

НАЦИОНАЛЬНАЯ АКАДЕМИЯ НАУК УКРАИНЫ
ФИЗИКО-ХИМИЧЕСКИЙ ИНСТИТУТ им. А. В. Богатского

На правах рукописи

МОКШИНА ЕЛЕНА ГЕОРГИЕВНА



УДК 547.898:544.16

**«КРИТИЧЕСКИЕ СВОЙСТВА И ВТОРЫЕ ВИРИАЛЬНЫЕ
КОЭФФИЦИЕНТЫ ОРГАНИЧЕСКИХ ВЕЩЕСТВ И ИХ СМЕСЕЙ. QSPR
АНАЛИЗ С УЧЕТОМ МЕЖМОЛЕКУЛЯРНЫХ ВЗАИМОДЕЙСТВИЙ»**

02.00.03. – Органическая химия

**Диссертация
на соискание ученой степени
кандидата химических наук**

Научный руководитель:
Кузьмин Виктор Евгеньевич,
Доктор химических наук,
профессор.

Одесса – 2016

Оглавление

СПИСОК ОСНОВНЫХ СОКРАЩЕНИЙ	4
ВВЕДЕНИЕ	5
РАЗДЕЛ 1 СОВРЕМЕННЫЕ МЕТОДЫ РАСЧЕТЫ ТЕРМОДИНАМИЧЕСКИХ СВОЙСТВ (ОБЗОР ЛИТЕРАТУРЫ)	11
1.1. Методы прогнозирования критических свойств и фактора Питцера	13
1.01. Аддитивные методы.....	15
1.02. Термодинамические корреляции	21
1.03. Технология решения задач установления количественной зависимости «структура-свойство»	22
1.03.01. Основы QSPR.....	22
1.03.02. Симплексное описание молекулярной структуры индивидуальных органических веществ.....	25
1.03.03. Симплексное описание молекулярной структуры смесей.....	27
1.03.03. QSPR методы прогнозирования критических свойств.....	29
1.2. Методы прогнозирования вторых вириальных коэффициентов	32
РАЗДЕЛ 2. МЕТОДОЛОГИЯ ИССЛЕДОВАНИЯ ТЕРМОДИНАМИЧЕСКИХ СВОЙСТВ ОРГАНИЧЕСКИХ ВЕЩЕСТВ И ИХ БИНАРНЫХ СМЕСЕЙ В РАМКАХ QSPR ПОДХОДА НА ОСНОВЕ СПМС.....	38
2.1. Развитие симплексного представления молекулярной структуры.....	38
2.1.01. «Псевдосмесевой» подход	38
2.1.02. Парные межмолекулярные дескрипторы	40
2.1.03. Смеси и их дескрипторное описание	42
2.2. Технология решения задач QSPR	43
2.2.01. Отбор экспериментальных данных	43
2.2.02. Методы построения QSPR моделей	44
2.2.02.1. Структурная и физико-химическая интерпретация построенных моделей	47
2.2.03. Корреляционный и каузальный анализ.....	48
РАЗДЕЛ 3 КОЛИЧЕСТВЕННАЯ ОЦЕНКА ВЛИЯНИЯ СТРУКТУРЫ НА ТЕРМОДИНАМИЧЕСКИЕ СВОЙСТВА ИНДИВИДУАЛЬНЫХ ОРГАНИЧЕСКИХ ВЕЩЕСТВ И ИХ БИНАРНЫХ СМЕСЕЙ.....	51

	3
3.1. Критические свойства индивидуальных соединений.....	51
3.1.01. Экспериментальные данные	51
3.1.02. Результаты и обсуждение.....	52
3.2. Использование «псевдосмесового» подхода для моделирования критических свойств.....	60
3.3. Критические свойства бинарных смесей	69
3.3.01. Данные.....	69
3.3.02. Результаты и обсуждение.....	70
3.4. Каузальный анализ фактора Питцера: влияние параметров формы и размера молекул	75
3.4.01. Структурные параметры.....	76
3.4.02. Результаты и обсуждение.....	76
3.5. Вторые вириальные коэффициенты индивидуальных веществ	85
3.5.01. Данные.....	85
3.5.02. Схема моделирования.....	86
3.5.03. Результаты и обсуждение.....	88
3.6. Вторые вириальные кросс-коэффициенты бинарных смесей органических соединений.....	96
3.6.01. Данные.....	96
3.6.02. Результаты и обсуждение.....	98
РАЗДЕЛ 4. ЭКСПЕРТНАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ КРИТИЧЕСКИХ СВОЙСТВ И ВТОРЫХ ВИРИАЛЬНЫХ КОЭФФИЦИЕНТОВ ОРГАНИЧЕСКИХ ВЕЩЕСТВ И ИХ БИНАРНЫХ СМЕСЕЙ.....	104
РАЗДЕЛ 5. АЛГОРИТМЫ И ПРОГРАММЫ, ИСПОЛЬЗОВАННЫЕ ДЛЯ РЕШЕНИЯ ЗАДАЧ QSPR.	107
5.1. “Lattice&Simplex Modelling”	107
5.2. Открытая программная среда вычислений R.....	109
5.3. Python, OpenBabel, pybel.....	110
ВЫВОДЫ	112
ЛИТЕРАТУРА	115

СПИСОК ОСНОВНЫХ СОКРАЩЕНИЙ

МС – молекулярная структура
СПМС – симплексное представление молекулярной структуры
 T_c – критическая температура
 P_c – критическое давление
 V_c – критический объем
 ω – фактор Питцера
 B – второй вириальный коэффициент
GLM – метод обобщенного линейного моделирования
PLS – метод частичных наименьших квадратов
Random Forest – метод случайного леса
SVM – метод опорных векторов
GBM – метода градиентного бустинга
 R^2 – коэффициент детерминации
RMSE – стандартная ошибка прогноза
MAPE – средняя абсолютная процентная ошибка
AAD – среднее абсолютное отклонение
RMS – относительная стандартная ошибка прогноза

ВВЕДЕНИЕ¹

Актуальность темы. Термодинамические свойства органических соединений, в частности параметров критической точки, фактора Питцера (ω) и вторых вириальных коэффициентов, являются фундаментальными свойствами, на основе которых проводят расчеты при внедрении и проектировании энерго-химико-технологических систем.

Экспериментальные методы определения данных параметров отличаются высокой трудоемкостью (жесткие требования к проведению анализов, высокая чистота реагентов, сложность фиксирования термодинамического равновесия и т.д.) и себестоимостью. Кроме того, эти методы неприменимы к большим молекулам. Свойства же бинарных смесей являются определяющими при моделировании фазовых равновесий, поскольку именно бинарные смеси – наиболее широко применимые рабочие среды.

Большинство известных термодинамических методов прогнозирования критических свойств могут применяться к соединениям одного класса или требуют дополнительных экспериментальных данных. Существующие QSPR-модели (Quantitative Structure-Property Relationships – количественные соотношения «структура-свойство»), что описывают критические свойства, разработаны с использованием небольших выборок родственных соединений, и их польза ограничена. Небольшое число исследований, в которых QSPR моделирование проводилось для широкого диапазона органических соединений, использует 3D модели молекул и характеризуется крайне трудоемкими численными методами (например, конформационный поиск с использованием квантово-химических методов). Практически все из существующих работ по QSPR для индивидуальных соединений используют дескрипторы, которые сложно или невозможно интерпретировать. Что же касается QSPR моделирования критических свойств смесей, то более ранних работ не существует, поскольку выбор смесевых дескрипторов крайне ограничен. Для вторых вириальных коэффициентов как

¹ В руководстве диссертацией принимал участие профессор, д.т.н. Недоступ Вадим Иоаннович

индивидуальных соединений, так и смесей QSPR моделей до сегодняшнего дня не существовало.

2D QSPR модели на основе симплексного представления молекулярной структуры имеют ряд преимуществ над аналогами. Наличие разработанной системы смесевых дескрипторов позволяет построить модели для термодинамических свойств смесей органических соединений. Создание адекватных моделей для прогнозирования исследуемых свойств органических соединений разных классов и их бинарных смесей на основе структурной информации является, таким образом, актуальной задачей. Исследуемые в работе свойства и, в первую очередь, критические свойства, непосредственно связаны с энергией межмолекулярных взаимодействий. Это позволяет через оценку влияния структурных параметров на термодинамические характеристики выявить и проинтерпретировать структурные зависимости энергии межмолекулярных взаимодействий. Результаты моделирования также позволяют заменить длительные, не всегда эффективные, неэкономичные расчеты на новый более эффективный подход; выявить и проинтерпретировать структурные зависимости для критических свойств и вторых вириальных коэффициентов.

Связь работы с научными программами, планами, темами. Данная работа выполнялась в соответствии с планами ведомственной тематики отдела молекулярной структуры и хемоинформатики ФХИ им. А.В. Богатского НАН України «Экспертная система прогнозирования критических параметров органических соединений и их бинарных смесей» (2015-2016, № государственной регистрации 015U001060). Часть работы, посвященная каузальному анализу, выполнялась в рамках проекта NanoBridges (People Programme (Marie Curie Actions) of the European Union Seventh Framework Programme FP7/2007-2013/ under REA grant agreement no 295128").

Анализ влияния структуры органических соединений различных классов и состава их бинарных смесей на критические свойства, фактор Питцера и вторых

вириальных коэффициентов. Для достижения указанной цели необходимо было решить следующие задачи:

1. Сбор, анализ и верификация экспериментальной информации о критических свойствах, ω и вторых вириальных коэффициентах органических соединений. Формирование соответствующих баз данных.
2. Создание системы дескрипторов для межмолекулярных взаимодействий в индивидуальных органических веществах и их смесях.
3. Разработка и валидация QSPR моделей на основе симплексного подхода и дескрипторов межмолекулярных взаимодействий для исследуемых термодинамических свойств органических веществ и их бинарных смесей.
4. Определение структурных и физико-химических факторов, которые обуславливают исследуемые свойства.
5. Создание экспертной системы виртуального скрининга на основе построенных QSPR моделей.

Объект исследования. Системы структурного описания органических соединений и их бинарных смесей и модели для прогнозирования критических свойств, фактора Питцера и вторых вириальных коэффициентов. Структурные и физико-химические факторы, которые влияют на исследуемые свойства.

Предмет исследования. Органические соединения различных классов и их бинарные смеси.

Методы исследования. Генерация 2D структурных дескрипторов была выполнена методом симплексного представления молекулярной структуры, «псевдосмесевым» методом и методом парных дескрипторов взаимодействий. Для предварительной обработки данных использовались подходы т.н. «data curation». Зависимости «структура-свойство» строились с помощью методов машинного обучения – метода обобщенного линейного моделирования (GLM), метода частичных наименьших квадратов (PLS), метода случайного леса (Random Forest),

метода опорных векторов (SVM) и метода градиентного бустинга (GBM). Валидация моделей проводилась с помощью 5-кратной внешней кросс-валидации и (при возможности) внешней тестовой выборки.

Научная новизна полученных результатов.

1. Разработаны новые QSPR модели, которые прогнозируют важные термодинамические свойства индивидуальных органических веществ – критические параметры, ω и вторых вириальных коэффициентов для органических соединений различных классов.
2. Впервые на основе симплексного представления молекулярной структуры разработаны подходы, которые более глубоко отображают природу взаимодействия исследуемых веществ. Разработаны новые дескрипторы – «псевдосмесевые», которые моделируют структуру индивидуального вещества, как смеси двух идентичных соединений, и парные межмолекулярные дескрипторы, которые базируются на параметрах атом-атомных потенциалов. Оба типа дескрипторов позволяют учитывать возможные межмолекулярные взаимодействия.
3. Впервые проведен каузальный анализ зависимости ω от характеристик формы и размера молекулы.
4. Использован новый «двухслойный» подход к QSPR моделированию и вторых вириальных коэффициентов, в котором используется теоретическая зависимость и вторых вириальных коэффициентов от температуры.
5. Впервые построены QSPR модели для критических свойств и вторых вириальных коэффициентов бинарных смесей.
6. Показано влияние физико-химических и структурных факторов на исследуемые свойства.

Практическое значение полученных результатов. Полученные в работе результаты определяют пути целенаправленного поиска новых технических агентов на основе индивидуальных органических соединений и их смесей, в частности, теплоносителей, хладагентов и т.д. Полученные модели могут быть полезны при

вычислении параметров различного рода химико-технологических процессов. Разработанная на их основе экспертная система может использоваться для оценки критических свойств, ω и вторых вириальных коэффициентов органических веществ и их смесей. Экспертная система является открытой к использованию и доступной широкому кругу специалистов. Полученные результаты могут быть использованы при преподавании спецкурсов для студентов специальности «Органическая химия

Личный вклад соискателя. Самостоятельно разработана система дескрипторов для более детального описания межмолекулярных взаимодействий на 2D-уровне описания молекулярной структуры, реализована генерация дескрипторов в виде скриптов, построены QSPR модели и проведена их валидация, определено влияния структурных и физико-химических факторов на исследуемые свойства. В формировании заданий исследования, интерпретации, а также подготовке статей принимали участие научный руководитель и профессор, д.т.н. В.И. Недоступ. Часть исследования была исполнена в соавторстве с к.х.н. П.Г. Полищуком. Академик НАН Украины Г.Л. Камалов высказал полезные советы касательно оформления диссертации.

Всем вышеупомянутым коллегам автор искренне благодарна за помощь и сотрудничество.

Апробация результатов диссертации. Основные результаты работы докладывались и обсуждались на XIV, XVI и XVII конференціях молодых вчених та студентів-хіміків Південного регіону України (Одеса, 2012, 2014, 2015) ; IV и VII Всеукраїнських наукових конференціях студентів і аспірантів «Хімічні Каразінські Читання» (Харків, 2012 и 2015); 3rd Strasbourg Summer School in Chemoinformatics (2012); 5th International Symposium: Methods and Applications of Computational Chemistry (Харків, 2013); I и II Kazan Summer School on Chemoinformatics (Казань, 2013 и 2015); XVIII Scientific Youth Conference «Problems and Achievements of Modern Chemistry» (Одеса, 2016).

Публикации. По материалам диссертации опубликовано 7 статей, 10 тезисов докладов.

РАЗДЕЛ 1

СОВРЕМЕННЫЕ МЕТОДЫ РАСЧЕТЫ ТЕРМОДИНАМИЧЕСКИХ СВОЙСТВ

(ОБЗОР ЛИТЕРАТУРЫ)

Несмотря на то, что теоретический аппарат термодинамики является одним из наиболее детально разработанных, существует большое число практических приложений, для которых необходимо предварительное знание термодинамических свойств, которые не были ранее измерены и не могут быть достаточно точно предсказаны существующими теоретическими методами. Наблюдается снижение скорости накопления новых экспериментальных данных и снижение популярности массивных экспериментальных исследований, с одной стороны, и возрастающая потребность в точных данных, с другой. Происходит быстрая смена концепции «ограниченных данных», для которой характерны трудоемкие дорогостоящие эксперименты и имитационное моделирование, концепцией «избыточности данных», в условиях которой классические расчетные подходы неспособны удовлетворить растущую потребность в данных, ввиду своей трудоемкости.

Знание базовых термодинамических свойств необходимо для расчета различного рода эмпирических и полуэмпирических констант, в том числе для уравнений состояния. А применение уравнений состояния в свою очередь позволяет описывать свойства различных веществ в широком диапазоне условий.

Поскольку такие термодинамические свойства как критические температура и давление, критический объем и фактор Питцера, а также вторые вириальные коэффициенты, являются опорными свойствами при исследованиях фазового поведения веществ, существует необходимость в разработке быстрых и точных методов прогноза этих свойств.

В настоящее время существует большое число традиционных расчетных методов, позволяющих оценить характеристические термодинамические константы

чистых веществ. Существующие методы расчета и прогноза термодинамических свойств можно разделить на следующие группы:

1) Методы групповых вкладов или аддитивные методы. Общая черта этих методов – аддитивное представление молекул в виде некоторой совокупности молекулярных фрагментов. Такое представление позволяет достаточно приближенно оценить свойства сложных молекул, исходя из того, что группа атомов в молекуле определенным образом вносит определенный вклад в исследуемое свойство, независимо от природы других присутствующих групп атомов.

2) Эмпирические термодинамические корреляции – в основном имеют вид простых температурных зависимостей и позволяют предсказывать нужное свойство для ограниченного числа классов. Основной недостаток такого рода подходов, что они имеют частный характер и, в основном разрабатываются для каждого отдельного вещества индивидуально.

Эмпирические термодинамические корреляции основаны на факторном анализе, который позволяет строить корреляционные уравнения исходя из данных для других измеримых макроскопических свойств, таких как плотности, молекулярная масса, температура кипения и т.д. То есть они зависят от знания других характеристических констант.

К данной группе также стоит отнести методы, которые базируются на уравнениях состояния, в основе которых лежат константы, рассчитываемые из эмпирических корреляций.

3) Методы QSPR. Являясь, по сути, методами типа «черный ящик», не дают представлений о деталях протекания процессов, однако позволяют изучать большое количество молекул, явно учитывают влияние молекулярной структуры на исследуемое свойство и позволяют выделить главные влияющие факторы. Кроме того, отличаются вычислительной быстротой, легкостью в использовании итоговых моделей и строгой системой валидации и оценки прогностической способности моделей.

друг друга с учетом их интенсивных свойств. Однако, в общем, для смесей это состояние не соответствует ни наивысшему давлению, ни наивысшей температуре, при которой две фазы могут сосуществовать.

Основное отличие критических свойств бинарных смесей от критических свойств индивидуальных соединений в том, что для смесей наблюдается зависимость от состава. Таким образом, для смеси двух соединений при различных составах наблюдается ряд критических точек, для каждого состава – одна неинвариантная критическая точка.

Критические точки [4] бинарных смесей найденные экспериментально традиционно помещались в две различные категории. Для температур намного ниже критической температуры более летучего компонента наблюдается т.н. критическая точка «жидкость-жидкость», которая при представлении в виде функции от состава и проецировании в p, T -координатах, обычно дает locus практически параллельный оси p . Второй тип наблюдается, когда при температурах близких к критическим температурам обоих компонентов и проецировании в пространство p, V, T -координат, locus плавно проходит между критическими точками «газ-жидкость» двух индивидуальных компонентов.

Не существует абсолютного различия между критическими точками «жидкость-жидкость» и «газ-жидкость» для бинарных смесей, поскольку они обе ограничены пределами диффузионной стабильности. Некоторые системы обладают критической точкой промежуточной между критическими точками, описанными выше.

Важно также указать некоторые феномены, встречающиеся в смесях и нехарактерные для индивидуальных соединений. Около критической точки для смеси возможно проявление необычных свойств. Например, представим смесь, нагреваемую при постоянном давлении вдоль пути. При пересечении точки кипения и переходе в двухфазную область возникает паровая фаза, которая при нагревании обретает максимальный объем и затем уменьшается при дальнейшем нагревании, пока смесь опять полностью не конденсируется. Это явление известно как ретроградное испарение. Обратное явление – ретроградная конденсация происходит

в обратном направлении. При вхождении в двухфазную область возникает жидкая фаза, доля которой при увеличении давления увеличивается, затем уменьшается до тех пор, пока вся смесь не испарится полностью. Строго говоря, термины пар и жидкость не являются значимыми, т.к. состояние в точке D будет называться жидкостью в случае первого пути и паром в случае второго пути.

Другой интересный феномен, называемый баротропным эффектом, также наблюдается для некоторых смесей. В тех случаях, когда смесь сжимается изотермически, газовая фаза становится более плотной чем жидкая фаза и опускается на дно экспериментального сосуда. Такого рода инверсия фаз впервые была найдена Onnes [4] для системы гелий + водород. Смеси азот + аммиак [4], некоторые смеси фторзамещенных углеводородов с углеводородами [4] также демонстрируют этот эффект.

1.01. Аддитивные методы

Методы прогнозирования критических свойств и фактора Питцера индивидуальных органических соединений

Одними из наиболее популярных аддитивных методов за счет простоты расчетов и большого числа групп долгое время были метод Формана и Тодоса [5, 6] и метод Лидерсена [7]. Каждый из этих методов представляет собой простой аддитивный метод групповых составляющих. Метод Формана и Тодоса можно считать одним из первых интегрированных подходов, поскольку в качестве дескрипторов молекулярной структуры там используется один из простейших топологических индексов – индекс Виннера. Широко используемыми методами являются методы Joback (1984; 1987) [8], Constantinou и Gani (1994) [9], Wilson и Jasperson [10] (1996), и Marrero и Pardillo [11] (1999).

Помимо вышеупомянутых методов существует множество других аддитивных методов для критических свойств. Однако внимательное их изучение показывает, что они чаще всего ограничены конкретным классом веществ, например,

насыщенными углеводородами, перфторированными производными, спиртами и т.д., либо они отличаются меньшей расчетной точностью, чем описанные. В качестве некоторых примеров можно привести методы Sastri, и др. (1997); Tobler (1996) [12] и Daubert [13, 14].

Метод Joback

Joback (1984; 1987) [8] фактически обновил метод групповых вкладов Лидерсена, добавив несколько новых функциональных групп и определив новые значения вкладов.

$$T_c = T_b \left[0.584 + 0.965 \left\{ \sum_k N_k(tck) \right\} - \left\{ \sum_k N_k(tck) \right\}^2 \right]^{-1}$$

$$P_c = \left[0.113 + 0.0032 N_{atoms} - \sum_k N_k(pck) \right]^{-2}$$

$$V_c = 17.5 + \sum_k N_k(vck)$$

Где вклады групп обозначены как tck , pck и vck для критических температур, критических давлений и критических объемов соответственно; N_{atoms} - число атомов в молекуле; N_k – число групп определенного типа. Для расчета критической температуры необходимы данные по температурам кипения. Примерами групп в данном методе являются $—CH_3$, $-COH$, $-O-$ и другие.

Метод Constantinou and Gani (CG)

Constantinou и Gani [9] (1994) разработали расширенный и уточненный метод, основанный на группах UNIFAC, который позволяет описать нужные свойства с помощью более точных функций, а также использует фрагментные вклады второго

порядка. Более точные функции позволяют повысить корреляции, а приближение второго порядка – преодолеть ограничения UNIFAC, который не в состоянии различать изомеры, резонансные структуры и т.д. в пределах приближения первого порядка.

Общая формулировка CG функции $f[F]$ свойства F :

$$F = f \left[\sum_k N_k (F_{1k}) + W \sum_j M_j (F_{2j}) \right]$$

, где f может быть как линейной, так и нелинейной функцией, N_k – число групп первого порядка типа k в молекуле; F_{1k} – вклад группы первого порядка $1k$ в свойство, F ; M_j – число групп второго порядка типа j в молекуле; и F_{2j} – вклад группы второго порядка $2j$ в свойство, F . Значение W равно нулю для вычислений первого порядка и единице для вычислений второго порядка.

Для критических свойств формулировка CG выглядит следующим образом:

$$T_c = 181.128 \ln \left[\sum_k N_k (tc1k) + W \sum_j M_j (tc2j) \right]$$

$$P_c = \left[\sum_k N_k (pc1k) + W \sum_j M_j (pc2j) + 0.10022 \right]^{-2} + 1.3705$$

$$V_c = -0.00435 + \left[\sum_k N_k (vc1k) + W \sum_j M_j (vc2j) \right]$$

, где $tc1k$, $pc1k$, $vc1k$ - вклады для группы k для критических температур, критических давлений, критических объемов соответственно; $tc2j$, $pc2j$, $vc2j$ - вклады для группы j для критических температур, критических давлений, критических объемов соответственно; N_k - число групп типа k ; M_j - число групп типа j .

Метод Constantinou/Gani может быть чрезвычайно надежен для всех критических свойств, несмотря на то, что для малых молекул могут наблюдаться значительные ошибки. Стоит заметить, что расчет критической температуры в

данном методе не требует знания температуры кипения. Это легко объясняется тем, что в рамках аддитивного подхода чаще всего невозможно точно предсказать малые молекулы, даже если они могут быть составлены из имеющихся групп [15]. В общем, наибольшие ошибки метода CG наблюдаются для наименьших и наибольших молекул, в особенности фторированных и циклических соединений. Также метод CG является единственным из методов, позволяющим наряду с критическими свойствами рассчитывать фактор Питцера. Основное соотношение:

$$\omega = 0.4085 \left\{ \ln \left[\sum_k N_k (w1k) + W \sum_j M_j (w2j) + 1.1507 \right] \right\}^{\left(\frac{1}{0.5050} \right)}$$

Где w_{1k} и w_{2j} – вклады для групп первого и второго порядка соответственно.

Примерами групп первого порядка являются $-\text{CH}_3$, $-\text{CH}_2$, $-\text{O}-$ и т.д.; примерами групп второго порядка – $(\text{CH}_3)_2\text{CH}-$, CHOH , $(\text{CH}_3)_3\text{C}-$ и т.д.

Метод Wilson and Jasperson

Wilson и Jasperson [10](1996) сообщили о трех методах для определения критических температур и давлений, применимых как для органических, так и для неорганических соединений. Метод нулевого порядка использует факторный анализ; плотность, температура кипения и молекулярный вес выступают в качестве дескрипторов. Метод первого порядка использует атомные вклады совместно с температурой кипения и числом циклов, в то время как метод второго порядка включает в себя групповые вклады. В методах первого и второго порядков используются следующие уравнения:

$$T_c = T_b / \left[0.048271 - 0.019846 N_r + \sum_k N_k (\Delta tck) + \sum_j M_j (\Delta tcj) \right]^{0.2}$$

$$P_c = 0.0186233 T_c / [-0.96601 + \exp(Y)]$$

$$Y = -0.00922295 - 0.0290403 N_r + 0.041 \left(\sum_k N_k (\Delta pck) + \sum_j M_j (\Delta pcj) \right)$$

Где T_b – температура кипения, Nr - число циклов в соединении, Nk - число атомов типа k с атомными вкладами первого порядка Δtck и Δpck ; Mj – число групп типа j с групповыми вкладами второго порядка Δtcj и Δpcj .

Примерами групп являются —ОН (С4 или меньше), —NH₂, >NH, >N— и т.д.

Метод Marrero and Pardillo

Marrero-Marejon и Pardillo-Fontdevila [11] описали метод для определения T_c , P_c и V_c , которые был назван ими техникой оценки вкладов групповых взаимодействий или метод вкладов связей. Стоит отметить, что для расчета T_c требуется T_b , но Marrero and Pardillo также разработан метод расчета T_b . Основные уравнения метода:

$$T_c = T_b / \left[0.5851 - 0.9286 \left(\sum_k N_k tcbk \right) - \left(\sum_k N_k tcbk \right)^2 \right]$$

$$P_c = \left[0.1285 - 0.0059 N_{atoms} - \sum_k N_k pcbk \right]^{-2}$$

$$V_c = 25.1 + \sum_k N_k vcbk$$

Где N_{atoms} – число атомов в соединении, Nk – число атомов типа k с вкладами $tcbk$, $pcbк$, и $vcbk$.

Метод Gharageizi

Из недавних работ, посвященных методу групповых вкладов, стоит отметить работу Gharagheizi и коллег [16], в которой использовался метод искусственных нейронных сетей в сочетании с методом групповых вкладов для расчета и оценки критических свойств индивидуальных веществ, включая критические давления, температуры, объемы и фактор Питцера. Авторами было исследовано около 1700 соединений различных классов. Были получены результаты, характеризующиеся

следующими средними абсолютными отклонениями (AAD): 1.1 % для критических давлений, 0.9 % для критических температур, 1.4 % для критических объемов и 3.7% для ацентрического фактора.

Методы прогнозирования критических свойств смесей

Для критических свойств (как и для всех рассматриваемых далее термодинамических свойств органических соединений) количество методов прогнозирования свойств смесей значительно меньше по сравнению с методами прогнозирования свойств индивидуальных соединений.

Существует ряд методов, позволяющих с использованием правил взвешивания по долям компонентов в смеси приближенно оценить критические параметры, которые в данном случае именуются «псевдокритическими». Однако такого рода методы имеют крайне высокий уровень ошибок, что позволяет использовать их только для грубых оценок параметров смеси.

Классическими методами групповых вкладов для критических свойств смесей являются методы Chueh и Prausnitz [17] (1967) и Li (1971), подробно описанные в [12]. Li и Kiran [18] разработали метод, основанный на групповых вкладах к факторам, описанным в простом методе Klincewicz и Reid [19]. Li and Kiran построили корреляцию для 41 системы, содержащих различные углеводороды, полярные производные и смеси углеводород-СО₂. Построенные корреляции были использованы для прогноза значений для 15 смесей. Для критических температур стандартное отклонение находилось в диапазоне от 1 до 20 К (или около 4%), что сопоставимо с методом Li [20]. Для критических объемов ошибки находились в диапазоне от 2 до 7%. Для критических давлений стандартные отклонения были от 6 до 17 бар (или около 15%). Метод Liu [21] позволяет рассчитать критические давления исходя из предварительно спрогнозированных критических температур и объемов. Его результаты для углеводородов сравнимы с результатами, которые могут быть получены с использованием уравнений состояния (Anselme и Teja

[22](1990)). Для быстрой оценки истинных критических объемов смеси также можно использовать метод Schick и Prausnitz [23](1968).

1.02. Термодинамические корреляции

Существует два основных типа уравнений, относящихся к прогнозу термодинамических свойств: уравнения состояния и термодинамические корреляции. Уравнения состояния представляют собой уравнения, связывающие термодинамические (макроскопические) параметры системы, такие, как объем, давление, температура, химический потенциал и т.д. Термодинамика предоставляет математические критерии для фазовой стабильности и критических точек индивидуальных компонентов и смесей. Для индивидуальных соединений критическая точка определяется из уравнения относящегося к изотермической сжимаемости, для смесей критерий включает в себя матрицы вторых и третьих частных производных функций энергии с учетом числа моль компонентов. В обзоре Heidemann [24] и работе Sadus [25] подробно рассматриваются кубические уравнения такие как модель Soave [26], метод Michelsen и Heidemann [27], Michelsen [28, 29], и Heidemann и Khalil [30]. Abu-Eishah с соавторами [31] реализовали модель прогнозирования критических свойств с использованием уравнения состояния Stryjek and Vera [32]. Из описанных классических термодинамических методов уравнения состояния наиболее надежны, метод Michelsen является наиболее эффективным для прогноза свойств смесей. Уравнения состояния точны, но сложны и непригодны для одновременного прогноза термодинамических свойств большого числа систем.

Термодинамические корреляции, в свою очередь, базируются на факторном анализе, с помощью которого устанавливаются зависимости между данными измерений других макроскопических свойств, таких как плотность, молекулярная масса, температура кипения и т.д. К этим методам можно отнести метод Klincewicz и Reid [19] и метод Vetere [33], применимые к многим классам веществ. Метод Somayajulu [34] применим только к алканам, но предлагает общую концепцию

расчета для гомологических серий. Как бы то ни было, этим методам присущи недостаточная точность, либо избыточная сложность, а также не хватает универсальности.

1.03. Технология решения задач установления количественной зависимости «структура-свойство»

1.03.01. Основы QSPR

В рамках “Quantitative Structure-Property Relationships” (QSPR) методологии рассматриваются методы оценки свойств, основанные на расчетных характеристиках молекулярной структуры – *молекулярных дескрипторах*. При QSPR анализе рассчитанные для каждой молекулы дескрипторы ставятся в соответствие проявляемому свойству. Затем полученный массив данных анализируется с помощью статистических методов. На данный момент существует большое количество методов QSPR, разработанных в ходе многолетних исследований [35-38], популярность которых, а также спектр решаемых с их помощью задач [39-41] возрастают из года в год.

Различные QSAR/QSPR модели отличаются, в основном, уровнем детализации описания молекулярной структуры (МС). В простейшем случае используется только информация, содержащаяся в брутто-формуле молекулы – 1D QSAR/QSPR. Очевидно, что с помощью таких подходов практически невозможно адекватно решать задачи “структура – активность”. По всей вероятности, эти подходы имеют только вспомогательную роль [42]. Весьма популярны 2D QSAR/QSPR модели [43, 44], которые используют информацию из структурной формулы молекулы, которая, по существу, представляет собой молекулярный граф. Работоспособность таких подходов обусловлена тем, что топологическая модель МС в неявном виде содержит в себе информацию о возможных конформациях исследуемого соединения. Широко распространены 3D QSAR/QSPR модели [45-47], в которых пространственная

структура молекулы учитывается в явном виде. Однако, выбор анализируемого конформера для каждой исследуемой молекулы, зачастую, случаен. Это, в частности, обусловлено тем, что для оптимизации геометрии рассматриваемого соединения используется произвольное начальное приближение. То есть, фактически, для каждой молекулы исследуется один из случайных локальных минимумов. Отметим, так же, что популярный CoMFA [48, 49] метод и его аналоги становятся малоэффективными в рядах структурно разнородных соединений из-за неоднозначности процедуры совмещения молекул обучающей выборки. Наиболее адекватно описание MC реализуется в 4D-QSAR моделях [50, 51]. Здесь, кроме трехмерного представления MC рассматривается не один конформер, а набор (четвертое измерение) наиболее устойчивых конформаций.

Невзирая на весьма существенные различия в алгоритмах, их общей методологической предпосылкой является предположение об объективном существовании связи между структурой вещества и его свойствами (активностью), т.е. некоторой функции $F: S \rightarrow P$, отображающей множество структур S в используемом представлении на множество значений свойств P , выраженных подходящим образом. Предполагается также, что найденная функция может в известных пределах быть экстраполирована на множество структур более обширное, чем то, которое использовалось для ее определения.

Следовательно, в каждом методе анализа связи между структурой и свойствами можно выделить три основные составляющие:

1. описание (измерение) исследуемого свойства Y ;
2. расчет структурных параметров (дескрипторов) X_j объекта;
3. построение математической модели, описывающей связь Y с X_j .

К третьей составляющей относятся различные методы определения вида связи структура-свойство. Среди них можно отметить наиболее популярные – обобщенную линейную регрессию (GLM) [52], метод частичных наименьших квадратов (PLS) [53, 54], метод опорных векторов (SVM) [55-57], метод случайного леса [58, 59], а также методы искусственных нейронных сетей [60-62].

Также общим в методах QSPR анализа является то, что все используемые разновидности молекулярных дескрипторов можно свести к трем типам:

- интегральные – отражают особенности структуры молекулы, как целого, например, ван-дер-ваальсовый объем, моменты инерции, дипольный момент и его составляющие вдоль осей координат, молярная рефракция, липофильность, параметры структурного подобия, различные квантово-химические характеристики (энергия наивысшей занятой и низшей свободной орбитали, потенциалы ионизации, различные индексы реакционной способности) [63, 64];
- локальные – отражают особенности структуры отдельных фрагментов молекулы, например, константы заместителей и физико-химические характеристики отдельных групп, заряды на атомах, липофильности, поляризуемости атомов и т.п. [65];
- полевые – отражают особенности воздействия молекулы на окружающее пространство, например, потенциал электростатического поля и поля липофильности, другие характеристики, отражающие особенности межмолекулярного взаимодействия. Могут быть как локальными, так и интегральными.

Все эти разновидности молекулярных дескрипторов могут быть получены из топологического или пространственного описания МС. Выбор того или иного способа описания структуры диктуется характером конкретной решаемой задачи и имеющимися ограничениями на доступность экспериментальных и расчетных данных.

