 174 Liang, C.K. QSPR prediction of vapor pressure from solely theoretically-derived descriptors / C.K. Liang, D.A. Gallagher // J. Chem. Inf. Comput. Sci. 1998. V.38. P. 321-324.
- 175 Smith, N.K. / N.K. Smith; R.C. Jr. Stewart, A.G. Osborn; D.W. Scott // J. Chem. Thermodyn. 1980, 12, 919-926.
- 176 Slayden, S.W / S.W. Slayden; J.F. Liebman // Chem. Rev. 2001, 101, 1541-1566.
- 177 Diogo, H.P. / H.P. Diogo, M.E.M. Piedade // J. Chem. Thermodyn. 2002, 34, 173-184.
- 178 Ed. Rumelhart, D.E. Parallel distributed processing: explorations in the microstructure of cognition / Ed. Rumelhart D.E., McClelland J.L. // Cambridge: MIT Press, 1986.
- 179 Баскин, И.И. Решение инверсной проблемы «структура-свойство» для случая топологических индексов / И.И. Баскин, Е.В. Гордеева, Р.О. Девдариани, Н.С. Зефилов, В.А. Палюлин, М.И. Станкевич // Докл. РАН. 1989. Т. 307. С. 613-617.
- 180 Зефилов, Н.С. Проблема генерации структур с заданными свойствами и решение инверсной проблемы для индекса Балабана / Н.С. Зефилов, В.А. Палюлин, Е.В. Радченко // Докл. РАН. 1991. Т. 316. С. 921-924.

- 181 Scvortsova, M.I. Inverse problem in QSAR/QSPR studies for the case of topological indices characterizing molecular shape (Kier indices) / M.I. Scvortsova, I.I. Baskin, O.L. Slovokhotova, V.A. Palyulin, N.C. Zefirov // J. Chem. Inf. Comput. Sci. 1993. V.33. P. 630-634.
- 182 Palyulin, V.A. Molecular field topology analysis method in QSAR studies of organic compounds / V.A. Palyulin, E.V. Radchenko, N.C. Zefirov // J. Chem. Inf. Comput. Sci. 2000. V.40. P. 659-667.
- 183 Basak, S.C. Topological indices: their nature and mutual relatedness / S.C. Basak, A.T. Balaban, G.D. Grunwald, B.D. Gute // J. Chem. Inf. Comput. Sci. 2000. V.40. P. 891-898.
- 184 Папулов, Ю.Г. Использование топологических индексов при построении корреляций структура-свойство / Ю.Г. Папулов, Т.Н. Чернова, В.М. Смоляков, М.И. Поляков // Журн. физ. химии. 1993. Т. 67. С. 203-209.
- 185 CRC Handbook of chemistry and physics: a ready-reference book of chemical and physical data (Ed. Lide D.R.) Boca Raton, Fla; London: CRC Press, 1991.
- 186 Cox, J.D. Thermochemistry of organic and organometallic compounds. / J.D. Cox, G. Pilcher // London, New York: Academic Press, 1970.
- 187 Термодинамические свойства индивидуальных веществ Под ред. Глушко В.П. // Справочное издание. М.: Наука, 1979.
- 188 Общая Органическая Химия, Том 1, Перевод с английского д.х.н. Л. Я. Яновской, под редакцией акад. Н. К. Кочеткова, Москва, Химия, 1981
- 189 Справочник химика. Л.: Химия, 1971.
- 190 Львовский, Е.Н. Статистические методы построения эмпирических формул. /Е.Н. Львовский //М.: Высшая школа, 1982.
- 191 Liu, J. A Laplacian-energy-like invariant of a graph / J. Liu, B. Liu // MATCH Commun Math Comput Chem, 2008, 59: 355-372.
- 192 Харари, Ф. Перечисление графов. / Ф. Харари, Э. Палмер // Пер. с англ. — М.: Мир, 1977.

- 193 Experimental Thermochemistry, ed. F. D. Rossini, Interscience, New York, 1956, vol. 1;
- 194 Cox, J.D. Thermochemistry of Organic and Organometallic Compounds ./ J.D. Cox, G. Pilcher // Academic, London, 1970
- 195 Stull, D.R. The Chemical Thermodynamics, of Organic Compounds / D.R. Stull, E.F. Westrum, Jr., G.C. Sinke // Wiley, New York, 1969
- 196 Smolenskii, E.A. Modeling of physicochemical characteristics of alkanes / E.A. Smolenskii, A.N. Ryzhov, A.L. Lapidus, N.S. Zefirov // Russian Chemical Bulletin. 2005. V. 54. №5. pp. 1084-1099
- 197 Smolenskii, E.A. Calculations of melting points of monoalkylalkanes / Smolenskii E.A., Maslova L.K., Shpil'kin S.A., Zefirov N.S. // Doklady Physical Chemistry, 1999 V. 368, №1-3 PP. 256-259
- 198 Smolenskii, E.A. Simulation of melting temperatures of normal alkanes / E.A. Smolenskii., A.N. Ryzhov, A.L. Lapidus, N.S. Zefirov. // Doklady Chemistry, 2002 V. 387, №1-3 PP: 297-301
- 199 Smolenskii, E.A. The Chemical Structure Matrix and a New Formalism for the QSPR Problem / E.A. Smolenskii, // Russian Chemical Bulletin, 2006 V.55, №9 PP. 1501-1507