При использовании топологического описания (для молекулы в целом ее первичное топологическое описание – это ее структурная формула) во внимание принимаются только типы атомов и связей и их последовательность, т.е. не рассматриваются расстояния между группировками и их взаимная ориентация. С математической точки зрения молекула в таком описании представляет собой граф [66] с мечеными вершинами (атомы разных типов) и мечеными ребрами (связи различной кратности) - молекулярный граф. В таком приближении наиболее часто в

виде дескрипторов выступают различного рода топологические индексы и дескрипторы молекулярных фрагментов [67].

В пространственном (3D) описании молекула представляется как совокупность атомов с известными координатами. Может также задаваться зависимость тех или иных физических свойств в точке от ее положения в той же системе координат. Современные квантово-химические методы и методы молекулярной механики позволяют с достаточной надежностью определять геометрию и энергетические характеристики молекул [68].

1.03.02. Симплексное описание молекулярной структуры индивидуальных органических веществ

Основным инструментом для решения задач в диссертационной работе было симплексное представление молекулярной структуры (СПМС) [50, 69-71], суть которого состоит в том, что каждая молекула представляется в виде системы различных симплексов – четырехатомных молекулярных фрагментов фиксированного состава и строения (табл. 1). Данный метод хорошо себя зарекомендовал при решении ряда различных QSAR/QSPR задач [71, 72].

Общее количество всех возможных симплексов N в n -атомной молекуле равно:

$$N = \frac{n!}{(n-4)!4!} \quad (2)$$

Молекулярная структура (МС) может быть описана на различных уровнях от 1D до 4D. На 1D уровне молекула представлена брутто-формулой, на 2D уровне – структурной формулой. В данной задаче использовались 1D и 2D-подходы. На 1D уровне учитывается лишь брутто-формула, симплекс является числом фиксированных четверок атомов. На 2D уровне учитываются связность атомов в

симплексе, тип атомов и природа связи (одинарная, двойная, тройная, ароматическая).

Таблица 1

Базовые типы молекулярных симплексов

Базовый тип	1	2	3	4	5	6	7	8	9	10	11
Фрагмент											
Пример											

Алгоритм расчета дескрипторов заключается в реализации следующих этапов:

1) Разбиение молекул на симплексы (наряду с симплексами возможно использование меньших фрагментов – пар, троек). Количество невзвешенных симплексов относится к топологическому типу;

2) Взвешивание вершин согласно различным физико-химическим свойствам (электроотрицательность, рефракция, параметры потенциала Леннарда-Джонса 6-12, частичные заряды на атомах [73], донор/акцептор водородной связи);

3) Разбиение весовых характеристик на диапазоны, присвоение каждому диапазону определенной метки, присвоение вершине метки согласно величине свойства, по которому она была взвешена. В настоящей работе атомам присвоены следующие метки (римские цифры обозначают тип вершины симплекса):

а) для частичных зарядов:

$$I \leq -0,1 < II \leq 0,05 < III \leq -0,01 < IV \leq 0,01 < V \leq 0,05 < VI \leq 0,1 < VII;$$

б) для рефракции:

$$I \leq 1,5 < II \leq 3 < III \leq 8 < IV;$$

в) способность к образованию водородной связи:

I – донор, II – акцептор, III – индифферентный центр.

г) кроме того, метки могут характеризовать тип атома (его природу или индивидуальность) в целом, например C_{sp3} , O_{sp2} и т.п.

4) Подсчет числа симплексов определенного типа. Фактически, количество симплексов определенного типа является дескриптором. Каждый дескриптор характеризует симплекс фиксированного типа. Количество дескрипторов может быть достаточно велико, в зависимости от структурного разнообразия исследуемых соединений. В качестве дескрипторов также были использованы интегральные параметры, такие как молекулярная масса, молекулярная рефракция и т.п.

1.03.03. Симплексное описание молекулярной структуры смесей

Смесевая модификация СПМС была разработана для QSAR/QSPR смесей произвольного состава [74]. Используются симплексные дескрипторы, характеризующие непосредственно смесь соединений. Главное отличие от классического симплексного (или любого другого) подхода состоит в том, что бинарная смесь представлена двумя молекулами одновременно. Генерируемые связанные дескрипторы описывают только один компонент смеси (компонент А или В), в то время как несвязные симплексы могут описывать оба компонента и смесь как единое целое. В связи с этим, необходимо отметить какие несвязные симплексы будут принадлежать одной и той же молекуле, а какие разным. В последнем случае такие несвязные симплексы отражают структуру не одной молекулы, а характеризуют пару разных молекул. При этом необходимо обозначить, принадлежат ли части несвязных симплексов к одной молекуле или к разным. В последнем случае, такие несвязные симплексы не будут отражать структуру единичной молекулы, но будут описывать пару разнородных молекул. Фактически, симплексы такого типа – это структурные дескрипторы смеси соединений (рис. 3). Для того, чтобы отличить такие «смесевые» симплексы от обычных, при генерации дескрипторов используются специальные метки. В данном подходе учитывается состав смеси, то есть дескрипторы отдельных компонентов (вещества А и В) взвешиваются согласно мольным долям компонентов в смеси, а дескрипторы смеси умножаются на минимальную долю одного из компонентов:

$$D = \begin{cases} [n_A\{D(A)\} + n_B\{D(B)\}]; \\ n_A\{D(A+B)\} \end{cases}$$

где D – дескрипторы смеси, $D(A)$ и $D(B)$ – симплексы индивидуальных молекул А и В,

$D(A+B)$ – симплексы смеси молекул А и В,

n_A и n_B – мольные доли компонентов А и В в смеси ($n_A < n_B$, $n_A + n_B = 1$).

Если в одной задаче изучаются и индивидуальные вещества, и их смеси, то дескрипторы индивидуальных веществ остаются неизменными.

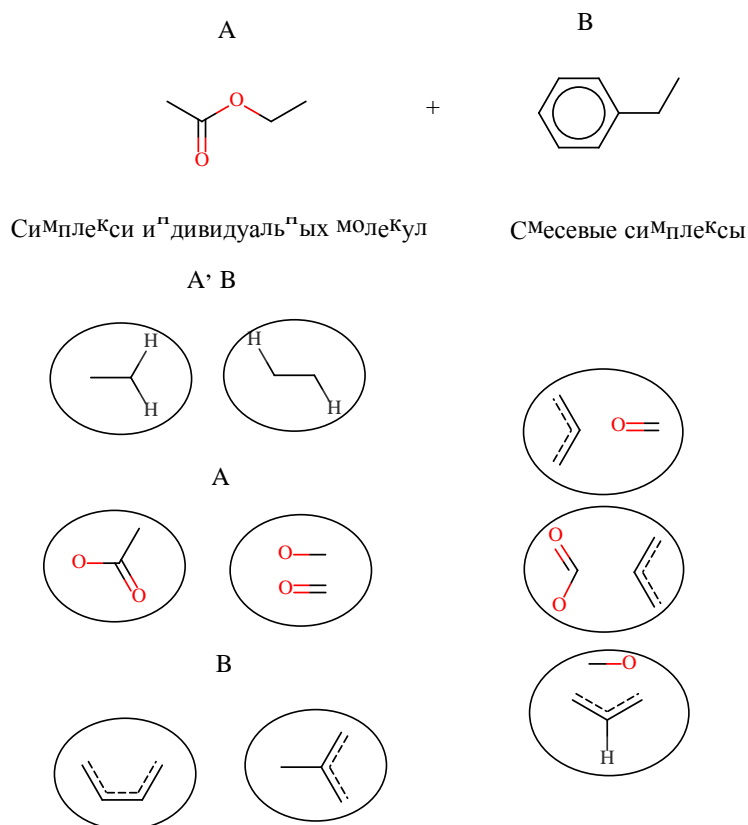


Рис. 2. Симплексное представление структуры смесей.

Состав смесей также учитывается, например, дескрипторы компонентов смеси (молекул A и B) взвешиваются в соответствии с их мольными долями, а количество смесевых дескрипторов умножается на удвоенную мольную долю меньшего по составу компонента. Таким образом, структура каждой смеси описывается одновременно и дескрипторами смеси как целого и дескрипторами ее индивидуальных компонентов.

1.03.03. QSPR методы прогнозирования критических свойств

Методы прогнозирования критических свойств, основанные на QSPR методологии, набирают все большую популярность. Можно выделить два типа моделей – локальные (способные прогнозировать критические параметры для соединений одного класса) и глобальные (использующие химически разнородные

выборки органических соединений). Локальные методы, например [75], имеют ограниченную область применения, не способны к генерализации и поэтому подробно обсуждаться не будут.

Memod Grigoros

Одной из первых попыток смоделировать критические свойства индивидуальных органических соединений была работа Grigoros [76]. В качестве основных дескрипторов были выбраны площадь поверхности молекул и электростатические параметры, полученные в результате квантово-механических расчетов. Примерами уравнений являются:

$$V_c = 2.217A - 93.0$$

$$T_c = 0.633A - 1.562A_- + 0.427A_+ + 9.914A_{HB} + 263.4$$

$$P_c = 2.9 + 20.2(T_c/V_c)$$

где A – площадь поверхности молекулы, A_- и A_+ – положительно и отрицательно заряженные доли площади поверхности молекулы и A_{HB} – площадь поверхности молекулы, участвующая в образовании водородной связи. В оригинальной работе было проведено моделирование и сравнение критических свойств для выборки из 137 соединений, включающей в себя соединения различных классов. Статистическая оценка показывает, что точность в данном случае крайне низка [15]. Поскольку единственным статистическим параметром, приведенным авторами, было качество аппроксимации, практически невозможно провести более детальную оценку результатов моделирования.

Memod Jurs

Jurs и соавторы построили ряд моделей для критических свойств [77, 78]. В отличие от работы Grigoros авторами использовались квантово-химические дескрипторы и более тщательный подход к статистическому моделированию. В частности, проводился целенаправленный отбор дескрипторов путем исключения

дескрипторов с низкой вариабельностью и высокими парными корреляциями. Например, в случае критических температур дескрипторами являлись дипольный момент, площадь поверхности, индекс связности, число вторичных углеродных атомов типа sp^3 , число атомов кислорода, гравитационный индекс, функция заряда атомов-акцепторов и средний положительный заряд на атомах углерода. Поскольку авторами не проводилась надежная валидация моделей, детальное сравнение результатов моделирования с другими подходами затруднено.

Memod Godavarthy

В работе Godavarthy и коллег [79] было проведено моделирование наибольшей выборки данных для органических соединений, в которой было представлено 1230 соединений различных классов. В рамках моделирования авторами были использованы топологические, геометрические и электростатические дескрипторы для 3D-структур. Для построения моделей использовалось как линейное моделирование, так и метод искусственных нейронных сетей. Полученные модели показали высокие статистические результаты, значения среднего абсолютного отклонения (AAD, %) составляют 0.9% для критических температур, 1.5% для критических давлений и 1.7% для критических объемов.

Несмотря на высокие характеристики полученных авторами моделей, данная работа имеет один существенный недостаток – не все использованные при моделировании значения являются экспериментальными, часть из них была спрогнозирована другими методами. Кроме того, авторами использовались трудоемкие методы расчета 3-Д дескрипторов.

Memod NIST. Исследовательской группой Национального Института Стандартов и Технологий (NIST) [80] был разработан ряд моделей для оценки и прогнозирования критических температур и давлений. Были использованы данные для более чем 850 соединений, особое внимание уделялось использованию исключительно экспериментальных значений. Для описания структуры соединений как квантово-химические, так и топологические и электростатические дескрипторы,

модели были построены с использованием метода опорных векторов. Полученные результаты отличаются большей точностью, чем референтные аддитивные методы, для большинства соединений ошибка прогноза составляет менее 3%.

Стоит отметить трудозатратность методологии построения моделей, применявшейся авторами – в рамках 3D-моделирования проводился конформационный поиск с использованием полуэмпирических квантово-химических методов. Кроме того, в ряде случаев, дескрипторы не всегда имеют ясную интерпретацию (например, площадь XY проекции).

1.2. Методы прогнозирования вторых вириальных коэффициентов

Как уже упоминалось, поведение веществ на всей фазовой поверхности может быть описано с использованием уравнений состояния. Тем не менее, даже наиболее простые уравнения состояния в основном носят эмпирический и полуэмпирический характер. Среди огромного разнообразия уравнений состояния, описывающих p - V - T поведение реального газа, существует только одно, имеющее строгое теоретическое объяснение при всех давлениях кроме наивысших. Упомянутое уравнение состояния – так называемое вириальное уравнение, в котором фактор сжимаемости (Z) выражается как разложение по плотности или давлению.

$$Z = pV_m/RT \quad (1)$$

$$pV_m/RT = 1 + B/V_m + C/V_m^2 + D/V_m^3 + \dots \quad (2)$$

Наиболее важно то, что вириальные коэффициенты напрямую связаны с межмолекулярными взаимодействиями и отклонениями поведения газа от идеального. Второй вириальный коэффициент описывает отклонение от идеальности связанное с парными взаимодействиями между молекулами. Так как парные взаимодействия статистически являются наиболее значимыми в любой системе, второй вириальный коэффициент наиболее важен. При всех температурах кроме наивысших второй вириальный коэффициент отрицателен. Третий

вириальный коэффициент описывает тройные взаимодействия и, следовательно, менее влиятелен.

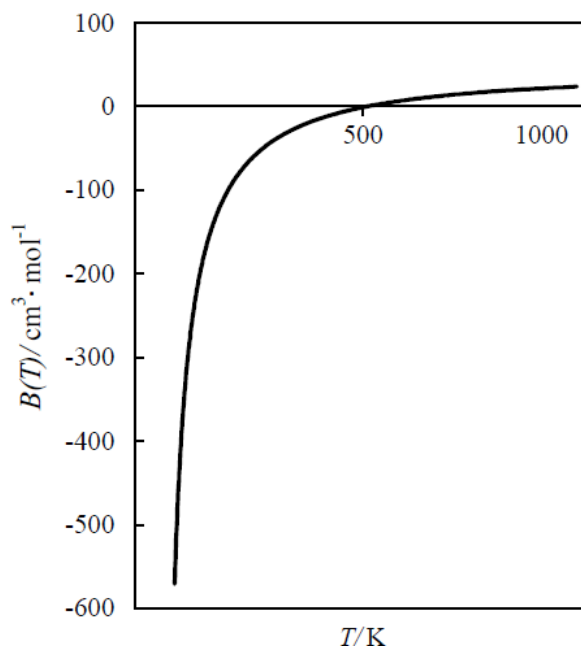


Рис. 3. Второй вириальный коэффициент метана.

Вириальное уравнение состояния для смесей является функцией не только от температуры T , но и от состава смеси, что видно из следующей серии уравнений.

$$Z = \frac{pV_m}{RT} = 1 + B_{mix}/V_m + C_{mix}/V_m^2 + \dots$$

Где V_m – молярный объем, T – абсолютная температура, R – газовая постоянная. B_{mix} , C_{mix} в разложении по плотности (или объему) определяются как второй, третий, четвертый и т.д. смесевые вириальные коэффициенты.

Коэффициенты вышеприведенных двух уравнений соотносятся между собой следующим образом:

$$B_{mix} = B_{mix}^* RT$$

Для бинарной смеси второй вириальный коэффициент может быть записан следующим образом:

$$B_{mix} = x_1^2 B_1 + x_2^2 B_2 + 2x_1 x_2 B_{12}$$

Где B_1 и B_2 – вторые вириальные коэффициенты чистых компонентов. B_{12} определяется как второй вириальный коэффициент взаимодействия, вириальный кросс-коэффициент или смешанный вириальный коэффициент. Он может быть отнесен к функции потенциальной энергии $U(R_{12})$, которая описывает взаимодействие молекул типа 1 с молекулами типа 2 в терминах молекулярного расстояния R_{12} , x_1 и x_2 – мольные доли компонента 1 и компонента 2 соответственно. Например, для случая взаимодействия молекул со сферически-симметричными функциями потенциальной энергии:

$$B_{12} = -2\pi N_A \int_0^\infty (e^{-U(R_{12})/kT} - 1) R_{12}^2 dR_{12}$$

Где N_A – число Авогадро.

Второй вириальный кросс-коэффициент является расчетным свойством, уникальность которого состоит в том, что он непосредственно и строго связан с гетерогенными парными взаимодействиями компонентов в смеси. В отличие от других свойств смесей второй вириальный кросс-коэффициент смеси молекул типа 1 с молекулами типа 2 характеризует только взаимодействия молекул типа 1 с молекулами типа 2. Поскольку свойства смесей в основном зависят именно от парных взаимодействий, второй вириальный кросс-коэффициент дает возможность прогнозировать pVT –свойства не только для бинарных, но также для многокомпонентных смесей.

Методы прогнозирования вторых вириальных коэффициентов индивидуальных органических соединений

Вириальный коэффициент – расчетный параметр, но экспериментальные подходы к получению исходных данных для его расчета сложны, время- и трудозатратны. Существующие теоретические методы расчета вторых вириальных коэффициентов являются методами, основанными на принципе соответственных состояний [81]. Из методов, имеющих практическое применение, для индивидуальных соединений стоит упомянуть следующие (Tsonopoulos [82], Hayden

и O'Connell [83], Tarakad и Danner [84], McCann и Danner [85], Orbey [86], Kis и Orbey [87]; Abusleme и Vera [88], Lee и Chen [89], Vetere [90]).

По существу все вышеуказанные методы могут быть записаны в следующей форме:

$$\frac{B(T)}{V^*} = \sum_i a_i f^{(i)}(T/T^*)$$

где V^* - характеристический объем, такой как V_c или P_c / RT_c , a_i – силовой параметр для различных межмолекулярных взаимодействий и $f(i)$ – наборы универсальных функций приведенной температуры T/T^* (T^* обычно является T_c). Таким образом, $f(0)$ – для простых сферических молекул, $f(1)$ вводит поправку на несферичность, $f(2)$ учитывает полярность (a_2 является функцией дипольного момента) и $f(3)$ учитывает ассоциацию молекул с помощью эмпирического параметра a_3 .

Немного иной подход к приведенным функциям используется в методах Hayden-O'Connell [83] и Abusleme-Vera [88]. Главное отличие метода Hayden-O'Connell состоит в том что: 1) Полярные эффекты и эффекты ассоциации не описываются в рамках критических свойств, и для описания сферичности используется ацентрический фактор, и 2) метод применяется напрямую к карбоновым кислотам путем отнесения константы равновесия димеризации к полученному второму вириальному коэффициенту в соответствии с «химической теорией» поведения неидеального газа.

Согласно указаниям NIST [91] единственный рекомендованный метод прогноза был разработан Xiang и коллегами [92]. Метод Xiang использует модифицированный принцип соответственных состояний, который требует в качестве входных параметров только значения критических параметров и ацентрического фактора. Для описания поведения соединений, чьи параметры силового поля сильно отклоняются от сферической симметрии используется дополнительный параметр. Дополнительный параметр определен в терминах отклонения критического фактора сжатия реальной молекулы от аналогичной сферической молекулы и является независимым от экспериментальных данных для любого заданного свойства.

Сравнение рассмотренных методов [15] показывает, что ни один из них не отличается значительно от других по точности. Все методы дают среднее отклонение в диапазоне от 10 до 100 см³/моль в зависимости от класса соединений. Некоторые методы дают более точные результаты для определенных классов, тем не менее никакой последовательности не наблюдается.

Методы прогнозирования вторых вириальных коэффициентов бинарных смесей органических соединений

Подход к кросс-коэффициентам смесей практически не отличается от подхода для индивидуальных соединений. Формулировка принципа соответственных состояний для всех пар компонентов в смеси i и j :

$$\frac{B_{ij}(T)}{V_{ij}^*} = \sum_m a_{mij} f^{(m)}(T/T_{ij}^*)$$

Где V^* - характеристический объем для пары, a_{mij} – силовые параметры для различных пар межмолекулярных взаимодействий, и $f^{(m)}$ – набор универсальных функций приведенной температуры, где T_{ij}^* – характеристическая температура для пары. Таким образом $f^{(0)}$ соответствует простым соединениям, $f^{(1)}$ вводит поправки на несферичность. Если один или оба компонента полярны, $f^{(2)}$ вводит поправку на полярность. И, как и для индивидуальных соединений, $f^{(3)}$ учитывает ассоциацию.

Для смешанных вириальных коэффициентов большинство методов использует комбинационные правила для T^* с постоянным бинарным параметром взаимодействия с k_{ij} . Существуют методы, которые не учитывают вклад полярных/ассоциационных взаимодействий для пар неполярное-полярное соединение, а также могут использовать эмпирические параметры для сольватации. Наиболее популярным и надежным из таких методов является метод Tsonopoulos [82].

Так же как и для индивидуальных соединений, ни один из перечисленных методов не обладает преимуществом перед другими, за исключением метода Hayden

и O'Connell (1975), который применим к системам, содержащим сильноассоциированные компоненты, такие как карбоновые кислоты.

Обобщая проведенный анализ, можно сделать вывод, что основным недостатком методов и подходов, базирующихся на принципе соответственных состояний является потребность в дополнительных экспериментальных данных (например, критических параметрах и ацентрическом факторе), а также использование дополнительных эмпирических подгоночных параметров.

РАЗДЕЛ 2.

МЕТОДОЛОГИЯ ИССЛЕДОВАНИЯ ТЕРМОДИНАМИЧЕСКИХ СВОЙСТВ ОРГАНИЧЕСКИХ ВЕЩЕСТВ И ИХ БИНАРНЫХ СМЕСЕЙ В РАМКАХ QSPR ПОДХОДА НА ОСНОВЕ СПМС

Исследование точных фундаментальных термодинамических свойств органических молекул и их смесей таких, как критические свойства, фактор Питцера и вторые вириальные коэффициенты, требует значительной модификации как методологии симплексного представления молекулярной структуры (см. раздел 1.03.02), так и стандартного процесса QSPR моделирования с целью подведения более строгой теоретической базы, согласующейся с теоретическими выводами и постулатами термодинамики.

2.1. Развитие симплексного представления молекулярной структуры.

Наиболее важным заданием из вышеперечисленных является модификация дескрипторного представления молекулярной структуры индивидуальных веществ и их смесей. В рамках исследования нами были разработаны два совершенно новых типа дескрипторов, направленные на более точное описание поведения органических соединений и более строгий и детальный учет межмолекулярных взаимодействий.

2.1.01. «Псевдосмесевой» подход

Особенности строения молекул в конденсированных фазах (например, жидкостях) позволяют представить фазу как макроскопическую «единую» молекулу с постоянно варьирующимся под влиянием теплового движения составом [93]. Кроме того, большинство изучаемых в данной работе термодинамических свойств непосредственно связаны с парными взаимодействиями между молекулами и энергией этих взаимодействий. Например, для ряда изученных молекул существует простое эмпирическое правило, связывающее критическую температуру, фактор Питцера и глубину потенциальной ямы [94]:

$$\frac{\varepsilon}{kT_c} = 0.7915 + 0.1693\omega$$

, где ε - глубина потенциальной ямы, k – постоянная Больцмана.

Следовательно, при QSPR моделировании также возникает необходимость описания несвязных взаимодействий не только в смесях, но и для индивидуальных соединений, более того описания их на 2D уровне представления молекулярной структуры. Основная идея предложенного нами подхода во многом перекликается с так называемыми «квази-химическими» термодинамическими методологиями [12].

В рамках «псевдосмесевое» подхода [95] индивидуальное органическое соединение представляется в качестве бинарной смеси двух идентичных молекул. Благодаря природе смесевых дескрипторов это не приводит к дублированию симплексов, но ведет к возникновению новых уникальных типов, характеризующих потенциальные межмолекулярные взаимодействия в индивидуальном веществе (рис. 6).

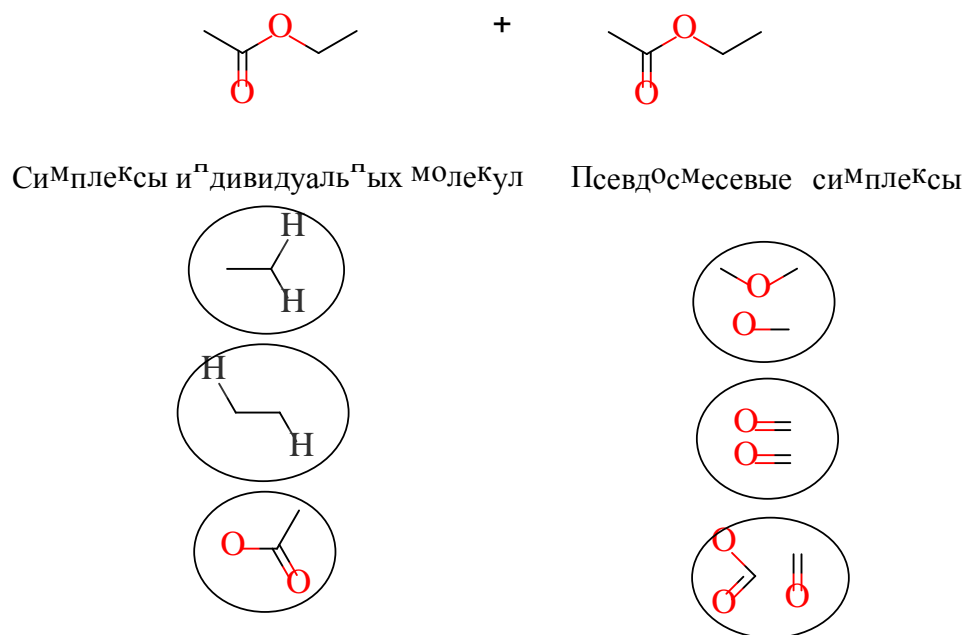


Рис. 4. Примеры симплексов индивидуального соединения (этилацетата) в «псевдосмесевом» подходе.

2.1.02. Парные межмолекулярные дескрипторы

Метод «атом-атом потенциалов» был предложен А.И. Китайгородским для неэмпирического расчета энергии решетки молекулярных кристаллов [96]. В рамках данного подхода каждый индивидуальный атом в молекуле рассматривается как некоторый силовой центр. Сумма энергий парных взаимодействий неких атомов i и j , принадлежащих разным молекулам, составляет энергию взаимодействия этих молекул. Энергии взаимодействия E_{ij} зависят только от типа атомов и могут быть выражены с помощью различных эмпирических потенциалов (например, потенциала Букингема, потенциала Штокмайера и т.д.), простейшим из которых и наиболее популярным в различного рода вычислительных задачах является потенциал Леннарда-Джонса.

Данный метод связан с расчетом потенциальной энергии молекул, которая выражается с помощью т.н. потенциальной функции или силового поля. Общая форма для полной энергии взаимодействия для аддитивного силового поля может быть записана как $E_{total} = E_{bonded} + E_{nonbonded}$, где ковалентный и нековалентный компоненты могут быть выражены следующими простыми уравнениями.

$$E_{bonded} = E_{bond} + E_{angle} + E_{dihedral}$$

$$E_{nonbonded} = E_{electrostatic} + E_{van\ der\ Waals}$$

В дополнение к функциональной форме конкретного потенциала, существует набор табулированных параметров для каждого типа атомов. Параметры обычно получают путем аппроксимации наборов экспериментальных данных.

Атомные параметры для нековалентных взаимодействий, а именно энергия ванн-дер-ваальсовых взаимодействий непосредственно связана с потенциалом. Двумя основными параметрами являются ϵ – глубина потенциальной ямы, и σ – межатомное вандерваальсово расстояние при минимуме потенциала.

Ранее использовавшаяся схема расчета вандерваальсовых взаимодействий для СПМС использовала следующее уравнение для взвешивания дескрипторов [97]:

$$E_{vdw} = Ae^{-Bx} - C_6/r_6$$

Где B – экспоненциальный член, описывающий отталкивание; C – дисперсионный член, описывающий притяжение.

Для присвоения метки атому использовались значения констант B и C для однородных взаимодействий, например, углерод-углерод и т.д. Таким образом, при взвешивании дескрипторов проводился учет только лишь однородных взаимодействий. Несмотря на то, что такой подход является результативным в некоторых случаях, он не позволяет рассчитать взаимодействия между разнородными атомами в случае симплексного подхода и является, по факту, грубым приближением.

Рассматривая фактическую картину взаимодействий в бинарной смеси (см. рис. 7), видно, что для бинарной смеси молекул типа 1 и молекул типа 2 существует три возможных типа парных взаимодействий (гетерогенных и гомогенных): 1-1, 1-2, 2-2. Для описания свойств исследуемой смеси абсолютно необходимо учитывать также и разнородные взаимодействия.

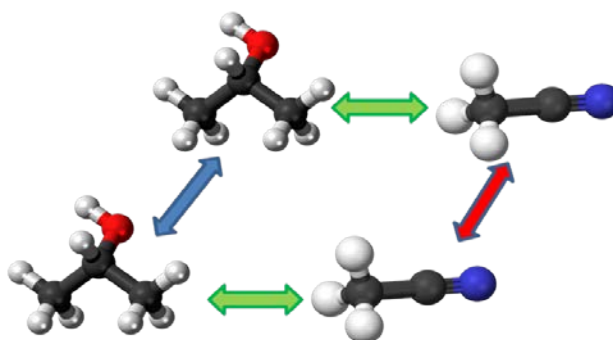


Рис. 5. Возможные парные взаимодействия в смеси пропан-2-ола и ацетонитрила.

Для описания такого рода взаимодействий нами была предложена схема расчета т.н. парных межмолекулярных дескрипторов [98]. Поскольку в рамках 2D-подхода не производится оптимизация структуры, а следовательно и межатомных расстояний, было принято решение для взвешивания дескрипторов использовать параметры потенциала, взвешенные согласно следующим правилам [97]:

$$\sigma_{ij} = \frac{1}{2(\sigma_i + \sigma_j)}$$

$$\varepsilon_{ij} = (\varepsilon_i \varepsilon_j)^{1/2}$$

, где ε – глубина минимума потенциала межмолекулярного взаимодействия; σ – расстояние, на котором энергия взаимодействия равна нулю.

Схема расчета парных межмолекулярных дескрипторов следующая. На первом шаге генерируются все возможные несвязные пары атомов ij , где i – первая молекула, j – вторая молекула (порядок молекул является незначимым). Затем для каждой пары атомов берутся рассчитанные взвешенные параметры потенциала (согласно выбранному силовому полю). На заключительном этапе взвешенные атом-атомные пары разбиваются на группы [A; B; C] согласно следующим интервалам – ε ($A \leq 0.1 < B \leq 0.3 < C$) и σ ($A \leq 3.5 < B \leq 4 < C$). Как и в СПМС число пар каждого типа и есть дескриптором. В данной работе для расчета парных межмолекулярных дескрипторов мы использовали универсальное силовое поле (UFF [97]), однако схема расчета позволяет использовать параметры и любого другого силового поля.

2.1.03. Смеси и их дескрипторное описание

Симплексное описание смесей претерпело незначительные изменения относительно [74, 99]. Прежде всего, для всех задач была модифицирована схема взвешивания дескрипторов. В рамках оригинальной схемы при учете мольных долей компонентов, схема взвешивания была следующей:

$$D = x_1 D_1 + 2x_2 D_{\text{смеси}} + x_2 D_2,$$

Где D – дескрипторы (смеси, компонента 1 и компонента 2 соответственно), x – мольные доли компонентов (мольная доля компонента 2 < мольной доли компонента 1).

Однако при рассмотрении бинарной смеси, одной из важнейших характеристик является количество возможных парных взаимодействий. В данном случае, как и для парных межмолекулярных потенциалов возможна реализация трех типов взаимодействий: 1-1, 1-2 и 2-2. Количество возможных парных взаимодействий в системе описывается биномом Ньютона, т.е. для системы компонент 1 – компонент 2 схема взвешивания будет выглядеть следующим образом:

$$D = x_1^2 D_1 + x_1 x_2 D_{\text{смеси}} + x_2^2 D_2$$

Поскольку такого рода взвешивание является теоретически обоснованным, было принято решение применять именно его при дальнейшем построении моделей.

Второе изменение стандартная схема смесевых дескрипторов претерпела при моделировании вторых вириальных коэффициентов для разнородных молекул. Как уже упоминалось ранее, второй вириальный коэффициент B_{12} описывает исключительно разнородные взаимодействия в смеси. Таким образом, информация о чистых компонентах в случае моделирования B_{12} органических бинарных смесей является избыточной и должна быть исключена из дальнейшего статистического моделирования.

2.2. Технология решения задач QSPR

2.2.01. Отбор экспериментальных данных

Большой массив накопленных экспериментальных данных по термодинамическим свойствам органических соединений и их бинарных смесей (в меньшей степени) позволяет составить представительные выборки, включающие информацию о большом числе соединений различных классов. Экспериментальные

данные для критических свойств индивидуальных органических соединений были взяты из наиболее полных справочников [100, 101]. Неполные или противоречивые значения критических свойств сверялись по открытой базе данных NIST WebBook[102]. Для критических свойств бинарных смесей были использованы данные из [4].

Данные по вторым вириальным коэффициентам были взяты из справочника серии Landolt-Börnstein [101], процедура проверки данных была проведена NIST. Процедура верификации данных, проводимая NIST, кратко описана ниже. Экспериментальные погрешности [101] данных были основаны на оценке информации, предоставленной авторами в оригинальных работах. Выбор данных был основан на сравнении экспериментальных погрешностей данных и распределении вириальных коэффициентов по температурам. Погрешность каждой точки данных в выборке сравнивалась с взвешенным средним всех других значений. Взвешивающий фактор являлся обратной экспоненциальной функцией разницы между температурой оцениваемого значения и температурами остальных значений. Основными параметрами, определявшими отбор данных, было количество точек для каждого соединения и температурный диапазон.

Дальнейшая проверка данных (или т.н. “data curation”), проведенная в работе, включала в себя поиск ошибочных структур, неправильных имен соединений, проверка структур на связность и уникальность, стандартизацию структур. Дубликаты были отсеяны с использованием и внутрилабораторного комплекса программ TheorChem. Собранные данные сохранялись и обрабатывались с использованием программы Instant JChem[103]. Все молекулярные структуры были стандартизованы с помощью ChemAxon Standardizer[104]. В рамках стандартизации проводилась ароматизация структур, добавление водородов в явном виде в структуры и т.д.

2.2.02. Методы построения QSPR моделей

Стандартная схема была следующей:

- Расчет симплексных дескрипторов.

- Предобработка исходного массива дескрипторов, включающая в себя нормирование и шкалирование дескрипторов.
- Построение статистических моделей с использованием современных методов машинного обучения.
- Валидация моделей с помощью процедуры 5-кратной внешней кросс-валидации. Оценка качества и области применимости моделей.
- Консенсусное моделирование – объединение ряда моделей для получения улучшенного усредненного прогноза.
- Интерпретация моделей с точки зрения структуры исследуемых органических соединений и влияющих физико-химических факторов.

Построение моделей проводилось в среде программирования с открытым кодом R. QSPR анализ проводился с использованием современных методов машинного обучения:

- Метод обобщенной линейной регрессии (Generalized Linear Model –GLM) [52] – универсальный метод построения регрессионных моделей, позволяющий учитывать взаимодействие между факторами, вид распределения зависимой переменной и предположения о характере регрессионной зависимости;
- Метод частичных наименьших квадратов (Partial Least Squares – PLS) [53, 105] – один из наиболее популярных методов, суть которого состоит в одновременном разложении матрицы данных **X** и матрицы концентраций **C** и переходе к новым переменным.;
- Метод случайного леса (Random Forest – Random Forest) [59] – алгоритм машинного обучения, заключающийся в использовании комитета (ансамбля) решающих деревьев;
- Метод градиентного бустинга (Gradient Boosting Machine – GBM)[106, 107] – метод, схожий с предыдущим, однако использующий дифференциальную функцию потерь при построении каждого следующего дерева;
- Метод опорных векторов (Support Vector Machine – SVM)[55, 56] – метод, основанный на переводе исходных векторов в пространство более высокой

размерности и поиск разделяющей гиперплоскости с максимальным зазором в этом пространстве.

Для оценки качества моделей использовались следующие параметры: R^2 – коэффициент детерминации (средняя ошибка 0.03 рассчитывалась в границах 95% доверительного интервала); RMSE – стандартная ошибка прогноза. Для критических свойств смесей также использовали показатели MAPE – средняя абсолютная процентная ошибка, AAD – среднее абсолютное отклонение, RMS – относительная стандартная ошибка прогноза:

$$RMS = 100 * \sqrt{((\sum(y_{obs} - y_{pred})) / y_{obs}) / (\text{количество точек}))}$$

, где y_{obs} – наблюдаемые значения, y_{pred} – предсказанные значения.

Коэффициент корреляции R между изучаемыми значениями свойства (y) и вычисленными (\hat{y}) по уравнению рассчитывается по формуле [108]:

$$R = \frac{m \sum_{i=2}^m y_i \hat{y}_i - \sum_{i=2}^m y_i \cdot \sum_{i=2}^m \hat{y}_i}{\sqrt{m \sum_{i=2}^m y_i^2 - \left(\sum_{i=2}^m \hat{y}_i \right)^2} \sqrt{m \sum_{i=2}^m \hat{y}_i^2 - \left(\sum_{i=2}^m y_i \right)^2}}$$

Где x_i – структурные параметры (независимые переменные).

Значение R^2 измеряет «долю общего разброса, относительно среднего \hat{y} , объясняемую регрессией».

Стандартная ошибка \hat{y} определяется по формуле [108]:

$$RMSE = \frac{\sum_{i=1}^m (y_i - \hat{y}_i)^2}{m} SD = \sum_{i=1}^m (y_i - \hat{y}_i)^2$$

где m – количество молекул в выборке; y_i – заданное значение активности, \hat{y}_i – вычисленное значение активности, \hat{y}_i^{cv} – значение активности, вычисленное в условиях кросс-валидации для i-той молекулы.

Для валидации моделей использовалась процедура 5-кратной внешней кросс-валидации [109], которая позволяет получить объективную прогнозирующую способность моделей даже в отсутствие внешнего теста (при условии достаточной репрезентативности выборки). Во время этой процедуры данные предварительно сортируются согласно значению свойства, затем каждое пятое соединение помещается в тестовую выборку. Процедуру повторяют со сдвигом до тех пор, пока не будет покрыт весь диапазон соединений. При валидации моделей для смесей использовалась модификация ранее разработанного подхода “mixtures out” [74]. В рамках данного подхода в тестовую выборку добавляют все точки смеси. В случае построенных моделей данные про индивидуальные компоненты не включаются в модель.

Определение области применимости моделей (domain applicability – DA). DA определяет область химического пространства, которая адекватно описывается соединениями обучающей выборки. Статистические модели могут надежно прогнозировать только те соединения, которые принадлежат этой области. В работе для определения DA был использован метод фрагментного контроля [110]. В этом подходе любая молекула, содержащая в себе фрагменты, которые не встречаются в обучающей выборке, находится за пределами DA.

2.2.02.1. Структурная и физико-химическая интерпретация построенных моделей

Основной целью физико-химической интерпретации является определение того, какие именно переменные, взвешенные согласно молекулярным свойствам (частичным зарядам, электроотрицательности, возможности образования водородных связей и т.д.) имеют наибольшее значение для исследуемого свойства. Для физико-химической интерпретации была использована оценка важности переменных по Бриману [59]. Для измерения важности j -й переменной после обучения модели, все значения j -й переменной подвергаются пермутации (т.е.

перемешиваются) для всех точек обучающей выборки. Важность переменной затем рассчитывается как разность между ошибками для оригинальной обучающей выборки и для перемутированных данных.

Для структурной интерпретации моделей для индивидуальных соединений использовался универсальный фрагментный подход [111], в котором вклад молекулярного фрагмента в исследуемое свойство рассчитывается как разность между значением для выбранного соединения и спрогнозированным значением для того же соединения за исключением этого фрагмента. Универсальный фрагментный подход является потенциально применимым к любому типу дескрипторов и позволяет оценивать разброс значений вкладов для разных соединений. Для структурной интерпретации моделей для смесей оценивалась (согласно схеме Бримана) важность отдельных смесевых фрагментов.

2.2.03. Корреляционный и каузальный анализ

Корреляционный анализ включает в себя ряд методов, позволяющих обнаружить зависимость между несколькими случайными величинами. Статистика говорит о корреляции между двумя переменными и указывает силу связи при помощи некоторого критерия взаимосвязи. Нами был использован коэффициент корреляции Пирсона (или линейный коэффициент корреляции) r , который является мерой линейной корреляции между двумя переменными X и Y . Коэффициент корреляции Пирсона (или линейный коэффициент корреляции) является мерой линейной корреляции между двумя переменными X и Y , и способен принимать значения между $+1$ и -1 включительно. При значении $+1$ наблюдается полностью положительная корреляция, при 0 – отсутствие корреляции, и при -1 – полностью отрицательная корреляция. Для определения статистической достоверности полученных коэффициентов корреляции использовался t -тест с количеством степеней свободы, равным размеру выборки $- 2$. Асимптотический доверительный интервал рассчитывался на основании Z трансформации Фишера [112].

Корреляционный анализ не дает возможности определить наличие причинно-следственных связей между переменными. Известная максима гласит: «Корреляция не обуславливает каузальности». Причинно-следственные взаимосвязи не могут быть получены или определены только в терминах статистических взаимосвязей [113]. Существует ряд математических методов каузального анализа, которые позволяют *in silico* воспроизвести рандомизированный контролируемый эксперимент, который является стандартным способом выявить причинно-следственные взаимосвязи. Для понимания каузальных методов необходимо ввести некоторые математические определения. Для двух случайных переменных X, Y , совместное распределение вероятностей – такое распределение вероятностей, которое дает вероятность что X, Y вместе попадают в определенный интервал дискретного набора значений. Две переменные X, Y являются независимыми, когда совместное распределение вероятности (X, Y) является произведением вероятности X и вероятности Y для всех значений X и Y . Каузальный анализ нашел применение в работах по биоинформатике [114] для определения наличия причинно-следственной связи между факторами транскрипции и генами. Каузальный анализ был ранее использован в задачах «структура-свойство» для описания свойств наночастиц [115].

Первый использованный в работе алгоритм каузального анализа – это алгоритм PC (названный так по инициалам его главных создателей), разработанный Spirtes, Glymour, и Scheines (1991). На первом шаге алгоритма для всех переменных строится полный неориентированный граф. Затем для наборов вершин графа проводятся тесты на условную независимость. В том случае, если две переменные независимы между собой, либо для двух переменных общей причиной является третья переменная – ребро удаляется. Затем для всех оставшихся ребер определяется ориентация исходя из определенного набора правил. Например, ребра не могут быть ориентированы таким образом, чтобы образовывался замкнутый цикл.

Второй алгоритм – алгоритм LTr (Linear Trace method) [116, 117]. Данный метод был разработан для выявления каузальных отношений в многомерных

данных. Основной идеей метода является использование асимметрии между распределениями предполагаемых причины и эффекта. Такого рода асимметрия наблюдается в том случае, если ковариационная матрица причины и матрица, связывающая причину с эффектом, вероятностно независимы. Этот метод одинаково применим к данным как с гауссовым (нормальным), так и с негауссовым распределением.

Для статистических расчетов было использовано программное обеспечение с открытым кодом R [118]. Алгоритмы каузального анализа находятся в открытом доступе, исходный код находится на <http://webdav.tuebingen.mpg.de/causality/>.

РАЗДЕЛ 3
КОЛИЧЕСТВЕННАЯ ОЦЕНКА ВЛИЯНИЯ СТРУКТУРЫ
НА ТЕРМОДИНАМИЧЕСКИЕ СВОЙСТВА
ИНДИВИДУАЛЬНЫХ ОРГАНИЧЕСКИХ ВЕЩЕСТВ
И ИХ БИНАРНЫХ СМЕСЕЙ

3.1. Критические свойства индивидуальных соединений

3.1.01. Экспериментальные данные

Большой массив накопленных экспериментальных данных по критическим свойствам позволяет составить представительную выборку, включающую информацию о большом числе соединений различных классов. Соединения, вошедшие в выборку, были взяты из справочника [100].

Таким образом, сформированы базы данных в которых количество изучаемых соединений: для критической температуры – 407, для критического давления – 382, для критического объема – 309, для фактора Питцера – 331. Наибольшее число данных было доступно, таким образом, для критических температур. Имеющиеся данные покрывают широкий диапазон значений (для критических температур от около 100 К до около 900 К, для критических давлений от 10 до около 90 бар, критических объемов от 100 до около 1000 см³/моль).

В обучающей выборке были представлены углерод-, галоген-, кислород-, азот- и серусодержащие органические соединения. Рассмотренные классы включают в себя насыщенные и ненасыщенные углеводороды, ароматические углеводороды, гетероциклические соединения, спирты, простые и сложные эфиры и их производные. Разнообразие выборки дает возможность построить достаточно универсальную QSPR-модель, не ограниченную соединениями лишь одного из представленных классов.

3.1.02. Результаты и обсуждение

Наилучшие результаты при моделировании [119, 120] были получены с использованием метода Random Forest. Все итоговые модели (табл. 2) характеризуются как высоким качеством аппроксимации ($R^2 = 0.97-0.99$), так и высокой прогнозирующей способностью в рамках процедуры 5-кратной внешней кросс-валидации ($R^2_{5f}=0.86-0.95$). Исходя из диаграмм «наблюдаемое - предсказанное» (рис. 9а), можно сделать вывод об увеличении ошибки прогноза в случае критической температуры для крайних точек (минимальные и максимальные значения температуры) тестовых выборок. Для критических давлений (рис. 9б) данные распределены более равномерно. Для критических объемов (рис. 9в) также наблюдается увеличение разброса для высоких значений свойства. На диаграмме «наблюдаемое-предсказанное» для фактора Питцера (рис. 9г) имеется довольно большой разброс для всего диапазона значений. Это может быть объяснено тем, что при расчете данного параметра используются значения давления насыщенного пара, которые в ряде данных могли быть получены путем аппроксимации экспериментальных, что вносит дополнительную ошибку.

Таблица 2

Параметры консенсусных моделей

(R^2 – коэффициент детерминации, S – среднеквадратичное отклонение,

AAE – средняя абсолютная ошибка)

	Параметры	Обучающая выборка	Консенсусная модель (5-кратная внешняя кросс-валидация)
T_c , К	R^2	0.99	0.89
	S	12.5	32.7
	AAE	9.4	23.4

$V_c, \text{см}^3/\text{моль}$	R^2	0.99	0.95
	S	11	38
	AAE	6	16
$P_c, \text{бар}$	R^2	0.98	0.95
	S	2.8	4.1
	AAE	2.4	3.8
ω	R^2	0.97	0.86
	S	0.027	0.060
	AAE	0.014	0.039

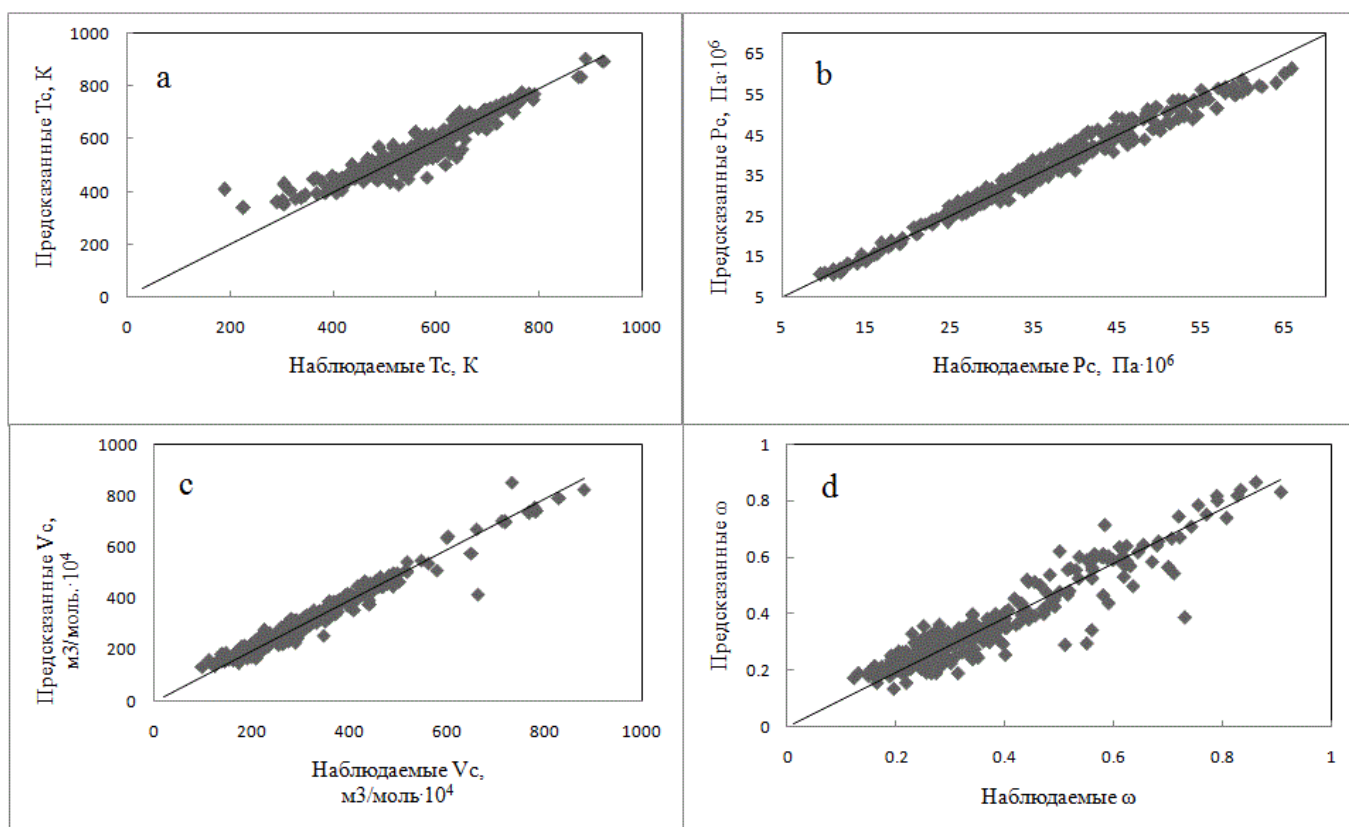


Рис. 6. Диаграммы «наблюдаемое-предсказанное» для исследуемых свойств:

- а) критическая температура, б) критическое давление, в) критический объем, д) фактор Питцера.

Анализ распределения ошибок прогноза для тестовой выборки (табл. 3) показывает, что для всех свойств максимум распределения приходится на диапазон допустимой экспериментальной ошибки – менее 10% [121]. Для критических температуры и объема большая часть соединений прогнозируется с ошибкой менее 3%, что соответствует высокой экспериментальной точности. Для критического давления и фактора Питцера большинство соединений прогнозируется с ошибкой от 3 до 10 %. Абсолютная ошибка, выраженная с помощью среднего абсолютного отклонения, также обладает максимумом в области близкой к экспериментальной ошибке.

Таблица 3

Количество соединений с различными ошибками прогноза
для критических температур из работы[80].

Выборка	Относительная ошибка, %	
	<3	3-10
Обучающая	90	10
Тестовая	83	19

Качество аппроксимации моделей, в сравнении с существующими методами (в частности, с методами групповых вкладов), достаточно высоко. Более подробное сравнение затруднено, из-за различия обучающих выборок, сравниваемых методов, однако возможно привести ориентировочные величины абсолютных ошибок аппроксимации для классических методов групповых вкладов[12]: для критической температуры – 15.9-43.6 К, для критического объема – 12-44 см³/моль, для критического давления – 1.39-2.88 бар, для фактора Питцера – 0.04-0.07 логарифмических единиц.

Существующие QSPR-работы, посвященные критическим свойствам, заслуживают отдельного внимания. Работы, посвященные отдельным классам веществ[75], имеют ограниченную область применения и поэтому подробно обсуждаться не будут. Из актуальных работ последних лет стоит упомянуть [80] и

[79], как ввиду большого количества изучаемых там соединений, так и ввиду сравнения двух наиболее известных баз данных по термодинамическим свойствам – DIPPR[122] и NIST[102]. В работе[80] авторами были построены QSPR-модели для критических температур, в работе [79] – модели для критических температур, давлений и объемов.

Стоит отметить трудозатратность методологии построения моделей, применявшейся авторами – в рамках изнурительного 3D-моделирования проводился конформационный поиск с использованием полуэмпирических квантово-химических методов. Кроме того, в ряде случаев, использованные дескрипторы[123] не имели ясной интерпретации.

Сравнивая полученные нами результаты с работой [80], можно сделать заключение о том, что распределение относительной ошибки в [80] близко к распределению относительной ошибки в СПМС QSPR моделях (см. рис. 9). Анализ результатов работы [79] показывает переобученность моделей, поскольку приведенная относительная ошибка прогноза $<2\%$ (меньше экспериментальной). Предположение о переобученности подтверждается тем фактом, что для таких классов соединений как сульфиды, сложные эфиры, ненасыщенные простые эфиры ошибка прогноза $>20\%$ и, согласно авторам, данные классы были исключены из итоговой модели.

Поскольку методология СПМС-QSPR предполагает использование, как меток атомов, так и взвешивания согласно их физико-химических свойств, было проведено сравнение степени влияния дескрипторов, относящихся к этим двум типам. Важность такого рода сравнения стоит подчеркнуть, поскольку для успешного описания зависимости «структура-свойства» и дальнейшего использования полученных моделей необходимо конкретизировать те специфические характеристики молекул, от которых зависит проявляемое свойство (например, поляризуемость или вандерваальсовы взаимодействия).

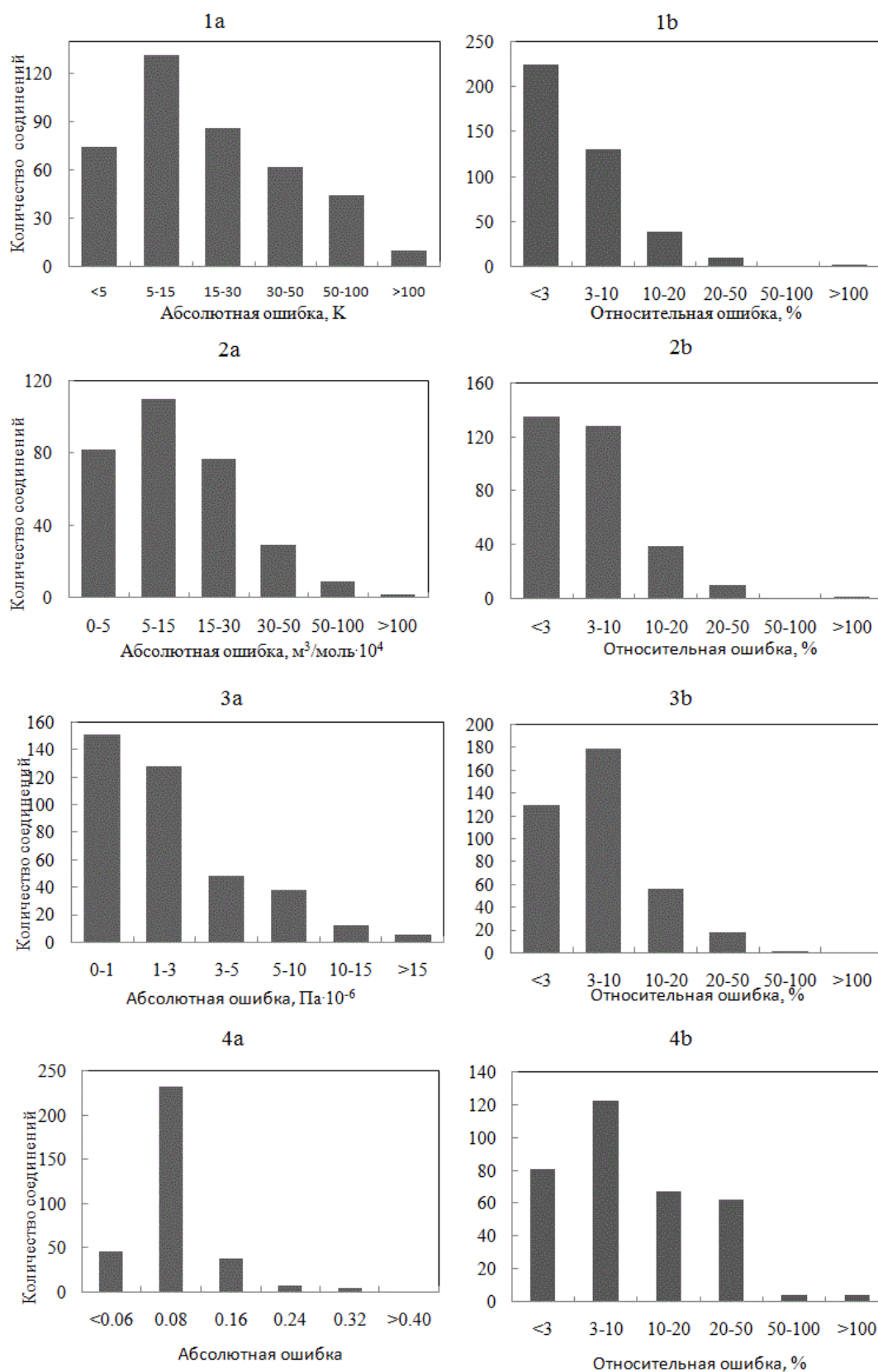


Рис.7. Распределение абсолютной (а) и относительной (б) ошибок прогноза (1 – критическая температура, 2 – критический объем, 3 – критическое давление, 4 – фактор Питцера).

Вклад дескрипторов, взвешенных согласно меткам атомов, свидетельствует о необходимости включения дополнительных физико-химических свойств атомов в процедуру взвешивания. Дескрипторы, основанные на метках атомов, в среднем описывают изучаемые свойства на 80 %. Подробный анализ вкладов характеристик атомов (см. рис. 11) подтверждает физическую обоснованность полученных моделей. Несмотря на то, что QSPR-модели являются частным случаем моделей типа «черного ящика», полученные вклады делают возможным анализ зависимости критических свойств от структуры молекул в рамках терминов, описывающих межмолекулярные взаимодействия. Предварительно стоит сделать некоторые уточнения. При расчете дескрипторов, описывающих вандерваальсовы взаимодействия, использовалось уравнение Леннарда-Джонса 6-12, которое в своей классической форме учитывает как лондоновские дисперсионные, так и обменные взаимодействия[124]. Такие параметры как электроотрицательность и частичные заряды, описывают распределение электронной плотности в молекуле и связаны как с дисперсионными, так и с электростатическими взаимодействиями.

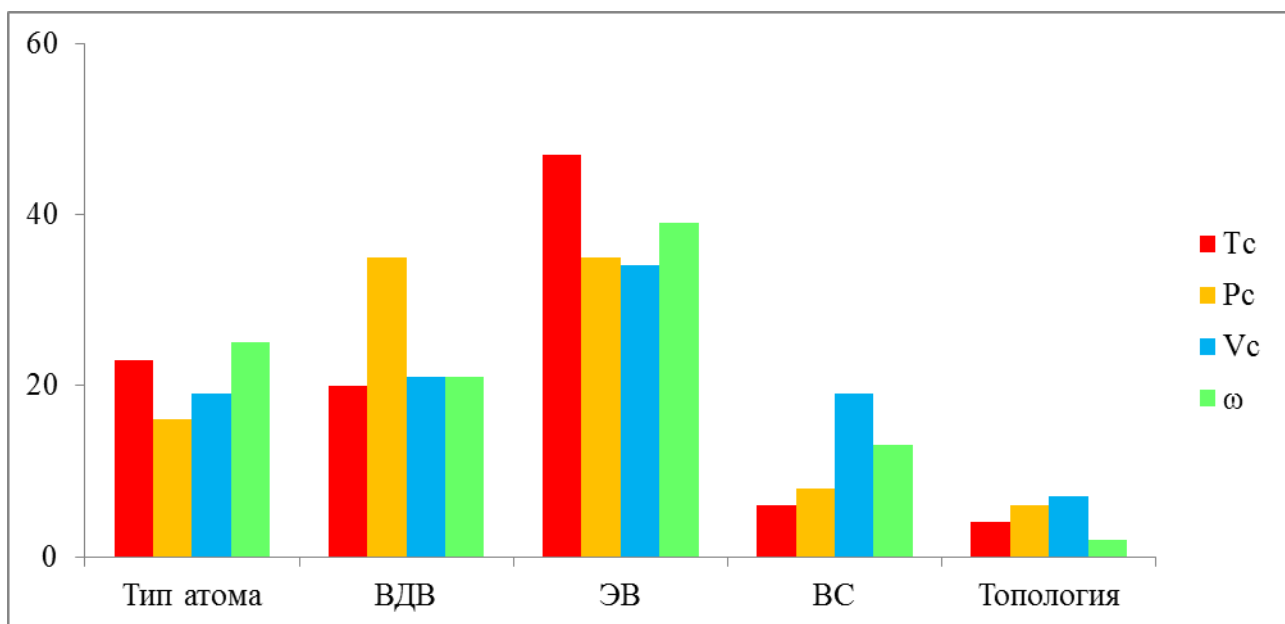


Рис.8. Влияние различных факторов на исследуемые свойства

(ВС – способность атома быть донором, либо акцептором водородной связи; ВДВ – вандерваальсовы взаимодействия; ЭВ – электростатические взаимодействия).

Из анализа влияния дескрипторов на исследуемые свойства (рис. 11) следует, что электростатические и вандерваальсовы взаимодействия вносят основной вклад. Это подтверждает тот факт, что параметры критической точки являются характеристиками, непосредственно зависящими от энергии межмолекулярных взаимодействий. В частности, критические температуры существенно зависят от энергии межмолекулярных взаимодействий и, соответственно, от тех параметров структуры, которые определяют эти взаимодействия. Критические объемы, как и следует ожидать, зависят от стерических свойств, а при переходе к энергии взаимодействия – от геометрических характеристик потенциала. Для критического давления и фактора Питцера важную роль играют как энергия межмолекулярных взаимодействий, так и стерические параметры. Для конкретного вида потенциала (в данной работе применялись параметры потенциала Леннард-Джонса 6-12) известны следующие соотношения, связывающие потенциалы параметра с критическими свойствами [100]:

$$\begin{aligned}\frac{\varepsilon}{kT_{\text{кр}}} &= 0.795 + 0.1963\omega \\ \sigma\left(\frac{p_{\text{кр}}}{T_{\text{кр}}}\right) &= 2.3551 - 0.0874\omega \\ \sigma &= 0.15894V_{\text{кр}}\end{aligned}$$

Вышеуказанные соотношения не только иллюстрируют полученные результаты, но и позволяют на основе значений критических параметров и фактора Питцера непосредственно перейти к параметрам потенциала Леннард-Джонса.

Анализ влияния взвешенных дескрипторов показывает доминирующий вклад электростатических и вандерваальсовых взаимодействий. Это подтверждается тем фактом, что параметры критической точки непосредственно зависят от энергии межмолекулярных взаимодействий. Роль топологии молекулы (разветвленность скелета без учета природы атомов) сравнительно мала для всех критических свойств. Стоит более детально рассмотреть некоторые индивидуальные закономерности, так для критических объемов существенно важными являются вандерваальсовы радиусы. На критические давления, по сравнению с другими критическими свойствами, значительное влияние оказывает способность

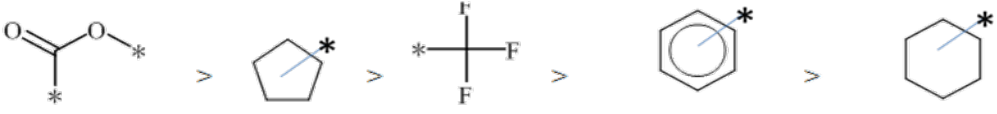
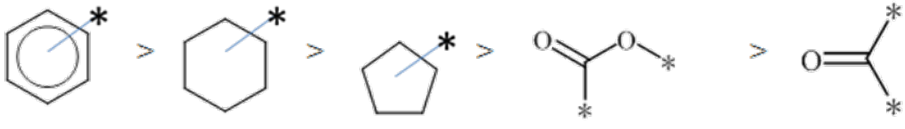
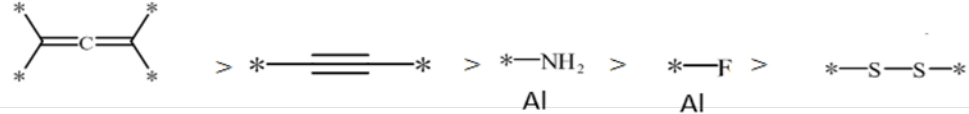
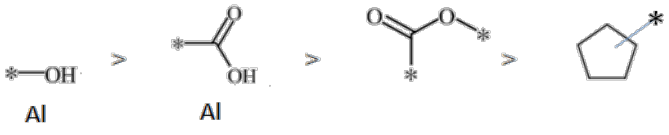
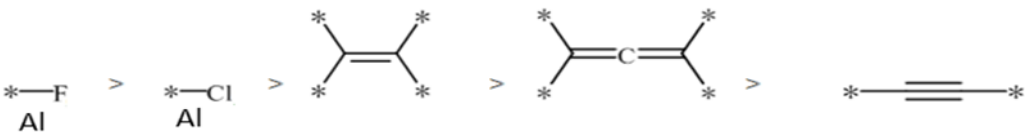
соединений образовывать водородные связи. На критические температуры оказывают большее влияние электронная поляризуемость (рефракция) и электроотрицательность, в то время как для критических объемов на первом месте находится совокупное влияние вандерваальсовых взаимодействий.

Структурная интерпретация (табл. 4) выявила наиболее влияющие молекулярные фрагменты. Для критических температур максимальным положительным влиянием обладают ароматические и алициклические фрагменты, а также фрагменты, содержащие карбонильную группу; наибольшим отрицательным влиянием – фторсодержащие фрагменты. Такие тенденции можно объяснить, исходя из того, что критическая температура связана с энергией межмолекулярного взаимодействия и возрастает при повышении способности молекул к взаимному притяжению. В свою очередь повышение способности молекул к отталкиванию отрицательно влияет на критическую температуру. Для критических давлений наибольшим положительным влиянием обладает дисульфидная группа, алифатический фтор, первичная аминогруппа, тройная $C\equiv C$ связь и алифатическая гидроксогруппа. Закономерно, что для критических объемов, наиболее объемные фрагменты обладают положительным влиянием.

Таблица 4

Структурная интерпретация для критических свойств и фактора Питцера индивидуальных органических веществ (АЛ – фрагмент в алифатическом окружении, АРОМ – фрагмент в ароматической системе)

T_c	
+	
-	
P_c	

+	$*-\text{S}-\text{S}-* > \underset{\text{Al}}{*-\text{F}} > *-\text{NH}_2 > *-\text{C}\equiv\text{C}-* > \underset{\text{Al}}{*-\text{OH}}$
-	
V_c	
+	
-	
ω	
+	
-	

3.2. Использование «псевдосмесового» подхода для моделирования критических свойств

Ряд моделей для критических свойств был построен с использованием «псевдосмесового подхода» [95]. В качестве сравнения были выбраны модели, основанные на традиционных симплексных дескрипторах – далее «модель одной молекулы», и модели, основанные на «псевдосмесовом» подходе – соответственно «псевдосмесевая модель».

Характеристики консенсусных моделей и их 95% доверительный интервал

		T_c	V_c	P_c
Модель “одной молекулы”	R^2_{cv}	0.89 ± 0.04	0.95 ± 0.01	0.95 ± 0.01
	$RMSE_{cv}$	32.70 ± 0.13 (K)	28.0 ± 0.10 ($cm^3/моль$)	4.10 ± 0.02 (бар)
“Псевдосмесевая” модель	R^2_{cv}	0.90 ± 0.01	0.92 ± 0.01	0.91 ± 0.01
	$RMSE_{cv}$	30.60 ± 0.15 (K)	24.0 ± 0.12 ($cm^3/моль$)	4.00 ± 0.02 (бар)

Статистические характеристики конечных моделей приведены в табл. 5. Несмотря на тот факт, что корреляционные коэффициенты полученных моделей близки и отличаются незначительно, разница в RMSE статистически значима в 95% доверительном интервале. В данном случае опущены и не будут обсуждаться «псевдосмесевые» модели для фактора Питцера по нескольким соображениям. Результирующие модели для фактора Питцера для обоих подходов оказались идентичны по своим статистическим характеристикам. И, более того, таковой результат был ожидаем, поскольку фактор Питцера является индивидуальной характеристикой молекулы – отклонения ее формы от сферической. При рассмотрении относительного уменьшения ошибок моделей RMSE (%) видно, что наибольшее улучшение наблюдается для V_c – около 15%, затем для P_c – 8% и T_c – около 7%. Улучшение моделей для V_c может быть объяснено аддитивной природой свойства. Характеристики полученных моделей позволяют включить их в экспертный комплекс прогнозирования термодинамических свойств органических соединений.

Графики «наблюдаемое-предсказанное» демонстрируют (как и для моделей «одной молекулы»), что наибольшие отклонения наблюдаются близко к границам графиков (рис. 13). Наибольшими являются ошибки предсказания для небольших значений критических свойств в связи со сложностью описания очень малых

молекул, таких как метан, поскольку для них существует слишком мало уникальных характерных дескрипторов. С другой стороны, большие ошибки прогноза для больших значений критических свойств могут быть обусловлены увеличением экспериментальной ошибки для более сложных молекул. Например, экспериментальные ошибки для критических температур метана и этана составляют только 0.3 К, в то время как для нонадекана и эйкозана ошибки около 8 К. Для кислородсодержащих соединений, например, этанола, ацетальдегида и уксусной кислоты экспериментальные ошибки составляют соответственно 7 К, 2 К и 30 К [102].

Для всех дескрипторов в «псевдосмесевых» моделях наибольший вклад принадлежит электростатическим и дисперсионным взаимодействиям, которые играют значительную роль и с теоретической точки зрения. В данном случае под дисперсионными взаимодействиями имеется в виду сумма важностей для всех дескрипторов, взвешенных согласно константам потенциала Леннарда-Джонса 6-12, под электростатическими взаимодействиями – сумму важностей для дескрипторов, взвешенных по рефракции, частичным зарядам и электроотрицательности (что соответствует в основном взаимодействиям типа диполь-диполь, диполь-индуцированный диполь). Роль топологии сравнительно мала для всех исследуемых критических свойств. Достаточно интересно то, что способность атома служить донором либо акцептором водородной связи является более важной для критических давлений – 19%, по сравнению с 6% и 9% для критических температур и критических давлений соответственно. Однозначно интерпретировать это наблюдение затруднительно в связи с тесной связью всех трех свойств.

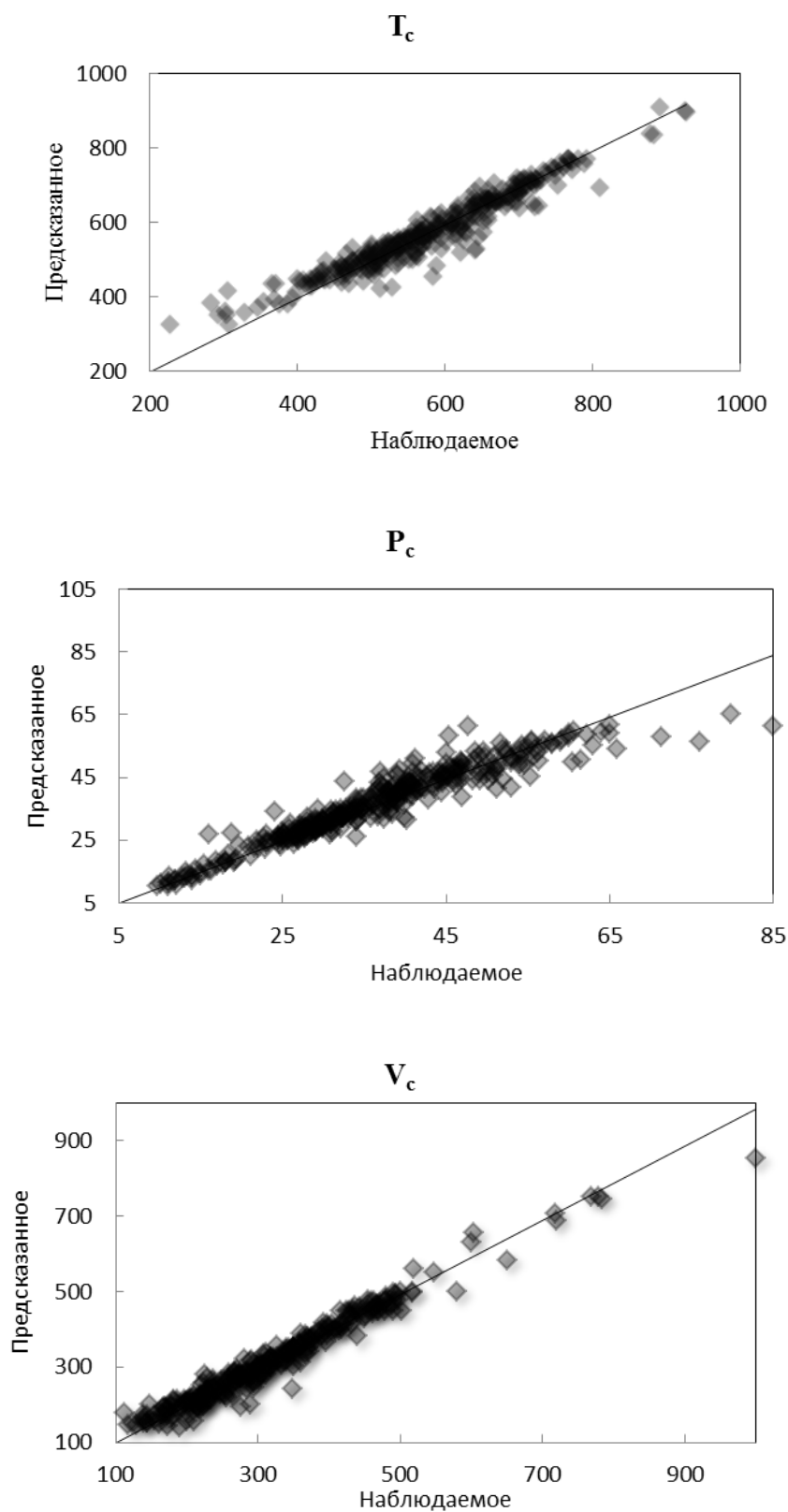


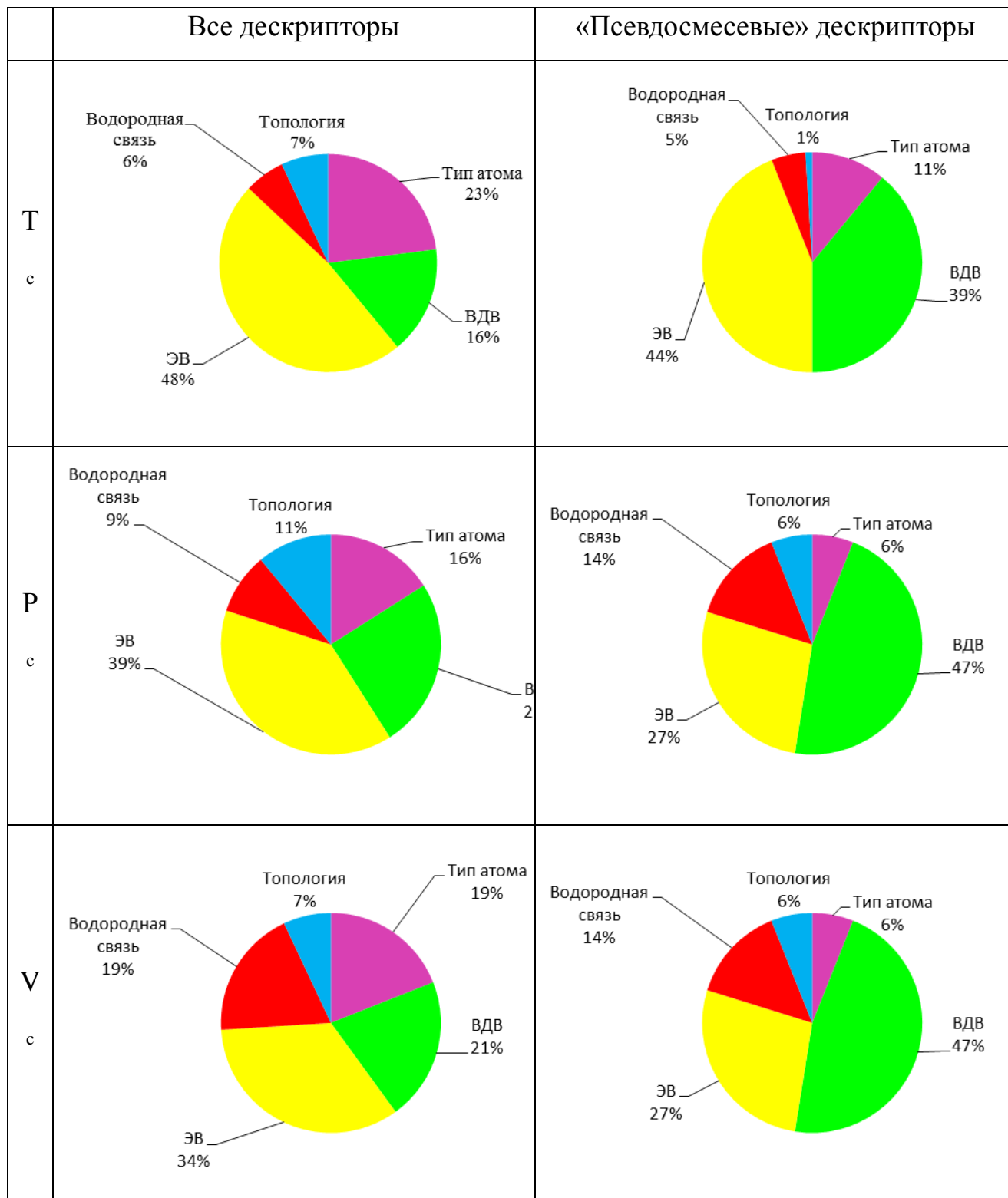
Рис. 9. Графики «наблюдаемое-предсказанное» для «псевдосмесевых» моделей.

Сравнение важностей молекулярных характеристик для всех дескрипторов и только «псевдосмесевых» (табл. 6) показывает, что для «псевдосмесевых»

дескрипторов выше важность дисперсионных взаимодействий. Для критических давлений в случае «псевдосмесевых» дескрипторов важность способности атома быть донором либо акцептором водородной связи значительно выше. Важность способности атома к образованию водородной связи может также увеличиваться благодаря учету возможного образования межмолекулярной водородной связи (что невозможно учесть в рамках модели «одной молекулы»), что отвечает за формирование ассоциатов, и таким образом влияет на критическое давление. Вклад топологии в случае «псевдосмесевой» модели намного меньше. Важность симплексов, взвешенных согласно типу атомов, также меньше, что демонстрирует большую физико-химическую интерпретируемость модели. Рассчитанный уровень вхождения «псевдосмесевых» дескрипторов в итоговые модели для всех трех свойств (T_c , P_c , V_c) составляет около 25%.

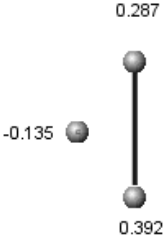
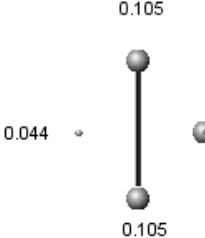
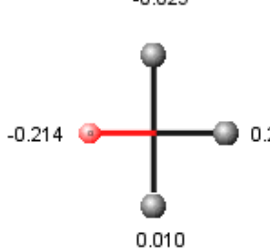
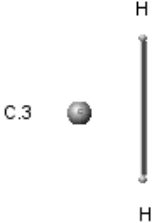
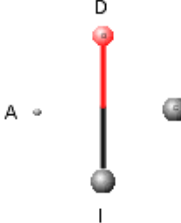
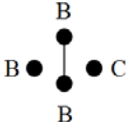
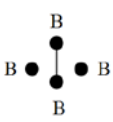
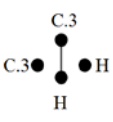
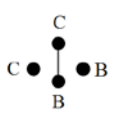
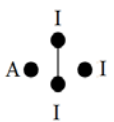
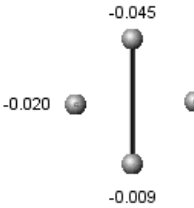
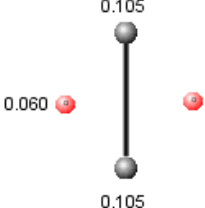
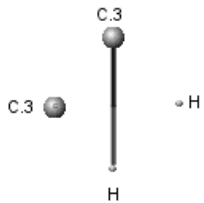

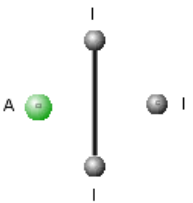
Большая часть дескрипторов, вошедших в итоговую модель, была представлена тетраатомными фрагментами. Наиболее важные симплексы представлены в табл. 7. Безусловно, эти фрагменты не могут быть интерпретируемы как четырех-центровые взаимодействия, но скорее как важность парного атом-атомного взаимодействия в определенном окружении. Можно увидеть, что наиболее важные симплексы относятся ко 2 и 3 типам (более подробно в табл. 1). Общие тенденции важности переменных слегка изменены. Например, для критических объемов и критических давлений одним из наиболее важных фрагментов являются симплексы 2 типа, взвешенные согласно способности атома быть донором/акцептором водородной связи, в то время как эти фрагменты не являются важными для критических температур. Для всех трех критических свойств наиболее важными симплексами являются симплексы 2 типа, описывающие электростатические взаимодействия, затем симплексы 2 типа, описывающие дисперсионные взаимодействия.

Важность физико-химических характеристик молекул
(ДВ – дисперсионные взаимодействия, ЭВ – электростатические
взаимодействия)



Примеры наиболее важных несвязных «псевдосмесевых» дескрипторов

с	Фр агмент					
	Тип *	ЭВ	ЭВ	ТА	ДВ	ТА
	Пр имер					
с	Фр агмент					
	Тип	ЭВ	ДВ	ЭВ	ТА**	Д/А

	Пример	    
с	Фрагмент	    
	Тип	ЭВ ДВ ТА ДВ Д/А
	Пример	    

* ЭВ – электростатические взаимодействия, ДВ – дисперсионные взаимодействия, ТА – тип атома, Д/А – способность атома быть донором/акцептором водородной связи

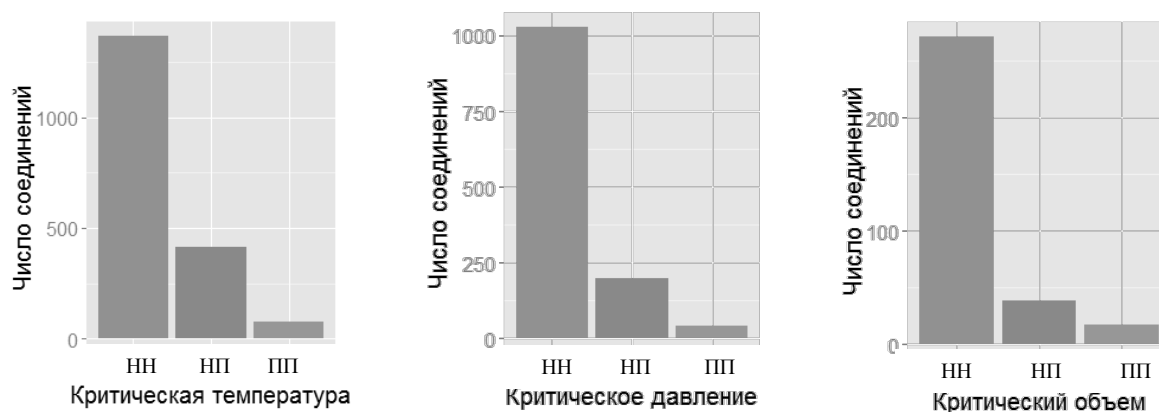
** С.ar – ароматический углерод,

С.3 – sp³ углерод

3.3. Критические свойства бинарных смесей

3.3.01. Данные

В сравнении с данными для критических свойств индивидуальных соединений, данные по критическим свойствам бинарных смесей органических соединений менее доступны [23], что объясняется сложностью экспериментального определения критических свойств смеси. В настоящей работе были использованы данные из обзора [4]. В этом обзоре есть как экспериментальные величины, так и данные, полученные экстраполяции и интерполяции других экспериментальных данных [125].



** НН – неполярное-неполярное; НП – неполярное-полярное; ПП – полярное-полярное

Рис. 10. Гистограммы распределения критических свойств бинарных смесей

В выборке представлены 94 соединения, около 300 смесей различных составов для каждого свойства, всего с учетом различных составов – около 1000 точек для каждого свойства. Диапазон величин также достаточно велик – от 20 бар до около 100 бар для критических давлений, от около 150 К до около 800 К для критических температур, от около 80 см³/моль до около 400 см³/моль для критических объемов.

Из литературных данных известно, что сложность описания свойств смесей возрастает при переходе от смесей неполярных соединений к смесям полярных соединений [23]. Для дальнейшей оценки выборки, соединения разбивались на группы согласно значениям дипольного момента, неполярными считались соединения с дипольным моментом меньше 0.5. Рассмотренные классы соединений включают в себя насыщенные и ненасыщенные углеводороды, ароматические углеводороды, гетероциклические соединения, спирты, простые и сложные эфиры и их производные. Как видно из рис. 14 наибольшее число точек данных существует для неполярных смесей, наименьшее – для полярных смесей, что объясняется меньшей устойчивостью последних в экспериментальных условиях..

3.3.02. *Результаты и обсуждение*

Был получен ряд моделей с помощью методов GBM, PLS, SVM и Random Forest [126]. Характеристики полученных моделей указаны в таблице 8. Жирным шрифтом выделены наилучшие модели. Как видно из таблицы, для критических температур и критических давлений статистические характеристики PLS-моделей являются неудовлетворительными, что может объясняться нелинейностью влияния структуры на данные свойства. Для критических свойств смесей, помимо симплексных дескрипторов, также были использованы парные межмолекулярные дескрипторы.

Выбранные наилучшие модели, полученные методом Random Forest, показывают высокие статистические характеристики, с наилучшим коэффициентом детерминации для критических объемов равным 0.93 и наихудшим коэффициентом детерминации для критических давлений равным 0.81.

Характеристики полученных моделей для 5-кратной внешней кросс-валидации

	Random Forest		PLS		GBM		SVM	
	R^2_{ts}	RMSE _{ts}	R^2_{ts}	RMSE _{ts}	R^2_{ts}	RMSE _{ts}	R^2_{ts}	RMSE _{ts}
T _c	0.93	23.2 (К)	0.66	52.9 (К)	0.91	26.8 (К)	0.88	31.4 (К)
P _c	0.81	24.6 (бар)	0.44	45.1 (бар)	0.80	25.2 (бар)	0.66	39.3 (бар)
V _c	0.92	37 (см ³ /моль)	0.88	45 (см ³ /моль)	0.73	69 (см ³ /моль)	0.93	33 (см ³ /моль)

Как можно заметить, в данном случае критические объемы и критические температуры являются наиболее точно моделируемыми свойствами. В случае критических температур это может быть объяснено более высокой экспериментальной точностью их измерения, критические же объемы, являясь аддитивным свойством, более просты в описании, чем другие критические свойства. Для экспериментальных измерений критических давлений, как и для моделей характерна меньшая экспериментальная точность. Сравнение расчетной ошибки с рекомендуемой экспериментальной [121] (см. табл. 9) показывает более чем удовлетворительные результаты моделирования. Характеристики полученных моделей позволяют включить их в экспертный комплекс прогнозирования термодинамических свойств органических соединений.

Таблица 9

Сравнение ошибок наилучших моделей и рекомендованных экспериментальных ошибок

	Модель	R^2_{ts}	RMSE _{ts}	MAPE _{ts} , %	MAPE _{exp} , %
T _c	Random Forest	0.93	23.2 (К)	6.8	~5

P_c	Random Forest	0.81	24.6 (бар)	11.5	~10
V_c	SVM	0.93	33 (см ³ /моль)	14.6	~10

Анализ важности переменных [59] для смесей показывает следующие тенденции (рис.15). Для критических объемов наибольшим значением важности обладает электронная поляризуемость. Как электронная поляризуемость, так и критические объемы являются величинами имеющими размерность объема, что может объяснять эту тенденцию.

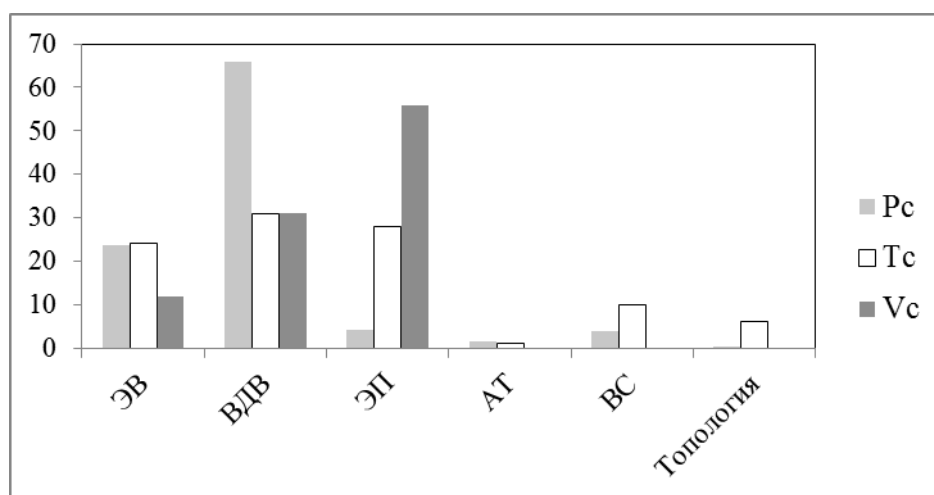


Рис.11. Относительная важность переменных (%).

(ЭВ – электростатические взаимодействия, ВДВ – ванн-дер-ваальсовы взаимодействия, ЭП – электронная поляризуемость, ВС – водородные связи, АТ – тип атома)

Важным преимуществом используемых дескрипторов является возможность их осмысленной интерпретации в рамках простых молекулярных свойств. Наиболее важным влияющим в случае критических температур и критических давлений факторами являются вандерваальсовы взаимодействия (включающие в себя парные дескрипторы), затем электростатические взаимодействия (частичные заряды). Поскольку критические температуры являются свойством непосредственно связанным с энергией межмолекулярных

взаимодействий, которая в свою очередь имеет существенную электростатическую и вандерваальсову составляющую, такого рода влияние является закономерным. Незначительный вклад дескрипторов, взвешенных по водородной связи, может быть объяснен небольшим количеством соединений, способных к ее образованию – от 11 до 22%. Сравнение с важностью переменных из моделей для чистых веществ [95, 119] показывает схожие тенденции – как для индивидуальных веществ, так и для бинарных смесей преобладающим является вклад дескрипторов, описывающих электростатические и вандерваальсовы взаимодействия.

Критические параметры непосредственно связаны с механизмом межмолекулярного взаимодействия в конденсированной фазе. Так, критическая температура определяет кинетическую энергию межмолекулярного взаимодействия в системе, где исчезают различия между жидкой и газовой фазами. Численно критические температуры близки к величине ε/k , где ε – глубина потенциальной ямы парного потенциала и связаны с параметрами потенциала простыми соотношениями: $\varepsilon \sim T_c, \sigma^3 \sim V_c, \frac{\varepsilon}{\sigma^3} \sim P_c$. Расхождения между этими величинами вызвано коллективными характеристиками критических явлений. Критический объем существенно зависит от собственного объема молекул и является величиной аддитивной. Связь координат критической точки с кривой межмолекулярного взаимодействия проявляется в простых соотношениях с параметрами уравнений Ван-дер-Ваальса [3].

Дополнительный анализ относительной важности различных типов дескрипторов, т.е. смесевых, несмесевых и парных, показывает влияние каждого из этих типов дескрипторов на изучаемые свойства. Для всех свойств основной вклад вносят симплексные дескрипторы. Вклад парных межмолекулярных дескрипторов (которые относятся к вандерваальсовым взаимодействиям) для критических давлений является преобладающим. Это может быть обусловлено тем, что критические давления являются свойством в значительной степени зависящим от образования ассоциатов (например, за

счет дисперсионных взаимодействий). Для всех критических свойств способность атома к образованию водородной связи, топология и тип атома играют незначительную роль.

Структурная интерпретация выявила наиболее влияющие фрагменты компонентов смеси (табл. 10). Для всех свойств наибольший вклад вносят разветвленные углеводородные фрагменты, что соответствует известным эмпирическим закономерностям, поскольку с возрастанием длины и разветвленности углеводородной цепи компонентов смеси, значения ее критических свойств возрастают. Для критических температур наибольшим положительным влиянием обладают кислородсодержащие фрагменты (а также совместное присутствие в смеси ароматических и кислородсодержащих фрагментов), при наличии которых в смеси могут протекать процессы ассоциации. Для критических давлений смесей, как и для критических давлений индивидуальных органических соединений, наиболее важны фрагменты с ненасыщенными углерод-углеродными связями. Для критических объемов, как и в случае индивидуальных соединений, наиболее вероятно влияние фрагментов обусловлено размером фрагмента.

Таблица 10

Структурная интерпретация для критических свойств бинарных смесей органических соединений.

T_c	
P_c	
V_c	

3.4. Каузальный анализ фактора Питцера: влияние параметров формы и размера молекул

Фактор Питцера является характеристикой подобия молекул. Основное применение фактор Питцера находит при использовании принципа соответственных состояний, согласно которому все вещества подчиняются одному уравнению состояния, выраженному через приведенные переменные. Приведенные переменные являются отношением величины параметра в данной точке к его величине в критической точке. Для расчета значения ω необходимо знать приведенную температуру $T_r = T/T_c = 0,7$, а также критическое давление.

Фактор Питцера первоначально был интерпретирован как связанный с ацентричностью или асферичностью молекулы [12]. Для благородных газов фактор Питцера считается равным нулю; для небольших близких к сферичности молекул (например, метана) принимает значения близкие к нулю, а для длинноцепочечных углеводородов его значение резко возрастает. Исходя из определения, фактор Питцера обладает нулевыми значениями для благородных газов аргона, криптона и ксенона. Также фактор Питцера интерпретируется как показатель несферичности поля молекулярных сил. Таким образом, фактор Питцера в основном использовался как показатель сложности молекулы, связанный в основном с ее строением и полярностью.

Тем не менее, связь фактора Питцера с геометрией молекулы и полярностью ранее имела исключительно эмпирическое объяснение. В немногочисленных работах [84, 94, 127, 128] была показана корреляция фактор Питцера с различного рода геометрическими параметрами, однако такого рода исследования не носили систематического характера. В рамках данной работы были применены методы корреляционного и каузального (причинно-следственного) анализа [115, 129] для определения основных физико-

химических факторов, которые влияют на фактор Питцера, и уточнения его интерпретации.

3.4.01. Структурные параметры

Исходя из известной классификации дескрипторов, проведенной Тодесчини и др. [67], для анализа был выбран ряд структурных параметров, связанных с формой и размерами молекул. Полный список параметров (дескрипторов) приведен в табл. 11. Дескрипторы были рассчитаны в программе Dragon [130].

3.4.02. Результаты и обсуждение

Среди использованных дескрипторов можно выделить 2 группы: дескрипторы, которые отвечают за асферичность формы молекулы – PJI2, DECC, MDDD, SPH, ASP, MECC, PJI3; и дескрипторы, которые связаны с размером молекулы – ECC, AECC, RGyr, SPAN. Для дескрипторов 1 группы SPH, ASP, MECC их значение совпадает с традиционной интерпретацией фактора Питцера – диапазон значений варьируется от 0 до 1, при этом 0 соответствует сферической форме молекулы, 1 показывает приближение формы к линейной. Дескриптор DECC является, в отличие от ECC, AECC дескриптором не размера, но формы, исходя из определения. Аналогично дескриптором формы является дескриптор MDDD, который описывает среднее смещение степеней расстояний вершин.

Список использованных дескрипторов [67].

Аббревиатура	Полное название	Описание
Топологические дескрипторы		
PJI2	2D Petitjean shape index	<p>2D Petitjean индекс рассчитывается для молекулярного графа как разница между топологическим диаметром и топологическим радиусом, деленная на радиус. Топологический диаметр определяется как максимальный атомный эксцентриситет; топологический радиус – как минимальный атомный эксцентриситет, где атомный эксцентриситет – максимальное топологическое расстояние от атома до другого атома.</p> $PJI2 = \frac{D-R}{R}$ <p>где D – топологический диаметр, R – топологический радиус.</p>
ECC	eccentricity	<p>Эксцентриситет – сумма атомных эксцентриситетов по всем неводородным атомам в молекуле.</p> $ECC = \sum_i \eta_i$ <p>где η_i – эксцентриситет <i>i</i>-го атома.</p>
AECC	average eccentricity	<p>Средний эксцентриситет рассчитывается как эксцентриситет ECC деленный на число неводородных атомов.</p> $AECC = \frac{ECC}{n_{noHAt}}$ <p>где n_{noHAt} – число неводородных атомов.</p>

DECC	eccentric	<p>Эксцентрик – мера среднего смещения атомных эксцентриситетов от среднего эксцентриситета AECC.</p> $DECC = \frac{\sum_i \eta_i - AECC}{n_{At}}$ <p>где η_i – эксцентриситет i-го атома, n_{At} – число атомов.</p>
MDDD	mean distance degree deviation	<p>Среднее отклонение степеней расстояний рассчитывается как среднее смещение степеней расстояний вершин от средней степени расстояний вершин, где степень расстояний вершин – среднее суммы строк матрицы расстояний молекулы, рассчитанной как сумма степеней расстояний деленных на количество неводородных атомов в молекуле.</p> $MDDD = \frac{\sum_i (\sum_{j=1}^{n_{noHAt}} d_{ij} - \bar{d})}{n_{noHAt}}$ <p>где \bar{d} – средняя степень расстояний вершин, $\sum_{j=1}^{n_{noHAt}} d_{ij}$ – среднее смещение степеней расстояний вершины i.</p>
Геометрические дескрипторы		
RGyr	radius of gyration	<p>Радиус гирации – дескриптор, описывающий распределение атомных масс в молекуле. Рассчитывается как:</p> $RGyr = \sqrt{\frac{\sum_{i=1}^{n_{AT}} m_i r_i^2}{MW}}$ <p>где r_i – расстояние до i-го атома от центра масс молекулы, m_i – соответствующая атомная масса, n_{AT} – число атомов и MW – молекулярная масса.</p>

SPAN	SPAN	<p>SPAN – дескриптор, определяемый как радиус наименьшей сферы, центрированной на центре масс молекулы, полностью закрывающей все атомы в молекуле:</p> $SPAN = \max_i(r_i)$ <p>где r_i – расстояние до iго атома от центра масс молекулы.</p>
MEcc	molecular eccentricity	<p>Молекулярный эксцентриситет – дескриптор формы молекулы, рассчитываемый из собственных значений λ молекулярной матрицы инерции:</p> $MEcc = \frac{(\lambda_1^2 - \lambda_2^2)^{1/2}}{\lambda_1}$ <p>Его значения находятся в диапазоне от 0 до 1, 0 соответствует сферическим молекулам, 1 – линейным молекулам.</p>
SPH	spherosity	<p>Сферичность – анизометрический дескриптор, рассчитываемый как функция собственных значений ковариационной матрицы, рассчитанной из молекулярной матрицы:</p> $SPH = \frac{3 \cdot \lambda_3}{\lambda_1 + \lambda_2 + \lambda_3}$ <p>Сферичность варьируется от 0 для плоских молекул, таких как бензол, до 1 для полностью сферичных молекул.</p>

ASP	asphericity	<p>Асферичность – анизометрический дескриптор, который измеряет отклонение конформации молекулы от сферической формы, рассчитывается из собственных значений λ расчетной молекулярной матрицы инерции следующим образом:</p> $ASP = \frac{1}{2} \cdot \frac{(\lambda_1 - \lambda_2)^2 + (\lambda_1 - \lambda_3)^2 + (\lambda_2 - \lambda_3)^2}{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}$ <p>Асферичность молекулы варьируется от 0 для сферичных молекул до 1 для линейных молекул.</p>
PJI3	3D Petitjean shape index	<p>3D индекс формы Petitjean является отношением разности геометрического диаметра молекулы и геометрического радиуса молекулы к геометрическому радиусу молекулы. Геометрический радиус молекулы определяется как минимальный геометрический эксцентриситет, а диаметр определяется как максимальный геометрический эксцентриситет. В свою очередь, геометрический атомный эксцентриситет – наибольшее геометрическое расстояние от атома до любого другого атома в молекуле.</p> $PJI3 = \frac{D_{3D} - R_{3D}}{R_{3D}}$ <p>где D_{3D} – геометрический диаметр молекулы, R_{3D} – геометрический радиус молекулы.</p>

Построенные нами корреляции (табл. 12) позволяют выделить наиболее значимые дескрипторы среди исследуемых. Наивысшие уровни корреляции ($r > 0.6$) наблюдаются для дескрипторов S0K, PHI, ECC, AECC, DECC, MDDD, ICR, RGyr, SPAN. Наименьшие корреляции ($r < 0.4$) наблюдаются для дескрипторов PJI2, SHP2, MECC, SPH, ASP, PJI3. В выборке были выделены несколько групп соединений – с точки зрения их способности к ассоциации – углеводороды, соединения с потенциальными водородными связями, полярные соединения, не образующие водородных связей (но способные к ассоциации за счет различного рода иных взаимодействий). Для каждой из этих групп были отдельно рассчитаны уровни корреляции. Результаты расчетов приведены в табл. 12. Коэффициенты корреляции для соединений, способных к образованию водородных связей, значительно ниже, чем для двух других исследуемых групп. Коэффициенты корреляции для соединений без водородных связей и углеводородов в большинстве случаев отличаются незначительно.

Таблица 12

Уровни корреляции и доверительные интервалы для наиболее значимых дескрипторов

Дескриптор	Коэффициент корреляции для всей выборки	Коэффициент корреляции для соединений без водородных связей	Коэффициент корреляции для углеводородов
S0K	0.73±0.05	0.88±0.03	0.91±0.02
PHI	0.68±0.06	0.84±0.04	0.89±0.02
ECC	0.73±0.05	0.91±0.02	0.93±0.02
AECC	0.75±0.04	0.92±0.02	0.96±0.01
DECC	0.73±0.05	0.89±0.03	0.93±0.02
MDDD	0.74±0.05	0.91±0.02	0.93±0.02

UNIP	0.73±0.05	0.92±0.02	0.94±0.01
CENT	0.69±0.05	0.85±0.03	0.89±0.04
ICR	0.67±0.06	0.82±0.06	0.87±0.03
RGyr	0.74±0.05	0.92±0.02	0.96±0.01
SPAN	0.74±0.05	0.91±0.02	0.96±0.02

Проведенный нами каузальный анализ позволил оценить причинно-следственные взаимосвязи между фактором Питцера и дескрипторами. Метод РС при доверительной вероятности 0.95 показывает, что дескриптор AECC каузально влияет на фактор Питцера. Для других дескрипторов каузальные связи не выявлены. Ввиду недостаточной чувствительности метода [117] РС для дополнительного анализа был привлечен метод LTr. Метод LTr отличается высокой робастностью, способностью к работе с большими объемами «зашумленных» данных и чувствительностью к наличию причинно-следственных связей. Метод LTr показал, что дескрипторы ECC, AECC, DECC значимо причинно влияют на фактор Питцера. В свою очередь для топологических дескрипторов – MDDD, UNIP, CENT, ICR – и геометрических дескрипторов – RGyr, SPAN, связанных с размером молекулы, наличие каузальной связи не подтверждается ни одним из используемых методов. Неожиданные, но интересные результаты были получены для дескриптора ASP, рассчитанный линейный коэффициент корреляции для которого 0.39. Метод LTr показал наличие значимого каузального влияния дескриптора ASP на фактор Питцера. При этом примечательно то, что интерпретация дескриптора ASP соответствует традиционной интерпретация фактора Питцера – он связан с отклонением формы молекулы от сферической формы, и диапазон его значений варьируется от 0 до 1. Отдельно был проведен каузальный анализ для группы углеводов. Для соединений класса углеводов дескрипторы формы не являются влияющими, что может быть связано со сферичностью оптимальной конформационной формы для молекул в неидеальных условиях. На рис. 16 приведены графики зависимости фактора Питцера для четырех дескрипторов, значимое причинно-следственное влияние которых на фактор Питцера показал

каузальный анализ. Из приведенного графика для исследуемых углеводов (рис. 1г) видно, что в отсутствие молекулярных ассоциатов дескриптор описывает фактор Питцера с высокой точностью. Для всей изучаемой выборки (рис. 1б а-в) значение фактора Питцера для соединений с водородной связью завышены, что связано с процессами молекулярной ассоциации. Поскольку размер молекулярных ассоциатов закономерно превышает размер индивидуальной молекулы, то расчетные значения молекулярных размеров не совпадают с известными нам экспериментальными данными. Метод LTg также позволяет непосредственно количественно рассчитать величину причинно-следственного эффекта для двух имеющихся переменных, что позволяет отсортировать изучаемые дескрипторы по величине каузального влияния. Несмотря на то, что в известных нам ранних работах была показана высокая корреляция радиуса гирации с фактором Питцера, каузальной связи между этими двумя величинами нами не было обнаружено. В порядке убывания абсолютного влияния дескрипторы образуют следующий ряд: ECC > AECC > DECC ~ASP. Как видно, молекулярные дескрипторы размеры гораздо более влиятельны, чем дескрипторы формы. Таким образом, традиционная интерпретация фактора Питцера была нами дополнена и уточнена.

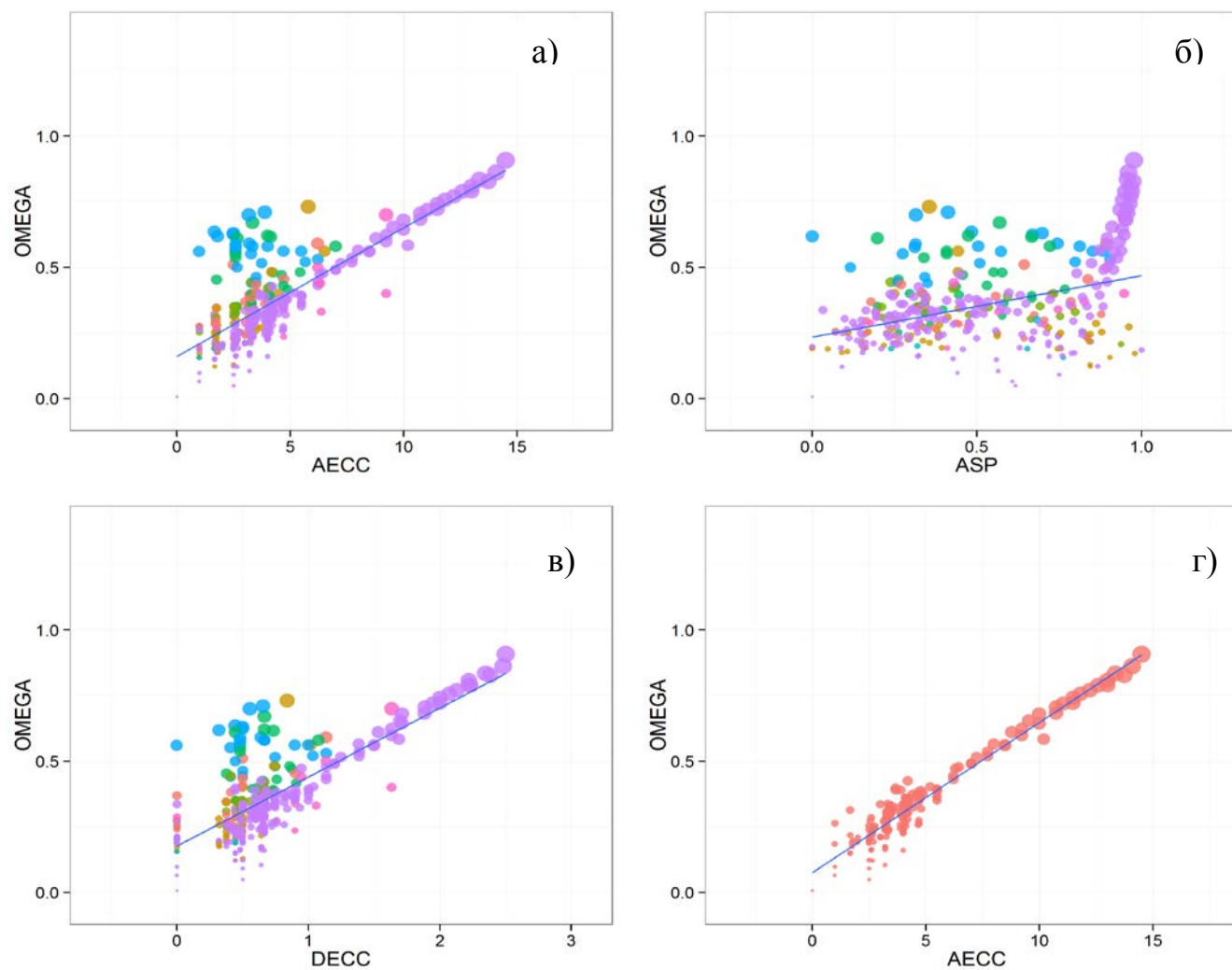


Рис.12. Зависимость фактора Питцера от различных дескрипторов (а,б,в – для всей выборки; г – для углеводов).

3.5. Вторые вириальные коэффициенты индивидуальных веществ

3.5.01. Данные

Выборка для вторых вириальных коэффициентов [101] индивидуальных органических соединений содержит данные по более чем 250 органическим соединениям. Для каждого исследуемого соединения дан ряд значений второго вириального коэффициента ($\text{см}^3/\text{моль}$) при разных температурах (K), таким образом общее число точек данных составляет более чем 4500 точек.

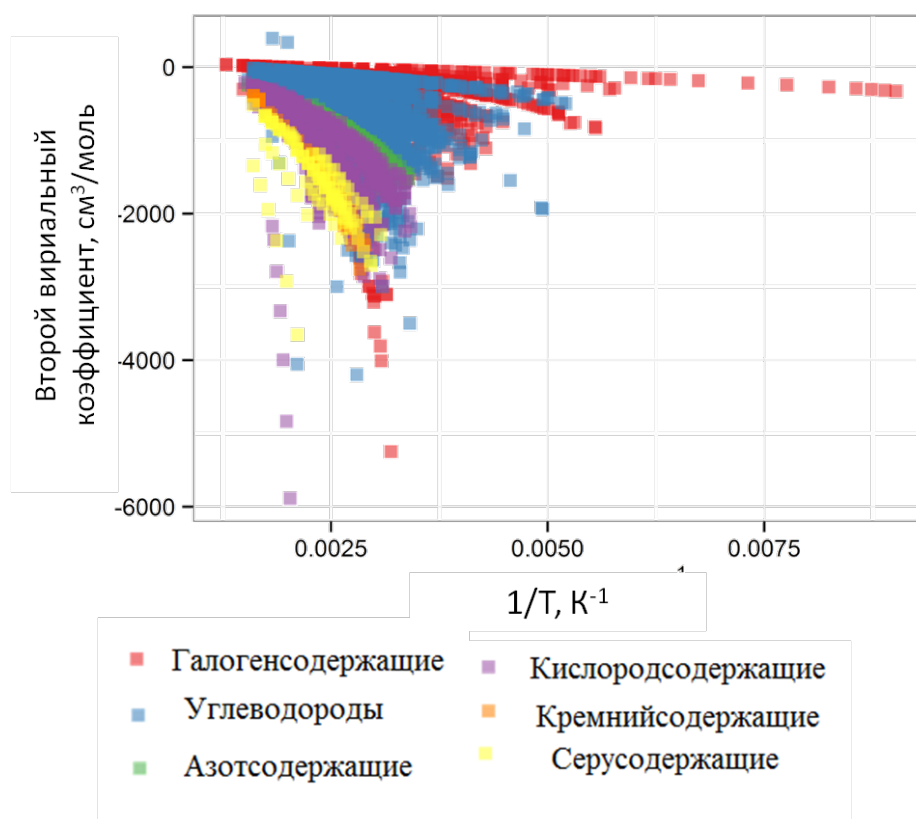


Рис. 13. Распределение данных для второго вириального коэффициента в зависимости от обратной температуры.

Предварительное изучение данных по вторым вириальным коэффициентам индивидуальных органических веществ показало значительную кластеризацию исследуемых данных (рис. 17). Можно заметить, что углеводороды и их галогенпроизводные в среднем имеют более высокие значения второго вириального коэффициента и для них наклон температурной кривой $B = f(T)$ невелик. Напротив, азот-, кислород- и серусодержащие соединения в среднем по

выборке имеют более низкие значения второго вириального коэффициента и для них наклон температурной кривой $B = f(T)$ достаточно велик. Линейная аппроксимация зависимости второго вириального коэффициента от температуры $B = f(T)$ показывает высокое качество аппроксимации с коэффициентом детерминации $R^2=0.96-0.98$, что позволяет нам использовать разработанный двухслойный подход к моделированию и, собственно, применять уравнение Клаузиуса.

3.5.02. *Схема моделирования*

Как и моделирование любого другого температурозависимого свойства, моделирование второго вириального коэффициента является совершенно нетривиальной задачей для QSAR/QSPR моделирования. В случае вторых вириальных коэффициентов подготовка данных к моделированию значительно затрудняется тем, что температурные интервалы для исследуемых данных неравны, иногда даже не пересекаются между собой и не существует опорных температурных точек (как, например, как комнатная температура для растворимости).

Может быть предложено несколько основных подходов к моделированию данных по второму вириальному коэффициенту. Первый подход (назовем его Т-моделью) – наиболее простой и применим к практически любому свойству, зависящему от температуры. Температура, в качестве независимого единичного дескриптора, включается в общий набор переменных так, что:

$$B = f(1/T, D_1, D_2, \dots, D_{mn})$$

Такой подход к моделированию температурозависимых свойств является наиболее простым и примитивным и, следовательно, является наиболее распространенным и популярным при описании данных свойств в рамках QSAR/QSPR-моделировании. Тем не менее, наиболее простой подход в данном случае не является наиболее корректным подходом. Как можно увидеть из вышеприведенного уравнения, все результирующие модели будут иметь

одинаковый вклад влияния температуры для всех исследуемых соединений, что является в корне неверным. Таким образом, данный подход не был применен, и в дальнейшем нами обсуждаться не будет.

Второй подход к моделированию температурозависимых свойств (далее называемый коэффициентной моделью или двухслойной моделью) основывается на строго физически обоснованной методологии, которая использует простое уравнение Клаузиуса, получаемое из фундаментального уравнения Ван-дер-Ваальса для реальных газов. Для сравнения качества моделирования было использовано две формы представления этого уравнения – линейная и экспоненциальная [131]:

$$B = b - a/RT$$

$$B = b - \exp(a/RT)$$

$$B = f(a, b)$$

$$a = f(D_1, D_2, \dots, D_n)$$

$$b = f(D_1, D_2, \dots, D_m)$$

, где a , b – коэффициенты вандерваальсова уравнения; D – дескрипторы; T – температура.

Несмотря на кажущуюся простоту приведенных уравнений, этот подход позволяет исследователю физически значимо интерпретировать полученную модели, так как оба коэффициента уравнения имеют четкий и строгий физический смысл. Таким образом, стандартная схема построения моделей была нами успешно модифицирована. В рамках нового двухслойного подхода на первом шаге расчетов путем аппроксимации экспериментальных данных к уравнению Клаузиуса рассчитываются коэффициенты a и b , затем рассчитанные коэффициенты моделируются (каждый отдельно) с помощью стандартной процедуры построения QSAR/QSPR моделей. Для получения спрогнозированных значений второго вириального предсказанные коэффициенты a и b подставляются в исходные уравнения и производится расчет. При выборе исходных данных, отбирались точки не менее чем с 4 значениями B при различных температурах.

3.5.03. Результаты и обсуждение

Используя вышеописанную схему двухслойного моделирования, был построен ряд моделей [132]. Помимо стандартных симплексных дескрипторов также был применен «псевдосмесевой» симплексного подход, поскольку вторые вириальные коэффициенты для индивидуальных веществ описывают парные взаимодействия однородных молекул, а «псевдосмесевые» дескрипторы направлены на определение всех возможных взаимодействующих несвязных фрагментов. В табл. 13 приведены основные статистические характеристики полученных моделей.

Таблица 13

Статистические характеристики консенсусных моделей

Модель «одной молекулы»

Уравнение	R^2_{ws}	$RMSE_{ws}$, см ³ /моль	R^2_{ts}	$RMSE_{ts}$, см ³ /моль
Линейная форма	0.98	25	0.76	260 ± 32
Экспоненциальная форма	0.98	20	0.78	255 ± 28

«Псевдосмесевая» модель

Уравнение	R^2_{ws}	$RMSE_{ws}$, см ³ /моль	R^2_{ts}	$RMSE_{ts}$, см ³ /моль
Линейная форма	0.98	20	0.79	190 ± 21
Экспоненциальная форма	0.99	18	0.81	185 ± 25

Как видно из вышеприведенной таблицы 13, аппроксимационная способность, описываемая R^2_{ws} и $RMSE_{ws}$, достаточно высока для всех используемых подходов. Использование экспоненциальной формы уравнения Клаузиуса дает незначительное уменьшение ошибки прогноза моделей. В то же время, добавление «псевдосмесевых» дескрипторов в модель позволяет

статистически значимо уменьшить значения RMSE прогноза для обеих форм уравнения Клаузиуса. Это статистическое улучшение легко объяснимо природой изучаемого свойства, которое зависит от взаимодействий между однородными молекулами вещества, как обсуждалось выше. Разница в значениях RMSE между моделью «одной молекулы» и «псевдосмесевой» моделью статистически значима в рамках 95% доверительного интервала, который рассчитывался между всеми тестовыми выборками, полученными в рамках процедуры 5-кратной внешней кросс-валидации. Разница между линейной и экспоненциальной формами уравнений Клаузиуса статистически незначима, что объясняется высоким качеством аппроксимации уравнения в обоих случаях – коэффициент аппроксимации $R^2=0.96-0.98$ (для линейной формы) и $R^2=0.97-0.99$ (для экспоненциальной формы). Высокие статистические характеристики полученных нами моделей несомненно позволяют включить их в созданный компьютерный экспертный комплекс прогнозирования термодинамических свойств органических соединений и их смесей.

Наилучшая из построенных модель, которой является «псевдосмесевая» модель для экспоненциальной формы уравнения Клаузиуса, была выбрана для анализа важности переменных (рис. 18). Примечательно, что полученные результаты по важности переменных очень схожи для обоих коэффициентов. Как было упомянуто выше, коэффициенты a и b имеют четкое и строгое физическое объяснение, a – мера отталкивания между частицами либо молекулами, b – исключенный объем 1 моля частиц либо молекул, который описывает межмолекулярное притяжение. Для обоих коэффициентов закономерно важны электростатические взаимодействия и поляризуемость, что хорошо коррелирует с изначальным физическим смыслом коэффициентов. Также проведенный нами анализ влияния физико-химических факторов на исследуемые коэффициенты показывает, что используя минимальный набор простых 2D дескрипторов, возможно описать сложные термодинамические свойства, непосредственно связанные с межмолекулярными взаимодействиями.

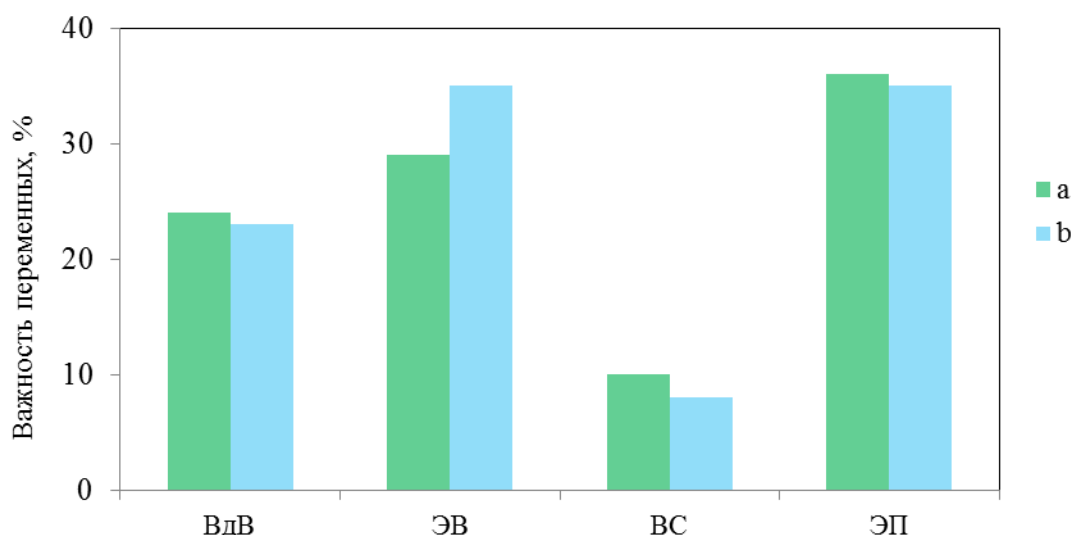


Рис. 14. Относительная важность переменных.

(ВС – водородная связь, ЭВ – электростатические взаимодействия, ВдВ – вандерваальсовы взаимодействия, ЭП – электронная поляризуемость)

При анализе полученных QSPR моделей по исследуемым термодинамическим свойствам, особое внимание нами уделялось тщательному сравнению ошибок прогноза моделей с ошибками исходных данных, а не только сравнению с ошибками других прогнозирующих методов. Поскольку каждое исследуемое свойство обладает своей собственной экспериментальной ошибкой, игнорирование этих ошибок может привести к неправильной и некорректной интерпретации и/или неправильному пониманию прогнозирующей способности и преимуществ полученных QSPR моделей. Исходя из этих соображений, мы отступили от традиционной парадигмы оценки ошибок QSPR моделей, основанной на сравнении QSPR моделей между собой, и провели поэтапное сравнение ошибок прогноза моделей и ошибок используемых данных. Такой подход к сравнению ошибок позволяет адекватно оценить реальную прогностическую способность QSPR моделей и полноценно учесть ошибку исходных данных. Детальное сравнение ошибок прогноза QSPR моделей и ошибок исследуемых экспериментальных данных показано на рис. 15 для основных классов исследуемых органических соединений (все значения RMSE рассчитаны для спрогнозированных значений из тестовых выборок).

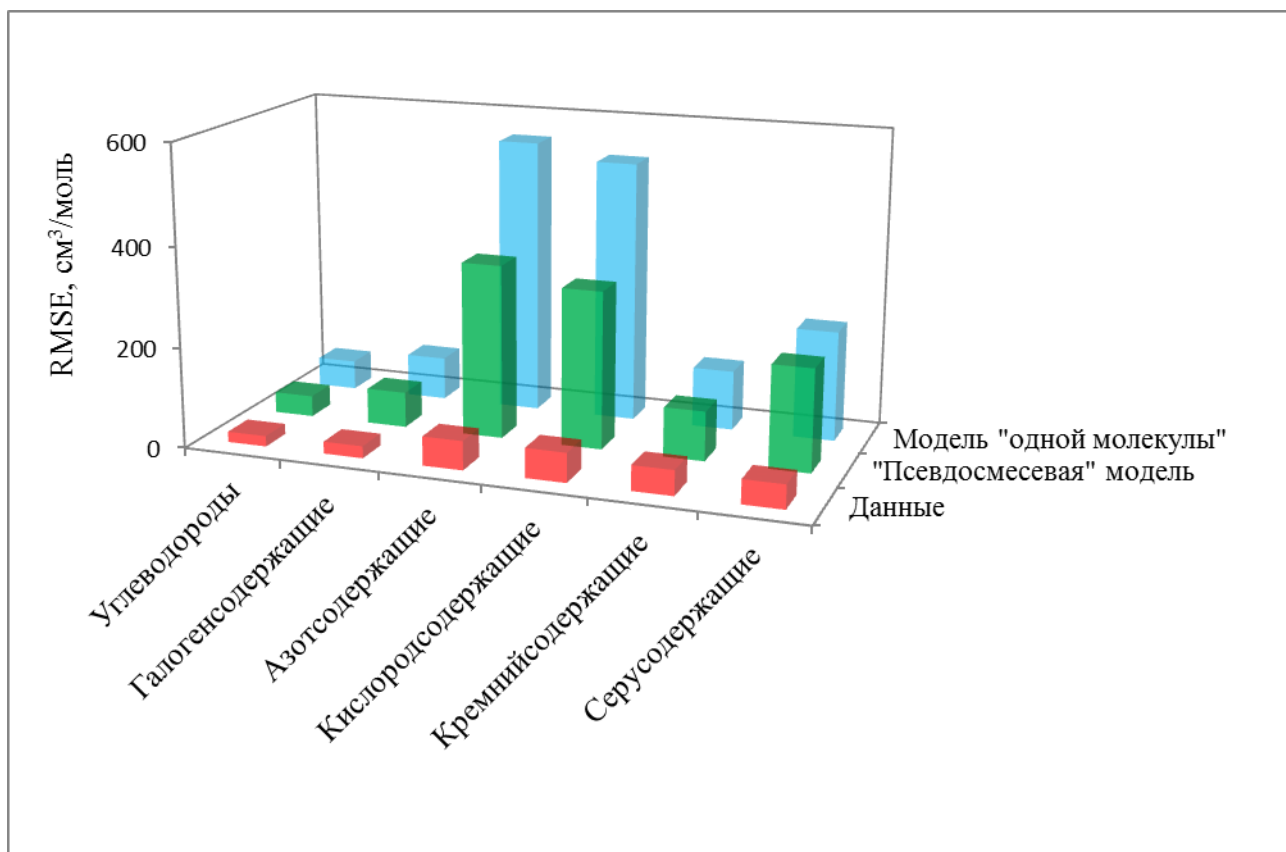


Рис. 15. Сравнение ошибок моделирования с ошибками данных[101].

Внимательное изучение соотношения между ошибками исследуемых данных и ошибками построенных QSPR моделей (рис. 19) показывает, что для исследуемых данных с меньшей экспериментальной ошибкой ошибка предсказания также невелика. Для азот- и кислородсодержащих соединений 2-кратное увеличение ошибки данных ведет к практически 4-кратному увеличению в ошибке прогноза полученных QSPR моделей. Большие экспериментальные ошибки для азот- и кислородсодержащих соединений могут быть вероятно объяснены как их меньшей температурной стабильностью, так и сложностью межмолекулярных взаимодействий и возможностью протекания реакций при экспериментальных условиях. Ошибка для полученных моделей «одной молекулы» в общем больше чем для полученных «псевдосмесевых» моделей.

Особый интерес представляет сравнение полученных QSPR моделей с предыдущими исследованиями. В работе Xiang с коллегами [92], основанной на

принципе соответственных состояний, редуцированный второй вириальный коэффициент рассчитывался по следующей формуле:

$$B_r = B\rho_c$$

Где B – второй вириальный коэффициент, см³/моль; B_r - приведенный второй вириальный коэффициент, ρ_c – критическая плотность, г/см³.

При этом метод Xiang был рекомендован NIST как стандартный метод для расчета второй вириальных коэффициентов. Для оценки качества прогноза в работе Xiang с коллегами[92] используются следующие статистические параметры:

$$AAD = 100 * \frac{|B_{obs} - B_{pred}| / B_{obs}}{\text{number of data points}}$$

$$RMS = 100 * \sqrt{\frac{|\sum (B_{obs} - B_{pred})| / B_{obs}}{\text{number of data points}}}$$

Поскольку используемые Xiang величины оценки ошибки не являются стандартными, то для сравнения результатов прогнозов наших моделей с моделью [92] мы конвертировали спрогнозированные значения в м³/моль и рассчитали B_r .

Сравнение ошибок предсказания для полученных симплексных QSPR моделей, основанных на экспоненциальной форме уравнения Клаузиуса, и моделей Xiang и др. показано в табл. 14 . Для большинства общих соединений ошибка прогноза симплексных моделей в общем намного ниже. Такая огромная разница может быть связана не только с эффективностью подходов, но также с разницей в использованных экспериментальных данных.

Таким образом, на примере QSPR моделей было показано, что использование простых физических закономерностей приводит к лучшим и более интерпретируемым результатам. Это показывает альтернативный путь моделирования температурозависимых свойств помимо добавления температуры

к общему массиву дескрипторов. Анализ ошибок по классам органических соединений подтверждает необходимость сравнения ошибок теоретического прогноза с ошибками экспериментальных данных для оценки реальных преимуществ и недостатков моделей. Более того, QSPR модели, построенные с использованием 2D дескрипторов сравнимы с одобренными термодинамическими методами, основанными на принципе соответственных состояний.

Структурная интерпретация проводилась отдельно для коэффициентов a и b (табл. 15). Наибольшим положительным влиянием на коэффициент a обладают фрагменты с ненасыщенными углерод-углеродными связями и галогенсодержащие фрагменты, наибольшим отрицательным влиянием обладают ароматические и алициклические, а также кислородсодержащие фрагменты. По своему физическому смыслу коэффициент a отображает межмолекулярное притяжение. В свою очередь для коэффициента b тенденции прямо противоположны, и фрагменты, способные к ассоциации (за счет образования водородных связей, стекинговых взаимодействий, и пр.) снижают энергию отталкивания. Поскольку коэффициент b описывает также исключенный молекулярный объем, объемные фрагменты также обладают для него положительным влиянием.

Сравнение ошибок моделей с ошибками метода Xiang [92]

Соединение	AAD PC	RMS PC	AAD TD	RMS TD	AAD ECS	RMS ECS	AAD СПМС экспоненциальная	RMS СПМС экспоненциальная
Этин	2.4	2.6	2.1	2.5	2.1	2.3	0.04	0.2
1,3- Диметилбензол	6.4	8	7	8.1	4.8	5.7	3.6	8
1,4- Диметилбензол	7.8	8.6	4.9	6.6	5.2	5.8	0.2	0.8
Фенол	10	10	6.2	6.3	4.9	5.5	0.2	0.8
Фторометан	11	11	7.2	7.4	9.4	9.5	0.1	0.3
1,2-Дихлороэтан	6	9.8	7.2	9.4	7.7	9	1.6	5.7
Бромметан	7.1	9.1	14	17	6.9	9.1	1.7	7.3
Диметиловый эфир	14	15	6.4	9.5	15	16	0.6	2.8

Метанол	24	29	15	17	22	25	0.2	1.5
Этанол	23	28	15	18	18	20	0.1	0.4
1-Пропанол	10	13	10	13	14	16	0.9	4.2
2-Пропанол	9.9	12	10	12	16	18	0.8	3.7
2-Метил-1-пропанол	4.9	7.3	23	23	19	20	3.2	6.5
2-Метил-2-пропанол	8	10	25	25	22	25	3.3	6.7
1-Бутанол	8.9	11	7.1	9.7	19	23	0.8	2.9
(RS)-2-Бутанол	15	16	17	17	9.7	11	3.6	7.2
Ацетонитрил	54	55	7.6	9.9	13	14	0.6	3.8
Нитрометан	49	49	3.7	4.5	14	14	0.02	0.1

PC = метод Pitzer и Curl [133], TD – метод Tarakad and Danner [84], ECS –расширенный принцип соответственных состояний Xiang [92]

Структурная интерпретация для вторых вириальных коэффициентов
индивидуальных органических соединений
(АЛ – группы в алифатическом окружении; АРОМ – группа в ароматической
системе)

<i>a</i>	
+	$R1 \equiv R2 > \underbrace{\{ -Cl \}}_{AL} > \{ -F \} > \begin{array}{c} R2 \quad R3 \\ \diagdown \quad / \\ R1 = R4 \end{array} > \underbrace{\{ -F \}}_{APOM}$
-	
<i>b</i>	
+	
-	

3.6. Вторые вириальные кросс-коэффициенты бинарных смесей органических соединений

3.6.01. Данные

Данные по вторым вириальным коэффициентам были выбраны из сборника [101]. При выборе нами отбирались данные не менее чем 4 значениями B_{12} при различных температурах.

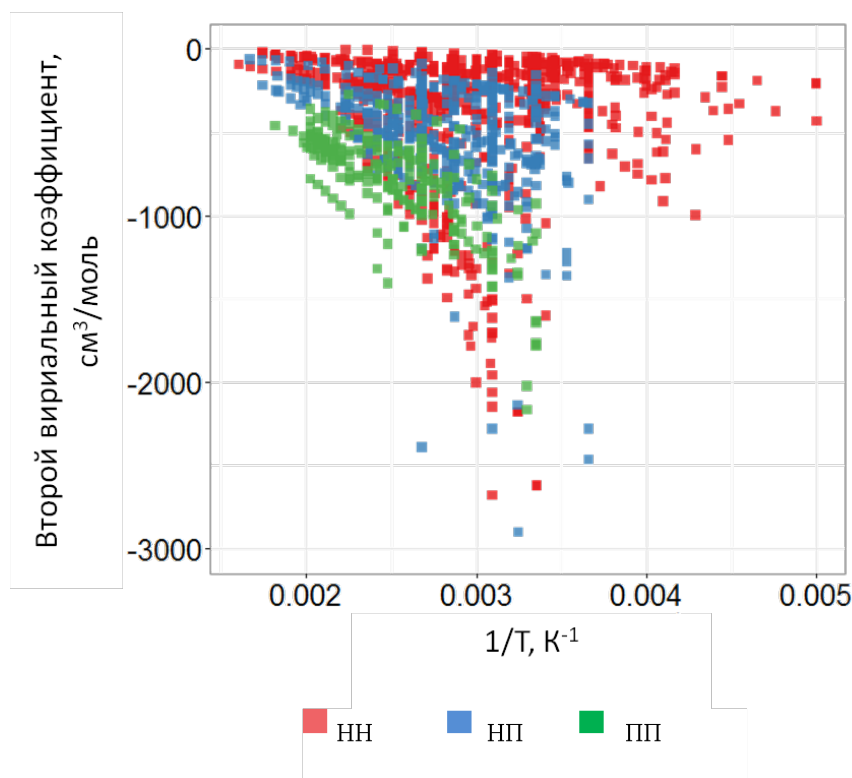


Рис.16. Зависимость вторых вириальных кросс-коэффициентов от обратной температуры.

(НН – смесь неполярное-неполярное, НП – смесь неполярное-полярное, ПП – смесь полярное-полярное)

Конечная выборка включала в себя 130 смесей, более чем 1500 точек при различных температурах (рис. 20). Для данных также наблюдается незначительная кластеризация, менее явно выраженная, чем для вторых вириальных коэффициентов индивидуальных соединений. Смеси полярных компонентов характеризуются наиболее низкими значениями вторых вириальных коэффициентов и наибольшим наклоном кривых, в то время как для смесей неполярных компонентов в среднем характерны более высокие значения. Наименьшее число данных для полярных смесей, что объясняется сложностью их экспериментального изучения. Смеси, для которых было меньше 4 значений, были добавлены во внешнюю тестовую выборку, которая включала 102 смеси с 188 значениями при различных температурах. Около 798 точек – для смесей НН-НН, 302 – для смесей НН-П, и 108 точек для смесей ПП.

В выборке были представлены как разнообразные классы органических соединений, так и различные типы смесей (рис.21). Наибольшее количество смесей типа неполярное соединение – неполярное соединение, наименее представленный тип смесей – смеси типа полярное соединение – полярное соединение. Ограниченность данных для полярных смесей связана со сложностью их экспериментального изучения. Линейная корреляция B_{12} с обратной температурой отличалась высокими значениями ($R^2 > 0.9$) коэффициента корреляции, что позволило использовать предложенную методологию двухслойного моделирования.

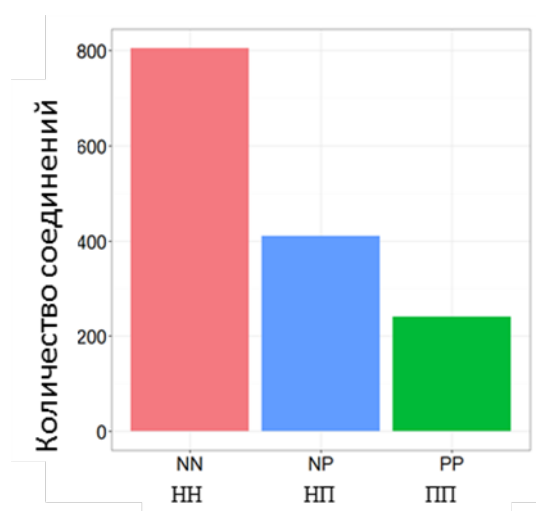


Рис. 17. Количество точек для различных типов смесей в выборке.

3.6.02. Результаты и обсуждение

Был построен ряд моделей [134] с использованием различных статистических методов, результаты консенсусного моделирования представлены на рис. 22. Качество линейной аппроксимации было достаточно высоким с коэффициентом детерминации около 0.90. Для коэффициентов a и b наилучшие модели были получены методом GBM с R^2_{sf} соответственно 0.52 и 0.56. Для наилучших моделей для B_{12} , построенных методом GBM, R^2_{sf} равен 0.75, $RMSE_{sf} = 253.1 \text{ см}^3/\text{моль}$. Для внешней тестовой выборки $R^2_{ts} = 0.52$, $RMSE_{ts} = 341.5 \text{ см}^3/\text{моль}$; в границах DA $R^2_{ts} = 0.65$, $RMSE_{ts} = 224.5 \text{ см}^3/\text{моль}$ (при этом покрытие тестовой выборки DA модели равно 92%). Характеристики полученных моделей позволяют включить их в

экспертный комплекс прогнозирования термодинамических свойств органических соединений.

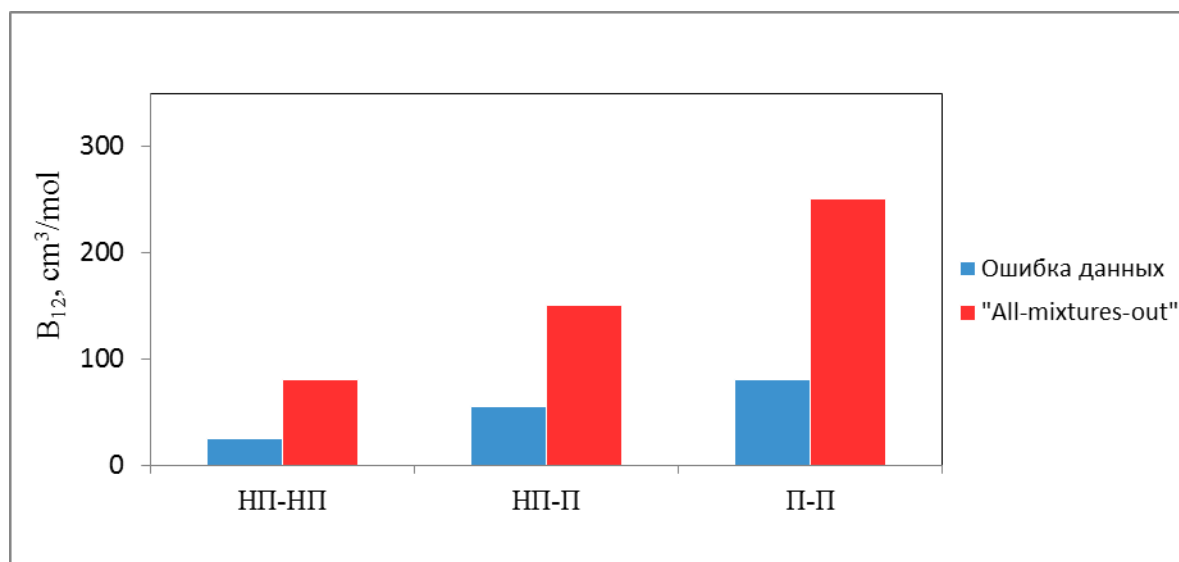


Рис. 18. Результаты консенсусного моделирования

Ошибка прогноза при 5-кратной внешней кросс-валидации увеличивается от неполярных смесей к полярным симбатно с ошибками данных. Для неполярных смесей ошибка данных составляет $25 \text{ см}^3/\text{моль}$, в то время как ошибка прогноза $80 \text{ см}^3/\text{моль}$. Средняя ошибка для смесей НП-П выше – около $55 \text{ см}^3/\text{моль}$, средняя ошибка прогноза – $150 \text{ см}^3/\text{моль}$. Итак, наибольшая ошибка данных для полярных смесей – около $90 \text{ см}^3/\text{моль}$, ошибка прогноза также наибольшая – около $250 \text{ см}^3/\text{моль}$. Диаграмма «наблюдаемое-предсказанное» (рис. 23) показывает общее высокое соответствие наблюдаемых и предсказанных значений, кроме некоторых точек выбросов.

Большинство точек выбросов принадлежат двум смесям: ацетонитрил-этаналь и бензол-перфторбензол. Более подробный анализ экспериментальных данных показывает, что данные для смеси ацетонитрил-этаналь находятся в довольно узком температурном диапазоне (от 313.15 К до 373.15 К), ошибка данных составляет около $150 \text{ см}^3/\text{моль}$. При этом значения температур позволяют предполагать

возможность протекания реакций в смеси (например, конденсации). Смесь бензол-перфторбензол также отличается нестандартным поведением при фазовых переходах (наличием двух азеотропов на фазовой кривой). Недавние исследования [135] показали, что в смеси такого рода за счет слабых взаимодействий при определенных условиях возможно образование устойчивого комплекса.

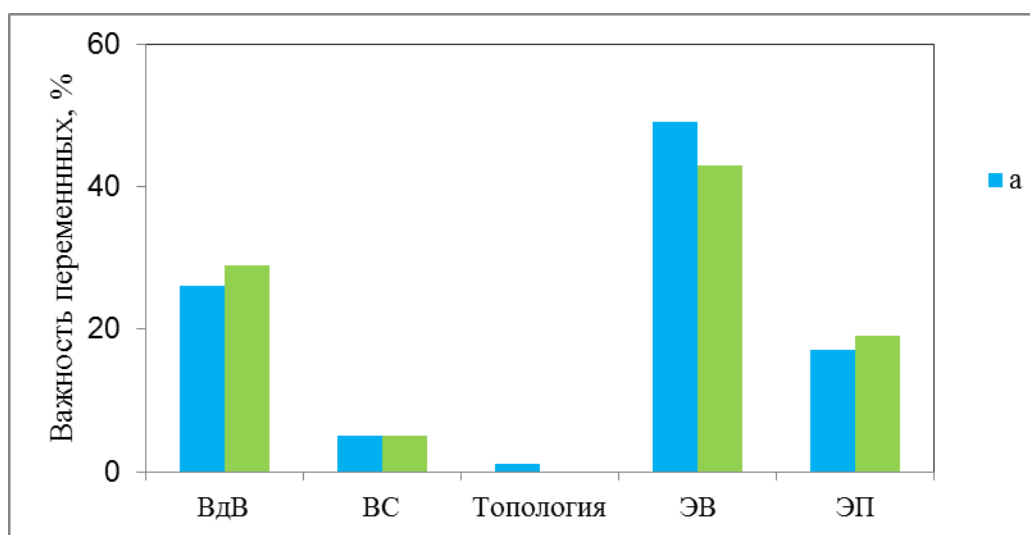


Рис. 19. Вклады различных параметров в изучаемое свойство. (ВС – водородная связь, ЭВ – электростатические взаимодействия, ВДВ – вандерваальсовы взаимодействия, ЭП – электронная поляризуемость).

Использование взвешенных дескрипторов позволяет проводить интерпретацию моделей. Интерпретация для консенсусной двухслойной модели показывает преобладающий вклад электростатических факторов, немного меньший вклад вносят вандерваальсовы взаимодействия, и третьим по важности наиболее влияющим фактором является электронная поляризуемость. Наблюдается схожая с индивидуальными веществами картина – важности различных факторов для *a* и *b* практически идентичны. Как и для индивидуальных соединений, для смесей первым по важности фактором являются электростатические взаимодействия, но за ними (в отличие от индивидуальных соединений) по важности следуют вандерваальсовы взаимодействия, которые для смесевых взаимодействий, по-видимому, играют более важную роль, чем поляризуемость. В отличие от индивидуальных соединений для

смесей вклад электронной поляризуемости наименьший. Второй вириальный кросс-коэффициент напрямую связан с энергией разнородных межмолекулярных взаимодействий $U(R_{12})$, которая в данном случае описывает взаимодействие молекулы 1 с молекулой 2 в терминах межмолекулярного расстояния R_{12} :

$$B_{12} = -2\pi N_A \int_0^\infty \left(e^{-\frac{U(R_{12})}{kT}} - 1 \right) R_{12} dR_{12}$$

Где N_A – число Авогадро.

Из теории межмолекулярных взаимодействий известно, что на больших расстояниях молекулы притягиваются в связи с дисперсионными взаимодействиями, а на малых расстояниях превалирует отталкивание, связанное с обменными взаимодействиями. В то же время постоянные дипольные моменты обладают наибольшим влиянием. Таким образом, рассчитанные значения важностей переменных согласуются с общетеоретическим описанием второго вириального кросс-коэффициента. В то же время способность атомов к образованию водородных связей обладает незначительной важностью, что может быть объяснено малым количеством полярных смесей, способных к образованию таких связей, в изучаемой выборке. Факторы, не связанные с межмолекулярными взаимодействиями, такие как топология или молекулярная масса, имеют незначительное влияние.

Из полученных моделей были восстановлены температурные кривые для B_{12} (рис. 25). Наилучшим образом воспроизводятся температурные кривые для неполярных смесей и смесей НП-П. Высокие ошибки для полярных смесей может быть объяснена как высокой ошибкой данных и малым числом точек, так и широким численным диапазоном значений для таких смесей. Результат расчета – гладкая кривая, что свидетельствует про корректное отображение теоретической зависимости.

Результаты структурной интерпретации приведены на рис. 26. Для коэффициента a для смесей, как и для индивидуальных соединений, наиболее влияющими являются фрагменты с ненасыщенными двойными/тройными углерод-углеродными связями, а также наличие в смеси одного ароматического и одного галогенсодержащего компонента, что может приводить к образованию

относительно стабильных π -комплексов. Для коэффициента b кроме ароматических и галогенсодержащих фрагментов важный вклад также вносят разветвленные углеводородные фрагменты, что может быть связано с их влиянием на суммарный объем смеси, что в общем повторяет тенденции влияния для индивидуальных соединений. Интересным является влияние наличия в смеси ароматического компонента и галогенсодержащего компонента, поскольку это может приводить к возможному образованию пи-комплексов.

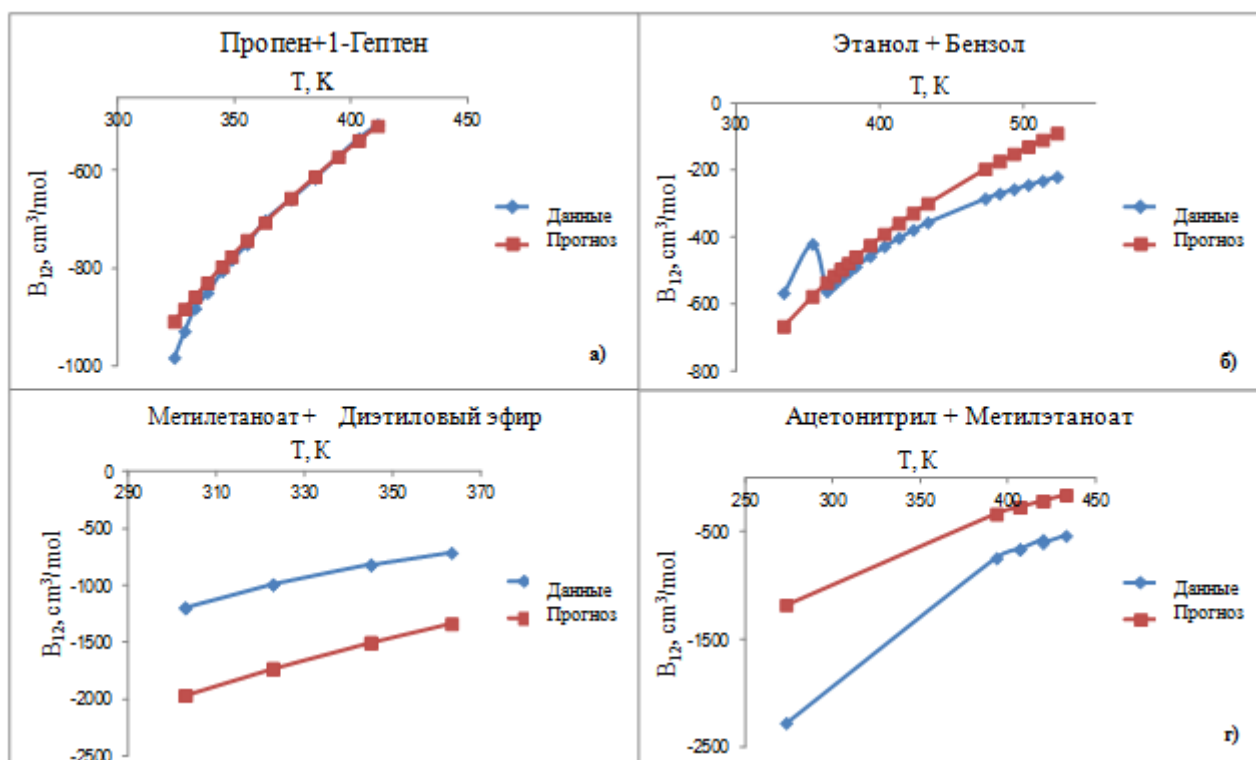


Рис. 20. Примеры восстановленных температурных кривых B_{12} для смесей разного типа (а – смесь НП-НП; б – смесь НП-ПЛ; в, г – ПЛ-ПЛ).

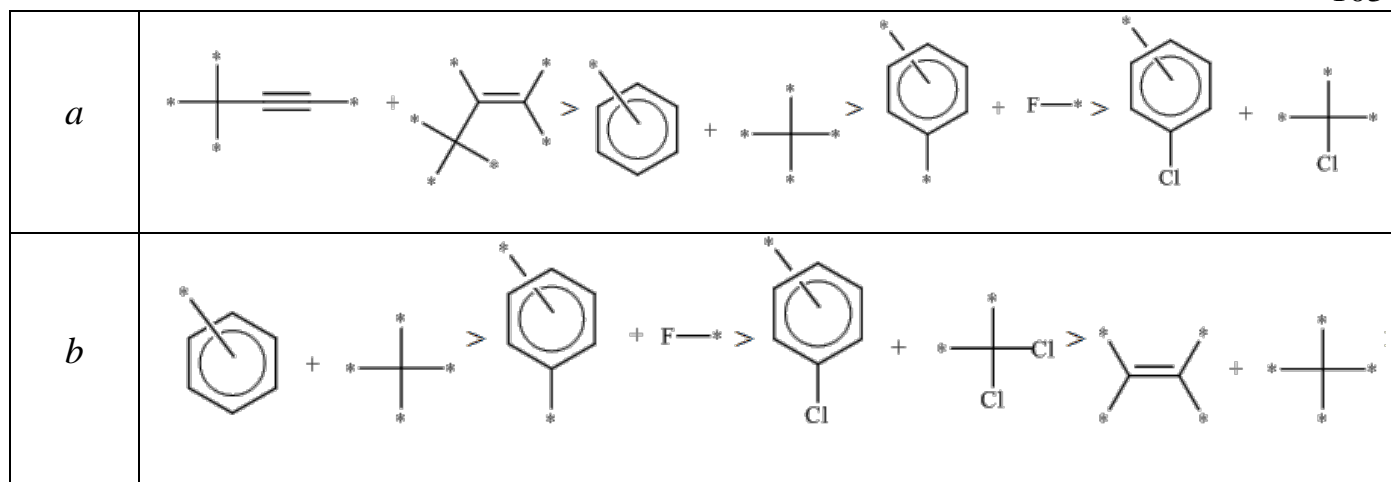


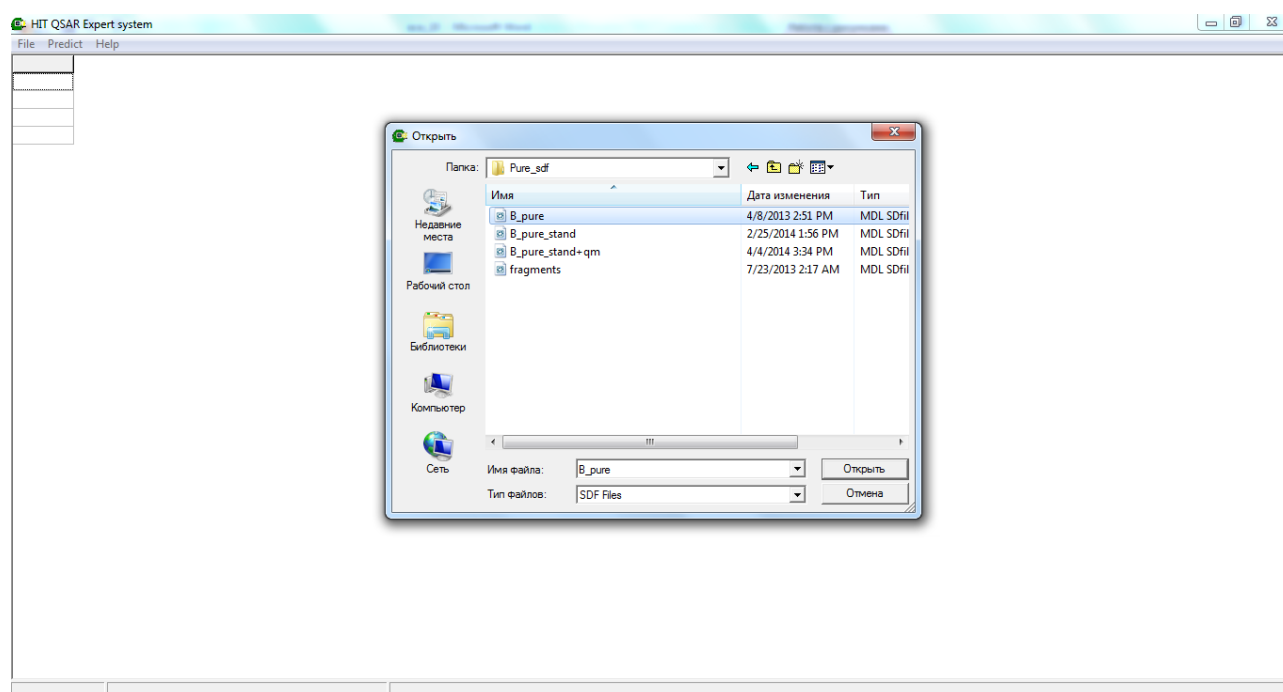
Рис. 21. Влияние некоторых фрагментов смесей органических соединений на второй вириальный кросс-коэффициент (B_{12}).

РАЗДЕЛ 4.

ЭКСПЕРТНАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ КРИТИЧЕСКИХ СВОЙСТВ И ВТОРЫХ ВИРИАЛЬНЫХ КОЭФФИЦИЕНТОВ ОРГАНИЧЕСКИХ ВЕЩЕСТВ И ИХ БИНАРНЫХ СМЕСЕЙ

Экспертная система представляет собой компьютерную программу, способную заменить специалиста-эксперта в разрешении проблемной ситуации. Особо широкое распространение экспертные системы получили при использовании статистических предиктивных моделей.

Для дальнейшего использования полученные в работе модели были объединены в единую экспертную систему, которая представляет собой программное обеспечение с визуальной оболочкой, совместимое с операционными системами Windows и Linux. Экспертная система рассчитана на широкий круг специалистов и обладает удобным, интуитивно понятным интерфейсом. Загрузка данных в экспертную систему происходит посредством диалога загрузки и выглядит следующим образом:



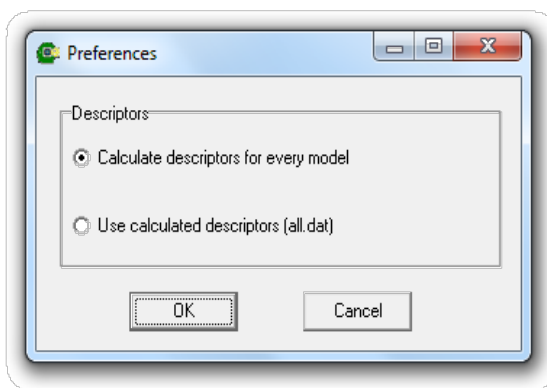
Программная система интуитивно понятна в управлении и предоставляет следующий развернутый функционал:

- 1) Возможность загрузки новых молекул из *.sdf файла;
- 2) Возможность просмотра молекул, входящих в обучающие выборки экспертной системы;
- 3) Расчет интегральных и симплексных фрагментных дескрипторов молекулярной структуры;
- 4) Прогноз свойств новых ранее неизученных молекул, расчет консенсусного прогноза по нескольким моделям;
- 5) Определение корректности предсказания путем проверки области применения моделей, сохранение результатов по области применения;
- 6) Экспорт результатов в отдельный *.csv файл.

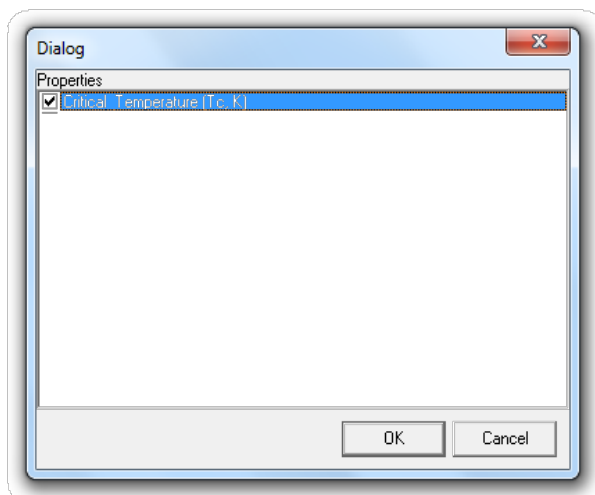
Окно загрузки *.sdf файла масштабируемо.

Используемые модели хранятся в отдельной папке в программном каталоге экспертной системы, однако существует возможность загрузки дополнительных моделей в программу. Таким образом, как исследователи могут обновлять модели в программе, так и продвинутый пользователь.

После загрузки пользователем молекул происходит запуск расчета дескрипторов. Существует возможность, как заново рассчитать дескрипторы, так и использовать предварительно рассчитанные:



Затем пользователь выбирает модели по интересующему его свойству из списка представленных моделей. В списке указано название свойства, а также краткое обозначение и единицы измерения. В данном случае пользователь выбирает модели для критического свойства.



Результат прогноза может быть, как просмотрен в основном окне программы, так и экспортирован в виде текстового файла с разделителями для дальнейшего использования в расчетах, либо как референтные данные.

HIT QSAR Expert system		
File Predict Help		
	Predicted (Tc, K)	Inside DA
v0001_Brom	345.791	true
v0002_Chlor	330.582	true
v0011_Trich	504.587	true
v0101_N,N-D	556.757	true
v0102_Trime	594.23	true
v0103_2,2,3	468.956	true
v0104_Triflu	468.577	true
v0105-Octal	449.743	true
v0106-Octal	427.359	true
v0107-Deca	528.497	true
v0108_1,1,1	563.129	true
v0109_1,1,1	608.556	true
v0110_2-Chl	565.715	true
v0012_Triflu	416.051	true
v0012_Triflu	539.565	true
v0111_Furar	438.797	true
v0112_Thiop	453.052	true
v0113_1H-P	448.773	true
v0114_1,3-B	460.148	true
v0115_1-But	435.955	true

РАЗДЕЛ 5.

АЛГОРИТМЫ И ПРОГРАММЫ, ИСПОЛЬЗОВАННЫЕ ДЛЯ РЕШЕНИЯ ЗАДАЧ QSPR.

5.1. “Lattice&Simplex Modelling”

Программа "Lattice & Simplex Modelling" входит в разработанный в отделе молекулярной структуры Физико-химического института им. А.В. Богатского НАН Украины программный комплекс "*THEORCHEM SOFTWARE*", написанный в среде визуального программирования BORLAND DELPHI 7.0 и предназначенный для эксплуатации под управлением операционной системы MS Windows. В комплексе выполнена графическая визуализация молекул, их решеточных моделей, полей и т.п. Программный комплекс "*THEORCHEM SOFTWARE*" состоит из двух взаимосвязанных программ "Lattice & Simplex Modelling" и "Methods of Data Analysis".

Программа "Lattice & Simplex Modelling" представляет собой программный инструмент для исследования трехмерной структуры молекулярных моделей. Она предназначена для создания выборки молекул, их суперпозиции (только для решеточных моделей), расчета симплексных, интегральных, гармонических и решеточных параметров, а также их графического отображения.

Возможности программы:

1) Для построения исследуемой серии молекул могут использоваться структурные файлы большинства программ молекулярного моделирования:

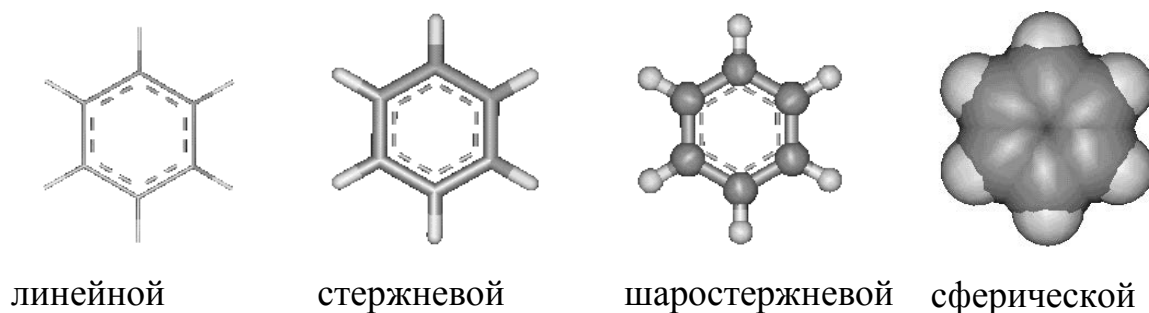
- SYBYL (TRIPOS) mol2- формат, применяющийся в программах фирмы Tripos Associates;
- SYBYL формат, в основном для поддержки устаревших программ, таких как PcModel компании Serena Software и Alchemy II фирмы Tripos Associates;
- MDL mol – широко распространенный формат, разработанный фирмой Molecular Design Limited; поддерживается ChemOffice, HyperChem; IsisDraw, Weblab Viewer.

- HIN, HCS - внутренний формат файлов программного комплекса HyperChem фирм Hypercube, Inc и Autodesk, Inc.
- Protein Data Bank (PDB) - усеченный формат хранения белковых структур;
- Card Coordinate - распространенный в настоящее время формат; поддерживается ChemOffice, HyperChem;
- MSI mol – формат хранения молекулярных данных фирмы ChemNote™; поддерживается ChemOffice;
- Standart Molecular Data (SMD), MacroModel Structure Files – распространенные форматы хранения молекулярных данных; поддерживаемые ChemOffice.

2) Расчет практически неограниченного количества структурных характеристик для каждой из молекул исследуемой серии (интегральные, симплексные, локальные, полевые, гармонические, информационные параметры, параметры СП/СР). Возможность настройки процедуры расчета дескрипторов.

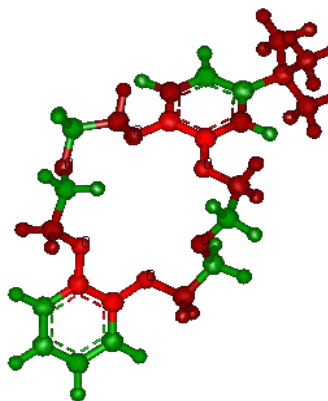
- Широкие возможности настройки визуализации молекул, решеточных моделей, молекулярных полей и т.д.

Изображение молекулы в виде моделей:



- Изображение ячеек в виде моделей:
- Создание ортогографических и перспективных проекций, стерео изображений, регулировка цвета и положения источника света, выбор качества изображения, отражение текстовых меток атомов и ячеек, различные режимы управления цветом.
- Визуализация срезов различных молекулярных полей.

- Вращение и изменение размеров рассматриваемых объектов.
- 3) Редактирование, удаление и добавления различных атомных свойств для всех элементов Периодической системы.
 - 4) Копирование изображения в буфер обмена (Clipboard), его печать, а также сохранения в графических форматах BMP и JPEG.
 - 5) Работа программы в качестве DDE-сервера, принимающего информацию о ячейках, влияющих на изучаемое свойство.
 - 6) В программу включен модуль расчета парциальных зарядов по методу Джолли-Пэрри [73]
 - 7) Реализован быстрый прогноз активности соединений (с указанием влияния каждого атома на исследуемую активность), как линейной комбинации вкладов симплексов, полученных в результате построения по результатам PLS-моделей в программ "Methods of Data Analysis".
 - 8) В программе предусмотрена возможность кодирования цветом атомов и молекулярных фрагментов в зависимости от направления и степени их влияния на исследуемую активность (свойство), что, в последствии, позволяет проводить молекулярный дизайн новых веществ с комплексом заданных свойств:



5.2. Открытая программная среда вычислений R

R — язык программирования для статистической обработки данных и работы с графикой, а также свободная программная среда вычислений с открытым исходным кодом [4]. Изначально R был разработан сотрудниками статистического факультета Оклендского университета Россом Айхэкой и Робертом Джентлменом

(первая буква их имён — R), на момент 2011 года язык и среда поддерживаются и развиваются организацией R Foundation.

R широко используется как статистическое программное обеспечение для анализа данных и фактически стал стандартом для статистических программ.

В R используется интерфейс командной строки, хотя доступны и несколько графических интерфейсов пользователя, например RStudio.

R поддерживает широкий спектр статистических и численных методов и обладает хорошей расширяемостью с помощью пакетов. Пакеты представляют собой библиотеки для работы специфических функций или специальных областей применения. В базовую поставку R включен основной набор пакетов, а всего по состоянию на 2013 год доступно более 4000 пакетов.

Ещё одной особенностью R являются графические возможности, заключающиеся в возможности создания качественной графики.

5.3. Python, OpenBabel, pybel

Python – высокоуровневый язык программирования общего назначения. Python поддерживает несколько парадигм программирования, в том числе объектно-ориентированное и функциональное. Основные черты – динамическая типизация, автоматическое управление памятью, поддержка многопоточных вычислений и удобные высокоуровневые структуры данных. Он распространяется под свободной лицензией Python Software Foundation License, позволяющей использовать его без ограничений в любых приложениях, включая проприетарные.

Python — активно развивающийся язык программирования, новые версии (с добавлением/изменением языковых свойств) выходят примерно раз в два с половиной года. Python также один из наиболее популярных языков программирования, используемых в хемоинформатике.

Open Babel – это программный химический инструментарий, созданный для конвертации различных химических форматов и работы с химическими данными. Это открытый проект, который делится на две основные части:

- Программы, готовые к использованию, например, для конвертации форматов, поиска, модификации и анализа химических файлов;
- Набор программиста – набор функций для полноценной разработки химических программ.

Rybel – реализация Open Babel для Python. Поскольку скриптовые языки, такие как Python, идеально подходят для многих программируемых задач в хемоинформатике, например, анализ данных или получение информации из файла. Rybel включает в себя обертку для главных функций, т.е. чтение и запись молекулярных файлов, расчет молекулярных фингерпринтов. Rybel имеет свой собственный формат молекул, который может быть легко конвертирован в OpenBabel OBMol. В общем, Rybel позволяет хемоинформатику быстро разрабатывать скрипты на Python для работы с химической информацией. Является кросс-платформенным обеспечением с открытым кодом.

ВЫВОДЫ

Разработана система моделей, что адекватно прогнозирует критические свойства органических веществ и их бинарных смесей. Показана эффективность применения новых дескрипторов межмолекулярных взаимодействий для построения моделей свойств, напрямую зависящих от энергии взаимодействия между молекулами. Показано, что 2D-дескрипторы способны описывать сложные фундаментальные свойства, прогнозировать их с ошибкой близкой к экспериментальной.

1. Использование «псевдосмесевых» дескрипторов в QSPR моделях позволяет уменьшить среднеквадратичную ошибку прогнозирования критических свойств и вторых вириальных коэффициентов индивидуальных органических соединений различных классов. Оценка важности дескрипторов парных взаимодействий в QSPR моделях по критическим свойствам смесей показала, что наибольший вклад эти дескрипторы вносят в описание критических давлений - около 30%. Для критических объемов наоборот их вклад не является значимым - около 1%.

2. На критические свойства веществ и их смесей наибольшее влияние оказывают электростатические и вандерваальсовы взаимодействия. Для критических температур превалирующую роль играют электростатические факторы, в то время как для критических объемов существенно возрастает роль вандерваальсовых взаимодействий. Анализ важности переменных для критических свойств смесей демонстрирует ведущую роль электронной поляризуемости для критических объемов (56%) и незначительное влияние электронной поляризуемости на критические давления (4%). Показано, что фактор Питцера зависит не только от асферичности молекулярной формы (как считалось ранее), а прежде всего от размера молекулы.

3. Для всех критических свойств наибольший вклад вносят разветвленные углеводородные фрагменты. Для критических температур наибольшее

положительное влияние оказывают циклические фрагменты и фрагменты, содержащие карбонильную группу, отрицательный – фрагменты, содержащие атомы фтора. Это можно объяснить, исходя из того, что критическая температура связана с энергией межмолекулярного взаимодействия и повышается при росте способности молекул к взаимному притяжению. Для критического давления наибольшее положительное влияние оказывает дисульфидная группа, алифатический фтор, первичная аминогруппа, тройная углерод-углеродная связь и алифатическая гидроксогруппа. Для критических объемов влияние фрагментов обусловлено размерами фрагментов. Для смесей наблюдаются аналогичные тенденции.

4. «Псевдосмесевые» QSPR модели для вторых вириальных коэффициентов индивидуальных веществ проявили высокую прогнозирующую способность, ошибка прогноза в среднем на порядок ниже, чем для других термодинамических методов. Благодаря использованию несвязанных симплексов и дескрипторов парных взаимодействий рассчитаны вторые вириальные кросс-коэффициенты (B_{12}), которые непосредственно характеризуют разнородные межмолекулярные взаимодействия в смесях. Рассчитанные кривые зависимости $B_{12} = f(T)$ в большинстве случаев адекватно воспроизводят данные эксперимента.

5. Структурная интерпретация параметров уравнения Ван-дер-Ваальса a и b , на основе которых рассчитываются вторые вириальные коэффициенты, показала следующие тенденции. Наибольшим положительным влиянием на коэффициент a обладают фрагменты с ненасыщенными связями углерод-углерод и галогенсодержащие фрагменты; наибольший отрицательный вклад вносят ароматические и неароматических циклические фрагменты и кислородсодержащие фрагменты, которые способствуют ассоциации молекул (за счет образования водородных связей, стекинг-взаимодействий, т.д.) и снижают энергию отталкивания. Объемные и / или способны к ассоциации фрагменты имеют положительное влияние на коэффициент b , который зависит от молекулярного объема. Для смесей влияние структурных фрагментов аналогично, при этом для коэффициента a является влияющим наличие в смеси одного ароматического

компонента и одного галогенсодержащего компонента, что может приводить к образованию относительно стабильных комплексов.

ЛИТЕРАТУРА

1. McGlashan M. L. Chemical thermodynamics/ McGlashan M. L. – Royal Society of Chemistry, 1973. – 350 C. – (T. 1).
2. Bett K. E. Thermodynamics for chemical engineers. / K. E. Bett, J. S. Rowlinson, G. Saville – Mit Press, 1975 – 425 C.
3. Moran M.J. Fundamentals of Engineering Thermodynamics. / Moran M.J. – H. N. S. – Wiley, 2000 – 338 C.
4. Hicks C. Gas-liquid critical properties of binary mixtures/ C. Hicks, C. Young // Chemical Reviews. – 1975. – T. 75, № 2. – C. 119-175.
5. Forman C. J. Critical temperatures and pressures of organic compounds/ C. J. Forman, G. Thodos// AIChE Journal. – 1960. – T. 6, № 2. – C. 206-209.
6. Thodos G. Critical constants of saturated aliphatic hydrocarbons/ Thodos G.// AIChE Journal. – 1955. – T. 1, № 2. – C. 168-173.
7. Lydersen A. L. Estimation of Critical Properties of Organic Compounds by the Method of Group Contributions. / A. L. Lydersen, R. A. Greenkorn, O. A. Hougen – University of Wisconsin, 1955 – 249 C.
8. Joback K. G. Estimation of pure-component properties from group-contributions/ K. G. Joback, R. C. Reid // Chemical Engineering Communications. – 1987. – T. 57, № 1-6. – C. 233-243.
9. Constantinou L.. New group contribution method for estimating properties of pure compounds/ L. Constantinou, R. Gani// AIChE Journal. – 1994. – T. 40, № 10. – C. 1697-1710.
10. Wilson G. Critical constants T_c , P_c , estimation based on zero, first and second order methods/ G. Wilson, L. Jasperson// AIChE Meeting –, 1996. – C.115.
11. Marrero-Morejón J. Estimation of pure compound properties using group-interaction contributions/ J. Marrero-Morejón, E. Pardillo-Fontdevila// AIChE journal. – 1999. – T. 45, № 3. – C. 615-621.

12. Poling B. E. The properties of gases and liquids. / B. E. Poling, J. M. Prausnitz, J. C. O'Connell – McGraw-Hill, 2001 – 458 C.
13. Jalowka J. W. Group contribution method to predict critical temperature and pressure of hydrocarbons/ J. W. Jalowka, T. E. Daubert// Industrial & Engineering Chemistry Process Design and Development. – 1986. – T. 25, № 1. – C. 139-142.
14. Daubert T. E. Prediction of critical temperature and pressure of organic compounds by group contribution/ T. E. Daubert, R. Bartakovits// Industrial & engineering chemistry research. – 1989. – T. 28, № 5. – C. 638-641.
15. Mason W. C. Properties of Gases, Liquids, and Solutions: Principles and Methods/ W. C. Mason – Academic Press, 2013 – 494 C.
16. Determination of critical properties and acentric factors of pure compounds using the artificial neural network group contribution algorithm/ [Gharagheizi F., Eslamimanesh A., Mohammadi A. H., Richon D.]// Journal of Chemical & Engineering Data. – 2011. – T. 56, № 5. – C. 2460-2476.
17. Chueh C. Vapor-liquid equilibria at high pressures: Calculation of critical temperatures, volumes, and pressures of nonpolar mixtures/ C. Chueh, J. Prausnitz// AIChE Journal. – 1967. – T. 13, № 6. – C. 1107-1113.
18. Li L. Estimation of critical properties of binary mixtures using group contribution methods/ L. Li, E. Kiran// Chemical Engineering Communications. – 1990. – T. 94, № 1. – C. 131-141.
19. Klincewicz K. Estimation of critical properties with group contribution methods/ K. Klincewicz, R. Reid // AIChE Journal. – 1984. – T. 30, № 1. – C. 137-142.
20. Li C. Critical temperature estimation for simple mixtures/ Li C.// The Canadian Journal of Chemical Engineering. – 1971. – T. 49, № 5. – C. 709-710.
21. Liu Z. Y. Prediction of critical pressure of dilute multicomponent mixtures/ Liu Z. Y.// AIChE journal. – 1998. – T. 44, № 7. – C. 1709-1712.
22. Correlation of the critical properties of alkanes and alkanols/ [Teja A., Lee R., Rosenthal D., Anselme M.] // Fluid Phase Equilibria. – 1990. – T. 56. – C. 153-169.
23. Schick L. On the correlation of critical properties of mixtures/ L. Schick, J. Prausnitz// AIChE Journal. – 1968. – T. 14, № 4. – C. 673-673.

24. Heidemann R. A. The classical theory of critical points/ Heidemann R. A.// Supercritical FluidsSpringer, 1994. – C. 39-64.
25. Sadus R. J. Predicting the Gas-Liquid Critical Properties of Binary Mixtures: An Alternative to Conventional Mixing Rules/ Sadus R. J.// Berichte der Bunsengesellschaft für physikalische Chemie. – 1992. – T. 96, № 10. – C. 1454-1459.
26. Soave G. Equilibrium constants from a modified Redlich-Kwong equation of state/ Soave G.// Chemical Engineering Science. – 1972. – T. 27, № 6. – C. 1197-1203.
27. Michelsen M. L. Calculation of critical points from cubic two-constant equations of state/ M. L. Michelsen, R. A. Heidemann// AIChE journal. – 1981. – T. 27, № 3. – C. 521-523.
28. Michelsen M. L. The isothermal flash problem. Part I. Stability/ Michelsen M. L.// Fluid phase equilibria. – 1982. – T. 9, № 1. – C. 1-19.
29. Michelsen M. L. The isothermal flash problem. Part II. Phase-split calculation/ Michelsen M. L.// Fluid Phase Equilibria. – 1982. – T. 9, № 1. – C. 21-40.
30. Heidemann R. A. The calculation of critical points/ R. A. Heidemann, A. M. Khalil// AIChE journal. – 1980. – T. 26, № 5. – C. 769-779.
31. Abu-Eishah S. Prediction of critical properties of binary mixtures using the PRSV-2 equation of state/ S. Abu-Eishah, N. Darwish, I. Aljundi// International journal of thermophysics. – 1998. – T. 19, № 1. – C. 239-258.
32. Stryjek R. PRSV: An improved Peng—Robinson equation of state for pure compounds and mixtures/ R. Stryjek, J. Vera // The canadian journal of chemical engineering. – 1986. – T. 64, № 2. – C. 323-333.
33. Vetere A. Methods to predict the critical constants of organic compounds/ Vetere A.// Fluid phase equilibria. – 1995. – T. 109, № 1. – C. 17-27.
34. Somayajulu G. The critical constants of long-chain normal paraffins/ Somayajulu G.// International journal of thermophysics. – 1991. – T. 12, № 6. – C. 1039-1062.
35. Comparison of a neural net-based QSAR algorithm (PCANN) with hologram-and multiple linear regression-based QSAR approaches: application to 1, 4-dihydropyridine-based calcium channel antagonists/ [Viswanadhan V. N., Mueller G. A., Basak S. C.,

- Weinstein J. N.] // Journal of chemical information and computer sciences. – 2001. – T. 41, № 3. – C. 505-511.
36. Mattioni B. E. Development of quantitative structure-activity relationship and classification models for a set of carbonic anhydrase inhibitors/ B. E. Mattioni, C. C. Jurs// Journal of chemical information and computer sciences. – 2002. – T. 42, № 1. – C. 94-102.
37. Three-dimensional quantitative structure-permeability relationship analysis for a series of inhibitors of rhinovirus replication/ [Ekins S., Durst G. L., Stratford R. E. et al.] // Journal of chemical information and computer sciences. – 2001. – T. 41, № 6. – C. 1578-1586.
38. Sutter J. M. Prediction of aqueous solubility for a diverse set of heteroatom-containing organic compounds using a quantitative structure-property relationship/ J. M. Sutter, C. C. Jurs// Journal of chemical information and computer sciences. – 1996. – T. 36, № 1. – C. 100-107.
39. Structure-antiviral activity relationship in the series of pyrimidine and purine N-[2-(2-phosphonomethoxy) ethyl] nucleotide analogues. 1. Derivatives substituted at the carbon atoms of the base/ [Holý A., Günter J., Dvorakova H. et al.] // Journal of medicinal chemistry. – 1999. – T. 42, № 12. – C. 2064-2086.
40. Maw H. H. E-state modeling of HIV-1 protease inhibitor binding independent of 3D information/ H. H. Maw, L. H. Hall// Journal of chemical information and computer sciences. – 2002. – T. 42, № 2. – C. 290-298.
41. Gironés X. Using molecular quantum similarity measures under stochastic transformation to describe physical properties of molecular systems/ X. Gironés, R. Carbó-Dorca// Journal of chemical information and computer sciences. – 2002. – T. 42, № 2. – C. 317-325.
42. Dearden J.C. QSAR prediction of environmental toxicity of oxygen- and sulfur-containing heterocycles/ Dearden J.C.// Book QSAR prediction of environmental toxicity of oxygen- and sulfur-containing heterocycles / Editor M.: IBS PRESS, 2003. – C. C. 43-57.

43. Voskresensky O. N. QSAR aspects of flavonoids as a plentiful source of new drugs/ O. N. Voskresensky, A. C. Levitsky// *Current medicinal chemistry*. – 2002. – T. 9, № 14. – C. 1367-1383.
44. Seel M. HQSAR—a highly predictive QSAR technique based on molecular holograms/ M. Seel, D. Turner, C. Wilett// *QSAR*. – 1999. – T. 18. – C. 245-252.
45. Pharmacophore model prediction, 3D-QSAR and molecular docking studies on vinyl sulfones targeting Nrf2-mediated gene transcription intended for anti-Parkinson drug design/ [Athar M., Lone M. Y., Khedkar V. M., Jha C. C.] // *Journal of Biomolecular Structure and Dynamics*. – 2015. – C. 1-16.
46. Thareja S. Steroidal 5 α -Reductase Inhibitors: A Comparative 3D-QSAR Study Review/ Thareja S.// *Chemical reviews*. – 2015. – T. 115, № 8. – C. 2883-2894.
47. Three-dimensional quantitative structure–activity relationship (3D-QSAR) analysis and molecular docking-based combined in silico rational approach to design potent and novel TRPV1 antagonists/ [Bhadoriya K. S., Sharma M. C., Jain S. V. et al.] // *Medicinal Chemistry Research*. – 2013. – T. 22, № 5. – C. 2312-2327.
48. Cramer R. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins/ R. D. Cramer, D. E. Patterson, J. D. Bunce// *Journal of the American Chemical Society*. – 1988. – T. 110, № 18. – C. 5959-5967.
49. Toward a pharmacophore for drugs inducing the long QT syndrome: insights from a CoMFA study of HERG K⁺ channel blockers/ [Cavalli A., Poluzzi E., De Ponti F., Recanatini M.] // *Journal of medicinal chemistry*. – 2002. – T. 45, № 18. – C. 3844-3853.
50. Hierarchic system of QSAR models (1D–4D) on the base of simplex representation of molecular structure/ [Kuz'min V. E., Artemenko A. G., Polischuk C. G et al.] // *Journal of molecular modeling*. – 2005. – T. 11, № 6. – C. 457-467.
51. Construction of 3D-QSAR models using the 4D-QSAR analysis formalism/ [Hopfinger A., Wang S., Tokarski J. S. et al.] // *Journal of the American Chemical Society*. – 1997. – T. 119, № 43. – C. 10509-10524.
52. Nelder J. A. Generalized linear models. / J. A. Nelder, R. Baker – Wiley Online Library, 1972 – 135 C.

53. Haenlein M. A beginner's guide to partial least squares analysis/ M. Haenlein, A. M. Kaplan// Understanding statistics. – 2004. – Т. 3, № 4. – С. 283-297.
54. Sheridan R. C. Extending the trend vector: the trend matrix and sample-based partial least squares / R. C. Sheridan, R. B. Nachbar, B. L. Bush// Journal of computer-aided molecular design. – 1994. – Т. 8, № 3. – С. 323-340.
55. Support vector regression machines/ [Drucker H., Burges C. J., Kaufman L. et al.] // Advances in neural information processing systems. – 1997. – Т. 9. – С. 155-161.
56. Cortes C. Support-vector networks / C. Cortes, V. Vapnik // Machine learning. – 1995. – Т. 20, № 3. – С. 273-297.
57. An introduction to kernel-based learning algorithms/ [Müller K.-R., Mika S., Rätsch G. et al.] // Neural Networks, IEEE Transactions on. – 2001. – Т. 12, № 2. – С. 181-201.
58. Application of Random Forest Approach to QSAR Prediction of Aquatic Toxicity / [Polishchuk C. G., Muratov E. N., Artemenko A. G. et al.] // Journal of Chemical Information and Modeling. – 2009. – Т. 49, № 11. – С. 2481-2488.
59. Breiman L. Random Forests / Breiman L. // Machine Learning. – 2001. – Т. 45, № 1. – С. 5-32.
60. Tetko I. V. Efficient partition of learning data sets for neural network training / I. V. Tetko, A. E. Villa// Neural Networks. – 1997. – Т. 10, № 8. – С. 1361-1374.
61. Livingstone D. J. Data modelling with neural networks: advantages and limitations / D. J. Livingstone, D. T. Manallack, I. V. Tetko// Journal of computer-aided molecular design. – 1997. – Т. 11, № 2. – С. 135-142.
62. Ripley B. D. Pattern recognition and neural networks. / Ripley B. D. – Cambridge university press, 1996 – 115 С.
63. Walters D. E. Case studies of the application of molecular shape analysis to elucidate drug action / D. E. Walters, A. Hopfinger// Journal of Molecular Structure: THEOCHEM. – 1986. – Т. 134, № 3. – С. 317-323.
64. Артеменко А.Г. Анализ связи структура-свойство на основе молекулярных решеточных моделей : дис. ... канд. хим. наук: 02.00.03./ Артеменко Анатолий Григорьевич – Одесса, 2000. – 151 с. с.

65. Amann A. Fundamental principles of molecular modeling. / A. Amann, J. C. Boeyens, W. Gans – Springer Science & Business Media, 2013 – 820 C.
66. Trudeau R. J. Introduction to graph theory. / Trudeau R. J. – Courier Dover Publications, 2013 – 224 C.
67. Consonni V. Handbook of molecular descriptors. Methods and principles in medicinal chemistry. / V. Consonni, R. Todeschini – Weinheim ; New York: Wiley-VCH, 2000. – 667 C. – (Methods and principles in medicinal chemistry; T. 11).
68. De Benedetti C. G. Multiscale quantum chemical approaches to QSAR modeling and drug design/ C. G. De Benedetti, F. Fanelli // Drug discovery today. – 2014. – T. 19, № 12. – C. 1921-1927.
69. Kuz'min V. E. Hierarchical QSAR technology based on the Simplex representation of molecular structure / V. E. Kuz'min, A. G. Artemenko, E. N. Muratov // Journal of computer-aided molecular design. – 2008. – T. 22, № 6-7. – C. 403-21.
70. Ognichenko L. N. New Advances in QSPR/QSAR Analysis of Nitrocompounds: Solubility, Lipophilicity, and Toxicity / Ognichenko L. N., V.E. Kuz'min, L.Gorb // Practical Aspects of Computational Chemistry II. – Springer Netherlands, 2012. – C. 279-334.
71. Per aspera ad astra: application of Simplex QSAR approach in antiviral research / [Muratov E. N., Artemenko A. G., Varlamova E. V. et al.] // Future medicinal chemistry. – 2010. – T. 2, № 7. – C. 1205-1226.
72. Identification of individual structural fragments of N,N'-(bis-5-nitropyrimidyl)dispirotripiperazine derivatives for cytotoxicity and antiherpetic activity allows the prediction of new highly active compounds / [Artemenko A. G., Muratov E. N., Kuz'min V. E. et al.] // The Journal of antimicrobial chemotherapy. – 2007. – T. 60, № 1. – C. 68-77.
73. Jolly W. L. Estimation of Atomic Charges by an Electronegativity Equalization Procedure Calibrated with Core Binding-Energies / W. L. Jolly, W. B. Perry // Journal of the American Chemical Society. – 1973. – T. 95, № 17. – C. 5442-5450.

74. Existing and Developing Approaches for QSAR Analysis of Mixtures / [Muratov E. N., Varlamova E. V., Artemenko A. G. et al.] // *Molecular Informatics*. – 2012. – T. 31, № 3-4. – C. 202-221.
75. Support vector regression based QSPR for the prediction of some physicochemical properties of alkyl benzenes / [Yang S., Lu W., Chen N., Hu Q.] // *Journal of Molecular Structure: THEOCHEM*. – 2005. – T. 719, № 1. – C. 119-127.
76. Grigoras S. A structural approach to calculate physical properties of pure organic substances: the critical temperature, critical volume and related properties / Grigoras S. // *Journal of Computational Chemistry*. – 1990. – T. 11, № 4. – C. 493-510.
77. Turner B. E. Prediction of critical temperatures and pressures of industrially important organic compounds from molecular structure / B. E. Turner, C. L. Costello, C. C. Jurs // *Journal of chemical information and computer sciences*. – 1998. – T. 38, № 4. – C. 639-645.
78. Egolf L. M. Prediction of boiling points and critical temperatures of industrially important organic compounds from molecular structure / L. M. Egolf, M. D. Wessel, C. C. Jurs // *Journal of Chemical Information and Computer Sciences*. – 1994. – T. 34, № 4. – C. 947-956.
79. Godavarthy S. S. A. Improved structure–property relationship models for prediction of critical properties / Godavarthy S. S., Robinson Jr R. L., Gasem K. A. // *Fluid Phase Equilibria*. – 2008. – T. 264, № 1. – C. 122-136.
80. Predictive correlations based on large experimental datasets: Critical constants for pure compounds / [Kazakov A., Muzny C. D., Diky V. et al.] // *Fluid Phase Equilibria*. – 2010. – T. 298, № 1. – C. 131-142.
81. Guggenheim E. A. The principle of corresponding states / Guggenheim E. A. // *The Journal of Chemical Physics*. – 1945. – T. 13, № 7. – C. 253-261.
82. Tsonopoulos C. An empirical correlation of second virial coefficients / Tsonopoulos C. // *AIChE Journal*. – 1974. – T. 20, № 2. – C. 263-272.
83. Hayden J. G. A generalized method for predicting second virial coefficients / Hayden J. G., O'Connell J. C. // *Industrial & Engineering Chemistry Process Design and Development*. – 1975. – T. 14, № 3. – C. 209-216.

84. Tarakad R. R. An improved corresponding states method for polar fluids: correlation of second virial coefficients / Tarakad R. R., Danner R. C. // *AIChE Journal*. – 1977. – Т. 23, № 5. – С. 685-695.
85. McCann D. W. Prediction of second virial coefficients of organic compounds by a group contribution method / D. W. McCann, R. C. Danner // *Industrial & Engineering Chemistry Process Design and Development*. – 1984. – Т. 23, № 3. – С. 529-533.
86. Orbey H. A four parameter Pitzer-Curl type correlation of second virial coefficients / Orbey H. // *Chemical Engineering Communications*. – 1988. – Т. 65, № 1. – С. 1-19.
87. Kiş K. An empirical method for the second virial coefficients of non-standard fluids / K. Kiş, H. Orbey // *The Chemical Engineering Journal*. – 1989. – Т. 41, № 3. – С. 149-154.
88. Abusleme J. A group contribution method for second virial coefficients / J. Abusleme, J. Vera // *AIChE journal*. – 1989. – Т. 35, № 3. – С. 481-489.
89. Lee M.-J. An Improved Model of Second Virial Coefficients for Polar Fluids and Fluid Mixtures / M.-J. Lee, J.-T. Chen // *Journal of chemical engineering of Japan*. – 1998. – Т. 31, № 4. – С. 518-526.
90. Vetere A. An empirical model for calculating second virial coefficients / Vetere A. // *Chemical engineering science*. – 1988. – Т. 43, № 12. – С. 3119-3127.
91. NIST. Recommended prediction methods for pure substances [Електронний ресурс] : [Веб-сайт]. – Електронні дані. – Режим доступу: // www.trc.nist.gov/TDE/TDE_Help/ - (дата звернення 30.03.2016) – Назва з екрана.
92. Xiang H. The new simple extended corresponding-states principle: vapor pressure and second virial coefficient / Xiang H. // *Chemical engineering science*. – 2002. – Т. 57, № 8. – С. 1439-1449.
93. Ben-Naim A. Solvation thermodynamics/ Ben-Naim A. – Springer Science & Business Media, 2013 – 246 С..
94. Недоступ В. И. Расчет термодинамических свойств газов и жидкостей методом идеальных кривых. / В. И. Недоступ, Е. П. Галькевич – Наукова думка, 1986 – 196 С.

95. 'Quasi-Mixture' Descriptors for QSPR Analysis of Molecular Macroscopic Properties. The Critical Properties of Organic Compounds / [Mokshyna E., Nedostup V., Polishchuk C., Kuzmin V.] // Molecular Informatics. – 2014. – Т. 33, № 10. – С. 647-654.
96. Китайгородский А. И. Молекулярные кристаллы. / Китайгородский А. И. – Наука. Гл. ред. физ.-мат. лит., 1971 – 175 С.
97. Uff, a Full Periodic-Table Force-Field for Molecular Mechanics and Molecular-Dynamics Simulations / [Rappe A. K., Casewit C. J., Colwell K. S. et al.] // Journal of the American Chemical Society. – 1992. – Т. 114, № 25. – С. 10024-10035.
98. Mokshyna E. Atom-pair descriptors, temperature-dependent property and binary mixtures: QSPR for second virial cross-coefficients/ E. Mokshyna, C. Polishchuk, V. Nedostup, V. Kuz'min // II Kazan Summer School on Chemoinformatics: тезисы докладов. – Казань, 2015. – С. 40.
99. QSPR Approach to Predict Nonadditive Properties of Mixtures. Application to Bubble Point Temperatures of Binary Mixtures of Liquids / [Oprisiu I., Varlamova E., Muratov E. et al.] // Molecular Informatics. – 2012. – Т. 31, № 6-7. – С. 491-502.
100. Reid R. C. The properties of gases and liquids. / R. C. Reid, J. M. Prausnitz, B. E. Poling – McGraw-Hill, 1987 – 487 С.
101. Virial Coefficients of Pure Gases / [Dymond J., Marsh K., Wilhoit R., Wong K.] – Landolt-Bornstein: Numerical Data and Functional Relationships in Science and Technology, 2002 – 630 С.
102. Mallard E. C. Thermodynamics Research Center N. B. L., M. Frenkel director. Thermodynamics Source Database // NIST Chemistry WebBook, NIST Standard Reference Database Number 69 / E. C. Mallard – National Institute of Standards and Technology, 2013 – 552 С.
103. JChem and InstantJChem Software from. ChemAxon, version 6.1.3. [Електронний ресурс] : [Інтернет-портал]. – Електронні дані. – [Одеса: Український Інтернет холдинг ТОВ "Тенет"] – Режим доступа: www.docs.chemaxon.com/display/docs/Instant+JChem+CT (дата звернення 30.03.2016).
– Назва з екрана.

104. Standardizer. Standardizer, JChem 6.1.3 [Електронний ресурс] : [Інтернет-портал]. – Електронні дані. – [Одеса: Український Інтернет холдинг ТОВ "Тенет"] – Режим доступу: www.docs.chemaxon.com/display/docs/Standardizer+-+Standardizing+Structure+Files/ (дата звернення 30.03.2016). – Назва з екрана.
105. Frank L. E. A statistical view of some chemometrics regression tools / L. E. Frank, J. Friedman // *Technometrics*. – 1993. – Т. 35, № 2. – С. 109-135.
106. Friedman J. H. Greedy function approximation: a gradient boosting machine / Friedman J. H. // *Annals of Statistics*. – 2001. – С. 1189-1232.
107. Friedman J. H. Stochastic gradient boosting / Friedman J. H. // *Computational Statistics & Data Analysis*. – 2002. – Т. 38, № 4. – С. 367-378.
108. Ферстер Э. Методы корреляционного и регрессионного анализа / Э. Ферстер, Б. Ренц // М.: Финансы и статистика. – 1983. – Т. 302.
109. Puzyn T. Recent advances in QSAR studies : methods and applications / T. Puzyn, J. Leszczynski, M. T. D. Cronin // *Challenges and advances in computational chemistry and physics*. – Dordrecht ; New York: Springer, 2010. – С. 423.
110. Quantitative structure–property relationship (QSPR) modeling of normal boiling point temperature and composition of binary azeotropes / [Solov'ev V. C., Oprisiu I., Marcou G., Varnek A.] // *Industrial & Engineering Chemistry Research*. – 2011. – Т. 50, № 24. – С. 14162-14167.
111. Universal approach for structural interpretation of QSAR/QSPR models / [Polishchuk C. G., Kuz'min V. E., Artemenko A. G., Muratov E. N.] // *Molecular Informatics*. – 2013. – Т. 32, № 9-10. – С. 843-853.
112. Fisher R. A. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population / Fisher R. A. // *Biometrika*. – 1915. – С. 507-521.
113. Pearl J. Causal inference in statistics: An overview / Pearl J. // *Statistics Surveys*. – 2009. – Т. 3. – С. 96-146.
114. New methods for separating causes from effects in genomics data / [Statnikov A., Henaff M., Lytkin N. I., Aliferis C. F.] // *Bmc Genomics*. – 2012. – Т. 13.

115. Causal inference methods to assist in mechanistic interpretation of classification nano-SAR models / [Sizochenko N., Rasulev B., Gajewicz A. et al.] // RSC Advances. – 2015. – Т. 5, № 95. – С. 77739-77745.
116. Probabilistic latent variable models for distinguishing between cause and effect / [Stegle O., Janzing D., Zhang K. et al.] // Advances in Neural Information Processing Systems –, 2010. – С. 1687-1695.
117. Robust learning via cause-effect models / [Schölkopf B., Janzing D., Peters J., Zhang K.] // arXiv, No. 1112. – 2011 – С. 225.
118. R: A Language and Environment for Statistical Computing [Електронний ресурс] : [Інтернет-портал]. – Електронні дані. – [Одеса: Український Інтернет холдинг ТОВ "Тенет"] – Режим доступу: www.R-project.org (дата звернення 30.03.2016). – Назва з екрана.
119. Mokshina E. QSPR Modeling of critical parameters of organic compounds belonging to different classes in terms of the simplex representation of molecular structure / E. Mokshina, V. Kuz'min, V. Nedostup // Russian Journal of Organic Chemistry. – 2014. – Т. 50, № 3. – С. 314-321.
120. Мокшина Е. Метод определения критических параметров и ацентрического фактора многоатомных экспериментально неизученных веществ / Е. Мокшина, В. Кузьмин, В. Недоступ // Технические газы. – 2014. № 1. – С. 52-58.
121. AIChE/DIPPR® Project 911 / [Kline A.A., Whitten C.R., Luehrs D.C. et al.] // Amer. Inst. of Chemical Engineers, 2011 – С. 320.
122. DIPPR 801 Database [Електронний ресурс] : [Інтернет-портал]. – Електронні дані. – [Одеса: Український Інтернет холдинг ТОВ "Тенет"] – Режим доступу: www.dippr.byu.edu (дата звернення 30.03.2016). – Назва з екрана.
123. Codessa descriptors [Електронний ресурс] : [Інтернет-портал]. – Електронні дані. – [Одеса: Український Інтернет холдинг ТОВ "Тенет"] – Режим доступу: www.semichem.com/codessa/ (дата звернення 30.03.2016). – Назва з екрана.
124. Kaplan I. G. Intermolecular interactions: physical picture, computational methods and model potentials. / Kaplan I. G. – John Wiley & Sons, 2006 – С. 258.

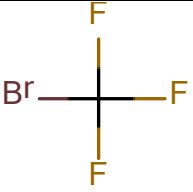
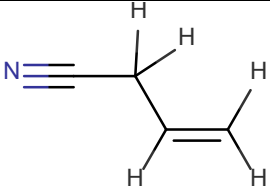
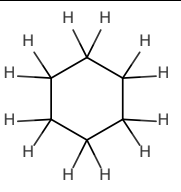
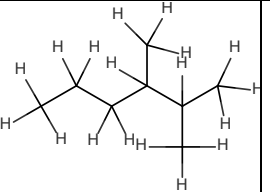
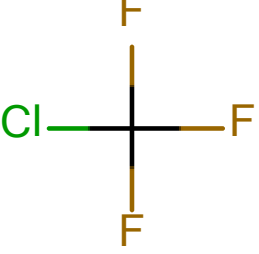
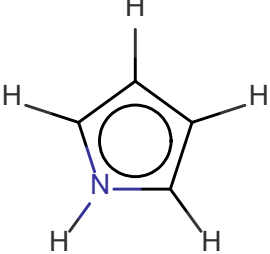
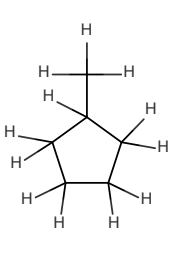
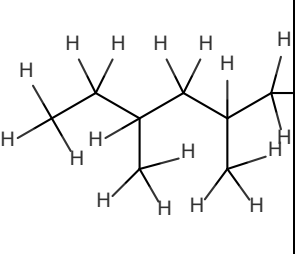
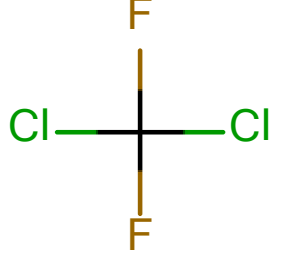
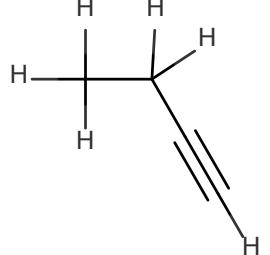
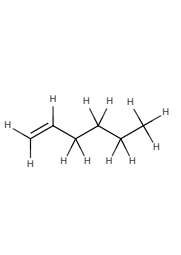
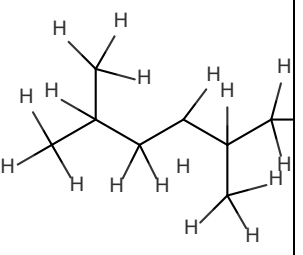
125. Sage B. H. Phase Equilibria in Hydrocarbon Systems II. Methane-Propane System / B. H. Sage, W. N. Lacey, J. G. Schaafsma // *Industrial & Engineering Chemistry*. – 1934. – Т. 26. – №. 2. – С. 214-217.
126. QSPR modeling of critical properties of organic binary mixtures / [Mokshyna E., Polishchuk C., Nedostup V., Kuz'min V.] // *Russian Journal of Organic Chemistry*. – 2016. – Т. 52, № 1. – С. 5-10.
127. Pitzer K. S. The volumetric and thermodynamic properties of fluids. I. Theoretical basis and virial coefficients/ Pitzer K. S. // *Journal of the American Chemical Society*. – 1955. – Т. 77, № 13. – С. 3427-3433.
128. Недоступ В.И. Структурно-молекулярный критерий термодинамического подобия / В.И. Недоступ, В.Е. Кузьмин, О.В. Недоступ // *Журнал физической химии*. – 1999. – Т. 73, № 9. – С. 1541-1544.
129. Mokshyna E. Influence of Molecular Shape and Size on Acentric Factor: Causal Analysis / Mokshyna E. // XVIII Scientific Youth Conference “Problems and Achievements of Modern Chemistry”: материалы конференции. – Одесса, 2016.
130. An easy approach to molecular descriptor calculations / [Todeschini R., Consonni V., Mauri A., Pavan M.] // *Match*. – 2006. – Т. 56. – №. 2. – С. 237-248.
131. Kaye & Laby Online Tables of Physical & Chemical Constants Version 1.1. [Електронний ресурс] : [Інтернет-портал]. – Електронні дані. – [Одеса: Український Інтернет холдинг ТОВ "Тенет"] – Режим доступу: www.kayelaby.npl.co.uk (дата звернення 30.03.2016). – Назва з екрана.
132. Predictive QSPR Modelling for the Second Virial Coefficient of the Pure Organic Compounds / [Mokshyna E., Polishchuk C., Nedostup V., Kuzmin V.] // *Molecular Informatics*. – 2015. – Т. 34, № 1. – С. 53-59.
133. Pitzer K. S. The volumetric and thermodynamic properties of fluids. III. Empirical equation for the second virial coefficient / K. S. Pitzer, Jr R. Curl // *Journal of the American Chemical Society*. – 1957. – Т. 79, № 10. – С. 2369-2370.
134. QSPR-Modeling for the Second Virial Cross-Coefficients of Binary Organic Mixtures / [Mokshyna E., Polishchuk C., Nedostup V., Kuz'min V.] // *International*

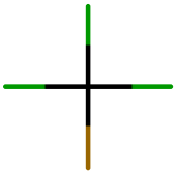
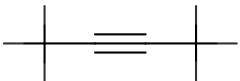
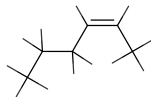
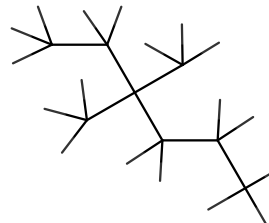
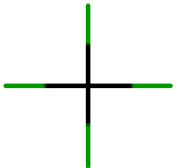
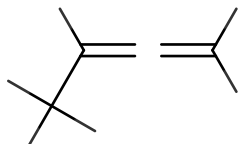
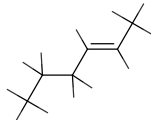
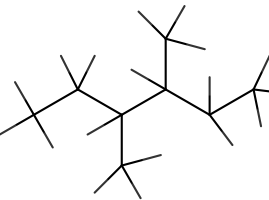
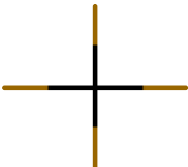
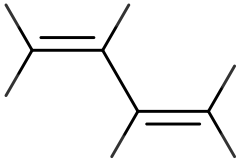
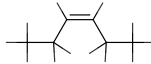
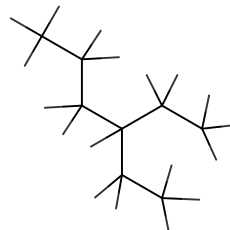
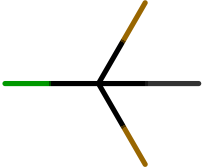
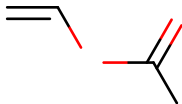
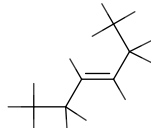
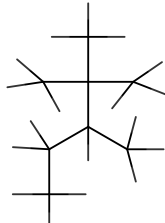
Journal of Quantitative Structure-Property Relationships (IJQSPR). – 2016. – T. 1, № 2. – C. 72-84.

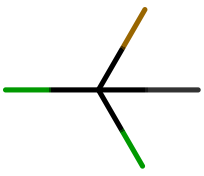
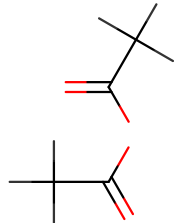
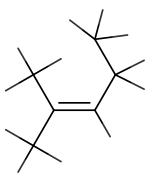
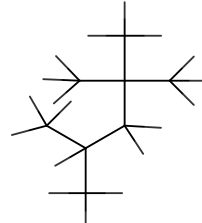
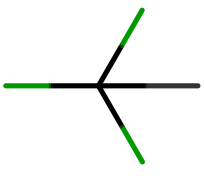
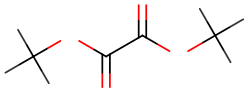
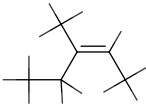
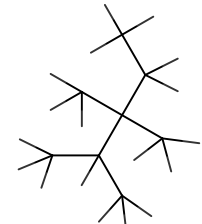
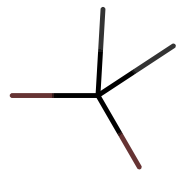
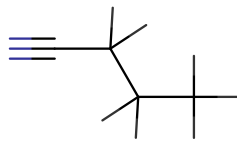
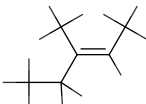
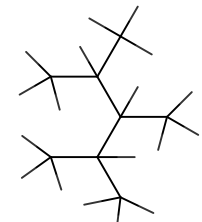
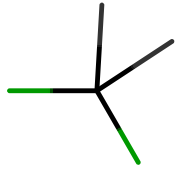
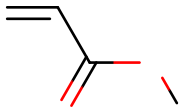
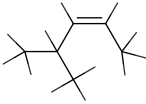
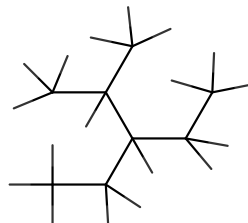
135. Tsuzuki S. Intermolecular interaction between hexafluorobenzene and benzene: Ab initio calculations including CCSD (T) level electron correlation correction / S. Tsuzuki, T. Uchimaru, M. Mikami // The Journal of Physical Chemistry A. – 2006. – T. 110, № 5. – C. 2027-2033.

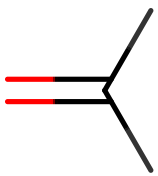
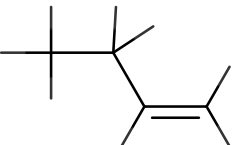
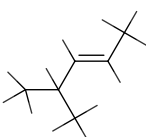
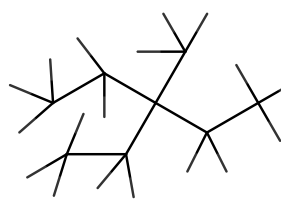
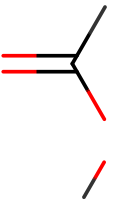
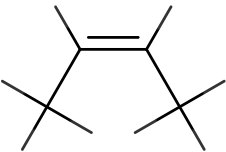
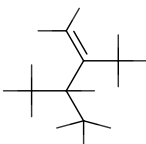

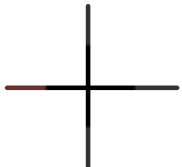
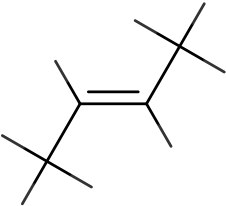
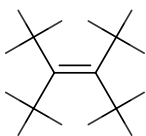

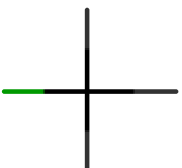
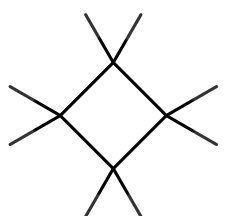
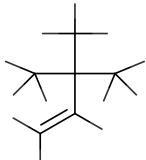
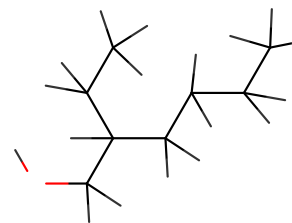
ПРИЛОЖЕНИЕ 1

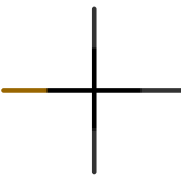
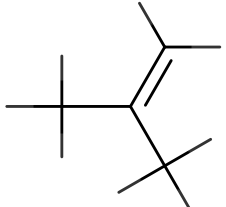
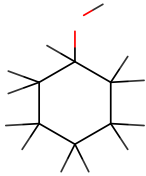
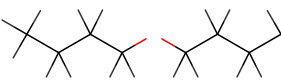
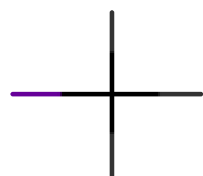
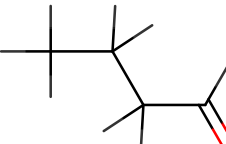
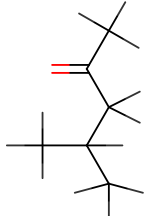
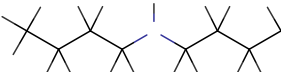
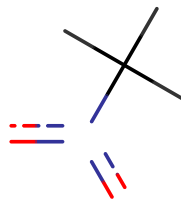
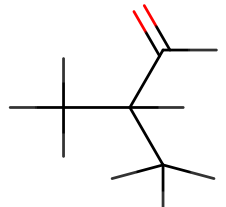
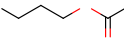
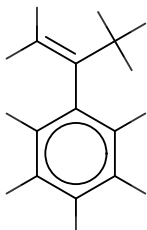
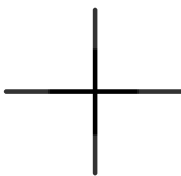
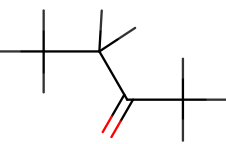
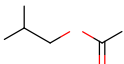
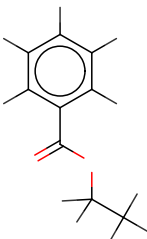
1.1. Молекулы обучающей выборки для критических свойств индивидуальных соединений

Структура	Имя	Структура	Имя	Структура	Имя	Структура	Имя
	бромтрифторметан		бут-3-еннитрил		циклогексан		2,3-диметилгексан
	хлортрифторэтан		1Н-пиррол		метилциклопентан		2,4-диметилгексан
	дихлордифторметан		бут-1-ин		гекс-1-ен		2,5-диметилгексан

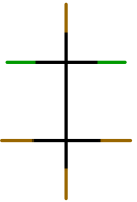
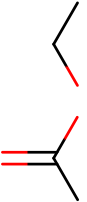
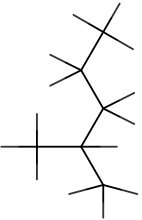
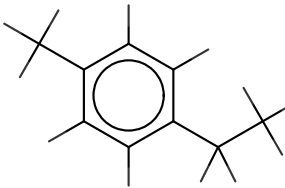
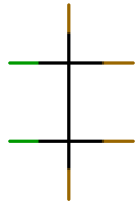
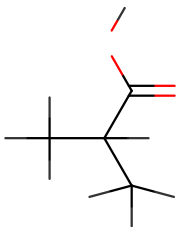
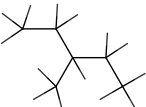
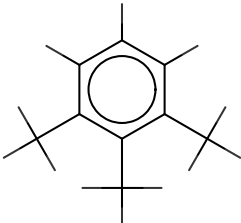
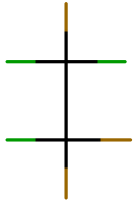
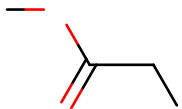
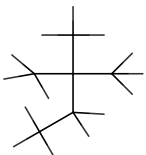
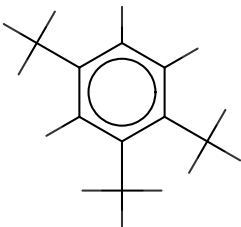
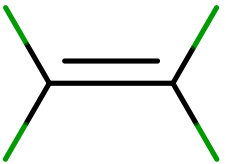

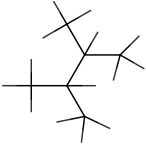
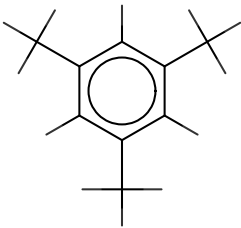
	трихлор (фтор) метан		бут-2-ин		(2Z) гекс-2-ен		3,3- диметилгекса н
	тетрахлормета н		буга-1,2- диен		(2E) гекс-2-ен		3,4- диметилгекса н
	тетрафтормет ан		буга-1,3- диен		(3Z) гекс-3-ен		3-этилгексан
	хлордифторме тан		бут-3-еноат		(3E) гекс-3-ен		2,2,3- триметилпент ан

	дихлор (фтор) метан		ацетилацетат		2-метилпент- 2-ен		2,2,4- триметилпент ан
	трихлорметан		диметилокса лат		(2Z) -3- метилпент-2- ен		2,3,3- триметилпент ан
	дибромметан		бутаннитрил		(2E) -3- метилпент-2- ен		2,3,4- триметилпент ан
	дихлорметан		2- метилпроп- 2-еновая кислота		(2Z) -4- метилпент-2- ен		3-этил-2- метилпентан

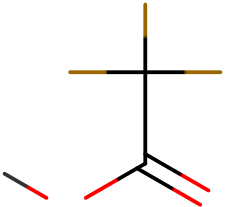
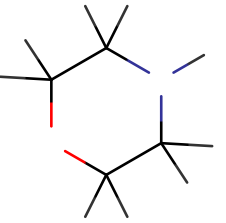
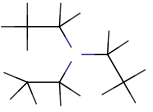
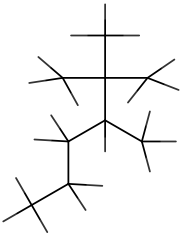
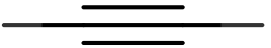
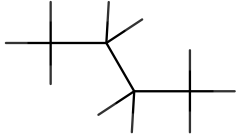
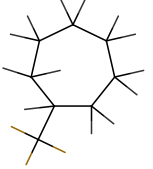
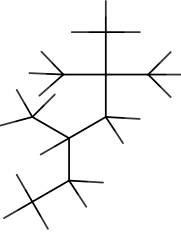
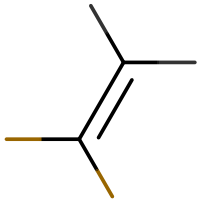
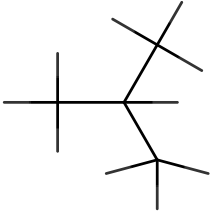
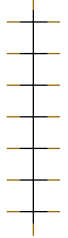
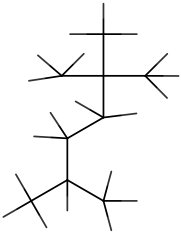
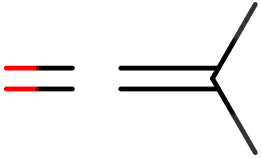
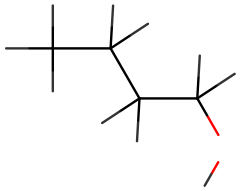
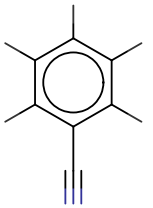
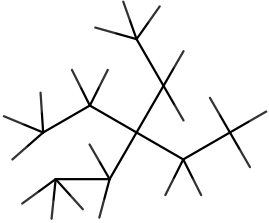
	формальдегид		бут-1-ен		(2E) -4-метилпент-2-ен		3-этил-3-метилпентан
	муравьиная кислота		(2Z) бут-2-ен		2,3-диметилбут-1-ен		октан-1-ол
	бромметан		(2E) бут-2-ен		2,3-диметилбут-2-ен		октан-2-ол
	хлорметан		циклобутан		3,3-диметилбут-1-ен		2-этилгексан-1-ол

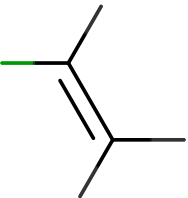
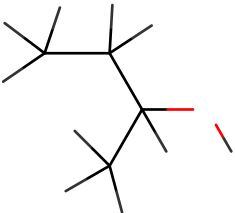
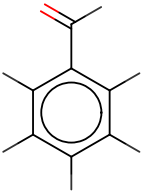
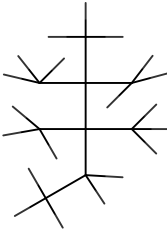
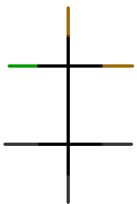
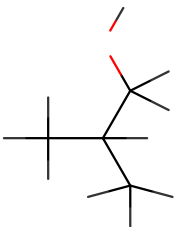
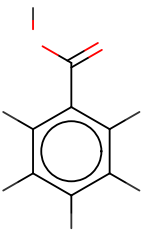
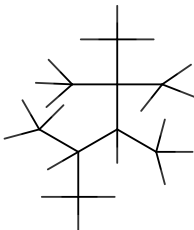
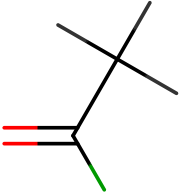
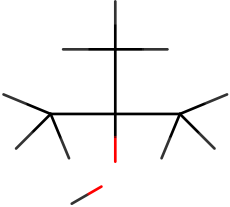
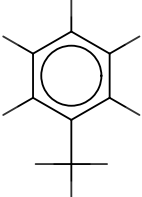
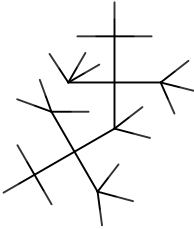
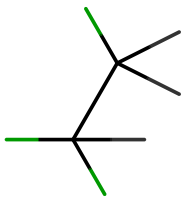
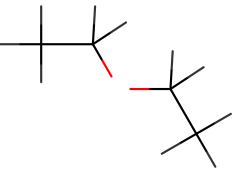
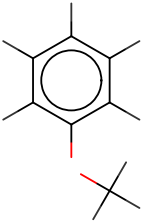
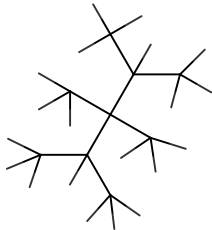
	фторметан		2- метилпроп- 1-ен		циклогексано л		1- бутоксибутан
	йодметан		бутанол		4- метилпентан- 2-он		дибутиламин
	нитрометан		2- метилпропан аль		гексаноат		проп-1-ен-2- илбензол
	метан		бутан-2-он		4- метилпентано ат		этилбензоат

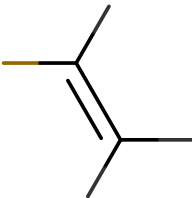

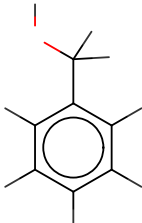
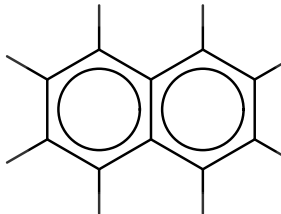
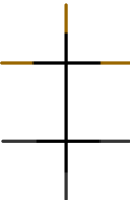
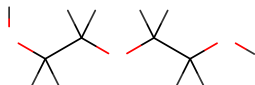
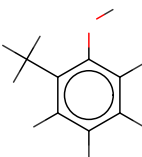
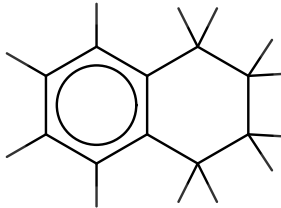
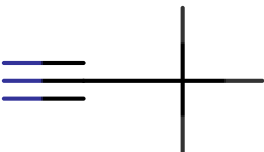
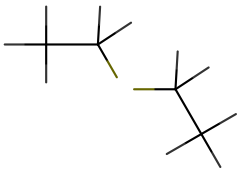
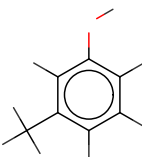
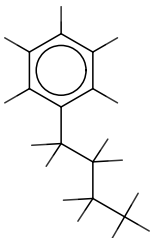
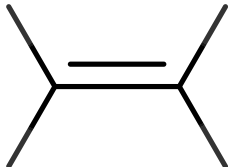
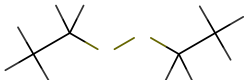
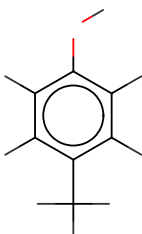
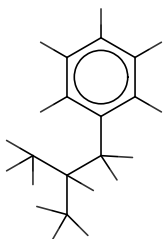
	метанол		оксолан		этил буганоат		пропилбензо л
	метантиол		4- (буг-3-ен- 1-илокси) буг-1-ен		этил-2- метилпропан оат		пропан-2- илбензол
	метанамин		бутановая кислота		пропил пропаноат		1-этил-2- метилбензол
	1-хлор- 1,1,2,2,2- пентафторэта н		1,4-диоксан		гексан		1-этил-3- метилбензол

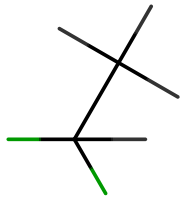
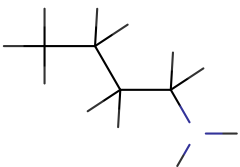
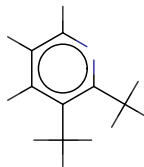
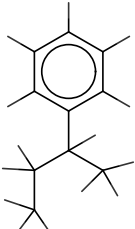
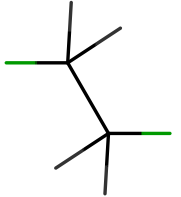
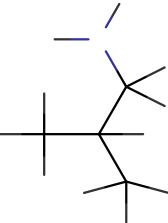
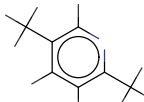
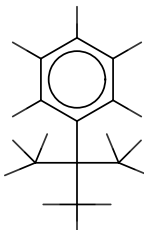
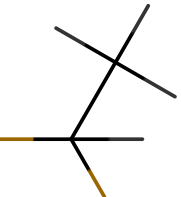
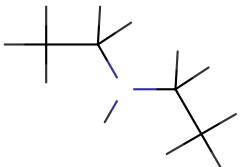
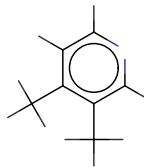
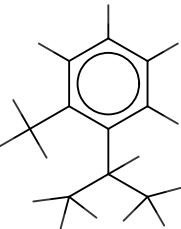
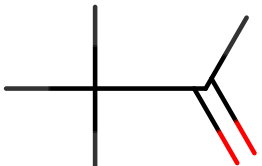
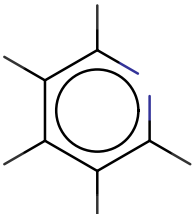
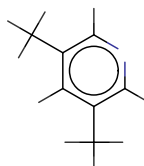
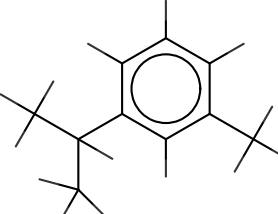
	1,1-дихлор- 1,2,2,2- тетрафторэтан		бутаноат		2- метилпентан		1-этил-4- метилбензол
	1,2-дихлор- 1,1,2,2- тетрафторэтан		2- метилпропан овая кислота		3- метилпентан		1,2,3- триметилбенз ол
	1,1,2-трифтор- 1,2,2- трихлорэтан		2- метилпропан оат		2,2- диметилбутан		1,2,4- триметилбенз ол
	тетрахлорэтен		бутаноат		2,3- диметилбутан		1,3,5- триметилбенз ол

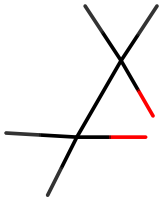
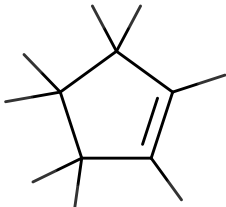
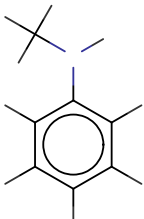
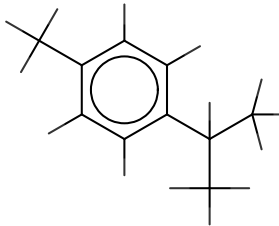
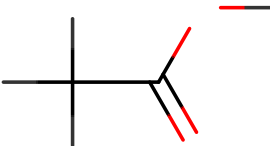
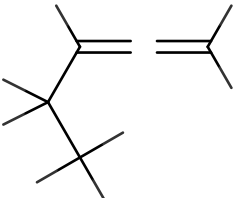
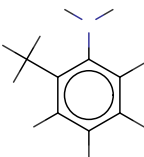
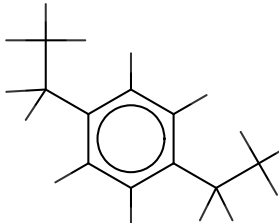
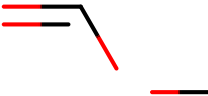
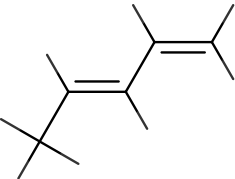
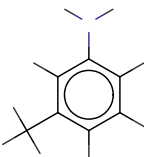
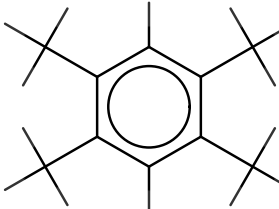
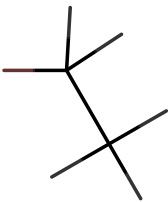
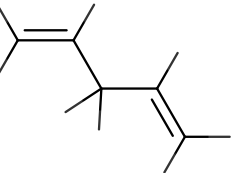
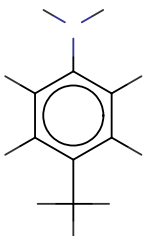
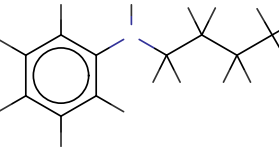
	1,1,2,2- тетрахлор-1,2- дифторэтан		1-хлорбутан		гексан-1-ол		пропилцикло гексан
	тетрафторэтен		2-хлорбутан		1- (гексилокси) гексан		проп-2- илциклогекса н
	гексафторэтан		2-хлор-2- метилпропан		2- (пропан-2- илокси) пропан		нон-1-ен
	1,1,2- трихлорэтен		пирролидин		дипропилами н		нонан

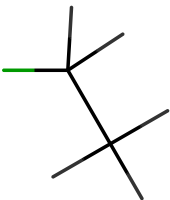
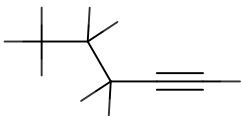
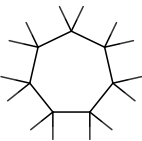

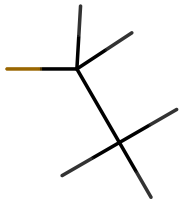
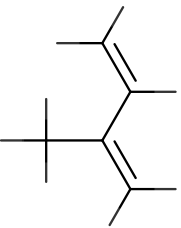
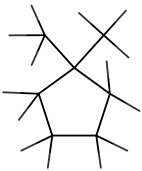
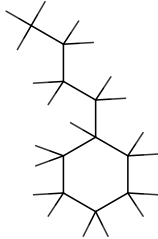
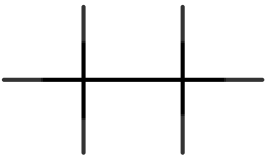
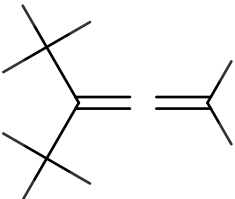
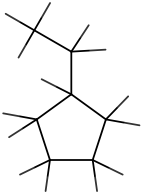
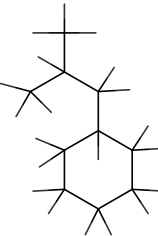
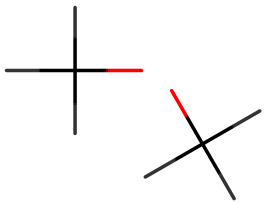
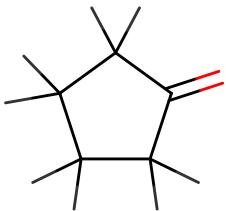
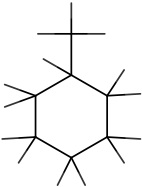
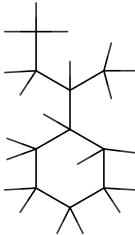
	трифторуксусная кислота		морфолин		триэтиламин		2,2,3-триметилгексан
	ацетилен		бутан		(Трифторметил) циклогептан		2,2,4-триметилгексан
	1,1-дифторэтен		2-метилпропан		гексадекафторептан		2,2,5-триметилгексан
	кетен		бутан-1-ол		бензонитрил		3,3-диэтилпентан

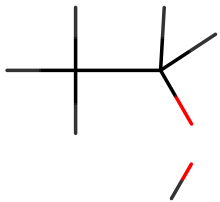
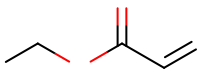

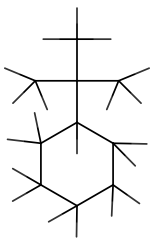
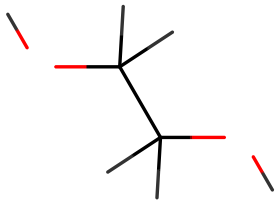
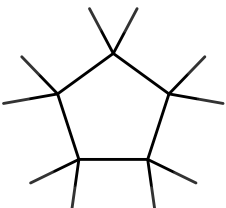
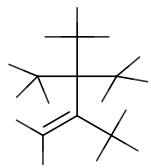

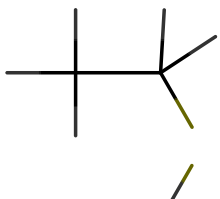
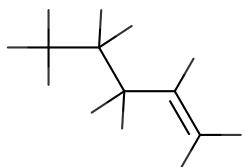


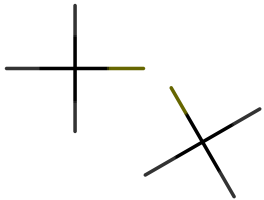
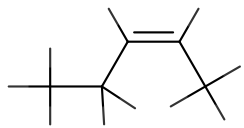
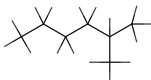
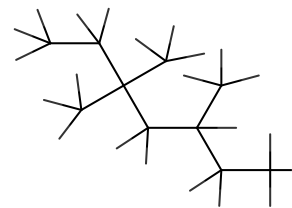
	хлорэтилен		бутан-2-ол		бензальдегид		2,2,3,3- тетраметилпе нтан
	1-хлор-1,1- дифторэтан		2- метилпропан -1-ол		бензойная кислота		2,2,3,4- тетраметилпе нтан
	ацетилхлорид		2- метилпропан -2-ол		толуол		2,2,4,4- тетраметилпе нтан
	1,1,2- трихлорэтан		этоксиэтан		анизол		2,3,3,4- тетраметилпе нтан

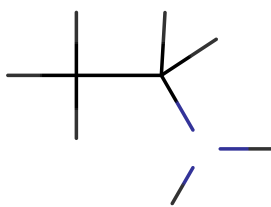
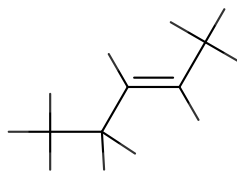
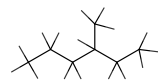
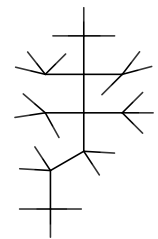
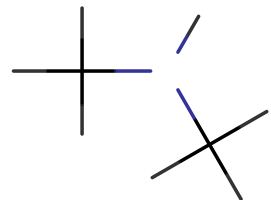
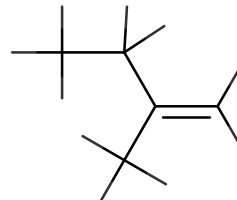
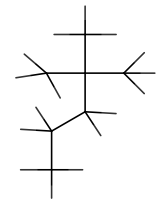
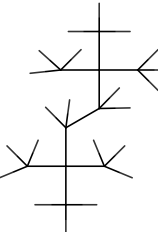
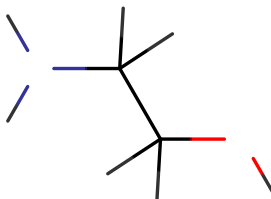
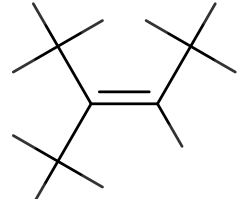
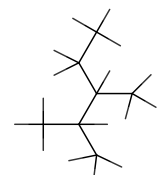

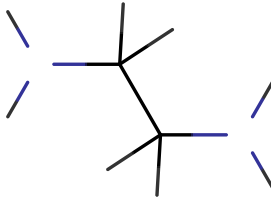
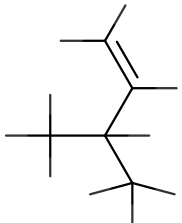
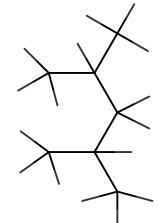
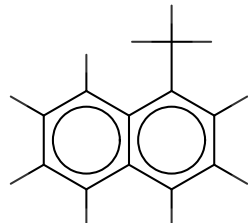
	фторэтен		1,2- диметоксиэтан		фенилметанол		нафталин
	1,1,1- трифторэтан		2- (2- гидроксиэтоксид) этан-1-ол		2-метилфенол		1,2,3,4- тетрагидронафталин
	ацетонитрил		(Этилсульфонил) этан		3-метилфенол		бутилбензол
	этен		(Этилдисульфонил) этан		4-метилфенол		(2- метилпропил) бензол

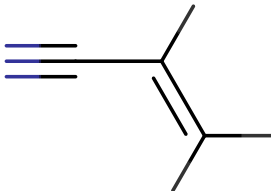


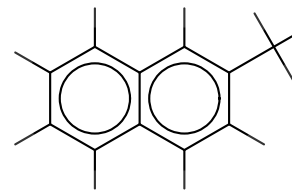
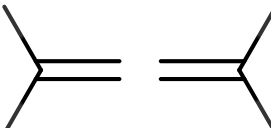
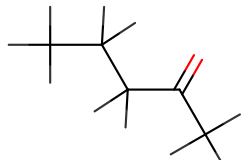
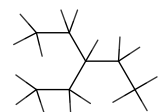
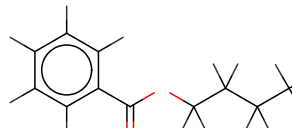
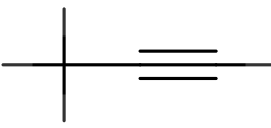
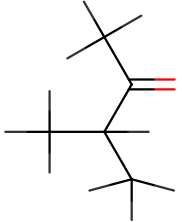
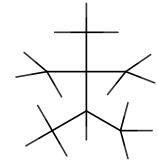
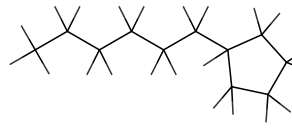
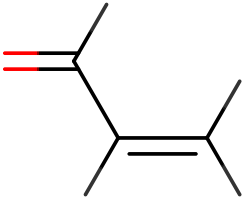
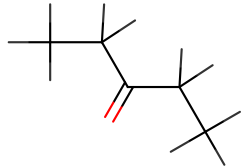


	1,1- дихлорэтан		бутан-1- амин		2,3- диметилпири дин		бутан-2- илбензол
	1,2- дихлорэтан		2- метилпропан -1-амин		2,5- диметилпири дин		трет- бутилбензол
	1,1- дифторэтан		диэтиламин		3,4- диметилпири дин		1-метил-2- (пропан-2-ил) бензол
	ацетальдегид		пиридин		3,5- диметилпири дин		1-метил-3- (пропан-2-ил) бензол

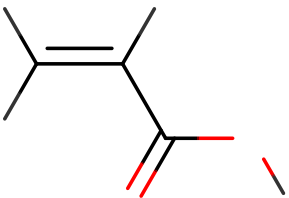

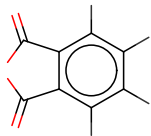

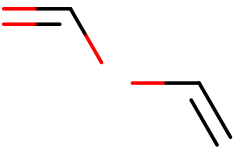
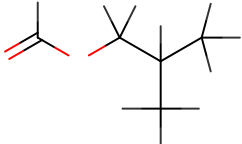
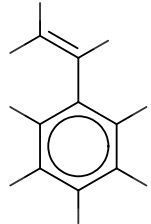
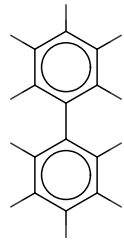
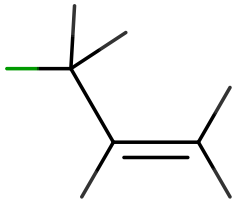
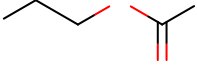
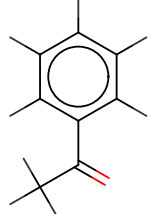
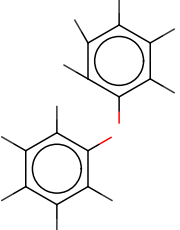
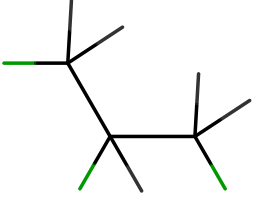
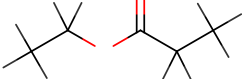
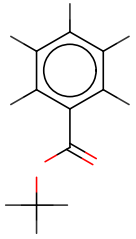
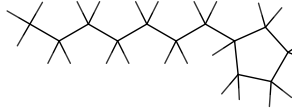
	оксиран		циклопентен		N-метиланилин		1-метил-4-(пропан-2-ил)бензол
	уксусная кислота		пента-1,2-диен		2-метиланилин		1,4-диэтилбензол
	ацетат		(3E) пента-1,3-диен		3-метиланилин		1,2,4,5-тетраметилбензол
	бромэтан		пента-1,4-диен		4-метиланилин		N-бутиланилин

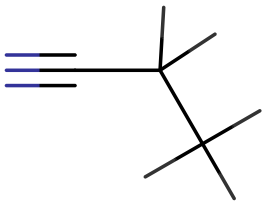
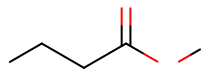
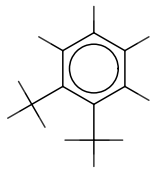

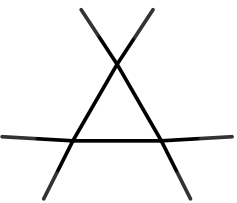
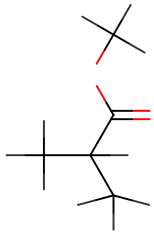
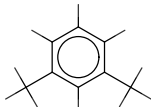

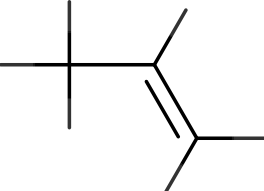
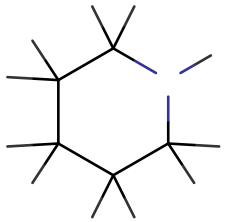
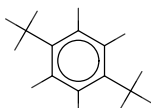
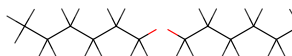
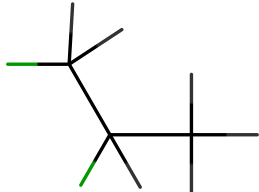
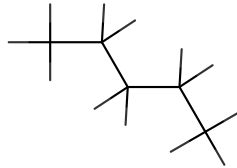
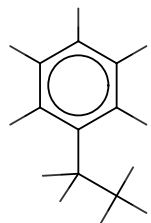

	хлорэтан		пент-1-ин		циклогептан		октаннитрил
	фторэтан		2-метилбута- 1,3-диен		1,1- диметилцикл опентан		бутилциклоге ксан
	этан		3-метилбута- 1,2-диен		этилциклопен тан		(2- метилпропил) циклогексан
	метоксиметан		циклопентан он		метилциклоге ксан		бут-2- илциклогекса н

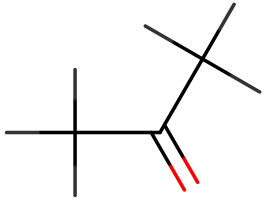
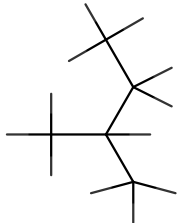
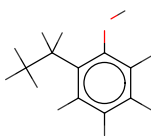
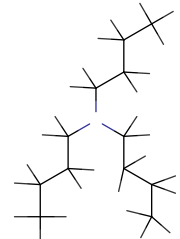
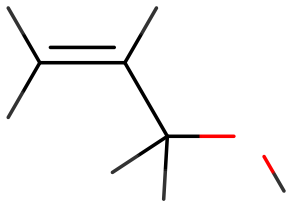
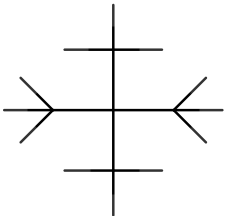
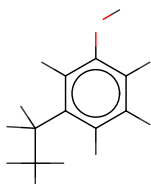
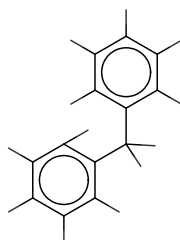
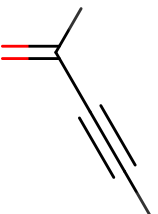
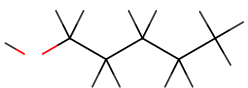
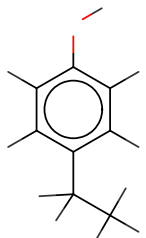
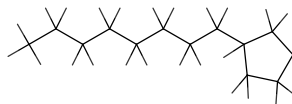
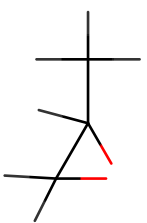
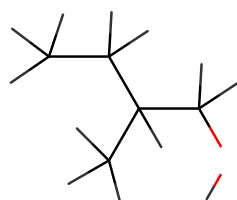
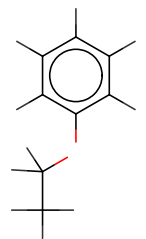

	спирт этиловый		2- метиленбу таноат		гепт-1-ен		трет- бутилциклоге ксан
	этан-1,2-диол		циклопентан		2,3,3- триметилбут- 1-ен		дек-1-ен
	этантиол		пент-1-ен		гептан		декан
	(Метилсульфа нил) метан		(2Z) пент-2- ен		2- метилгексан		3,3,5- триметилгепт ан

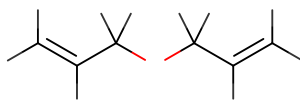
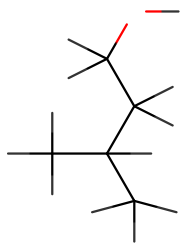
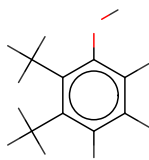
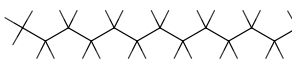
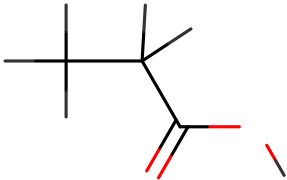
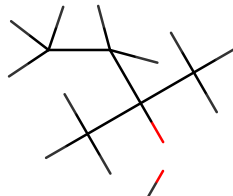
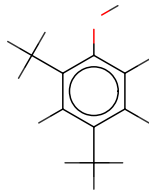
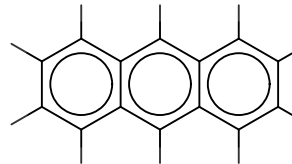
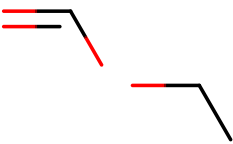
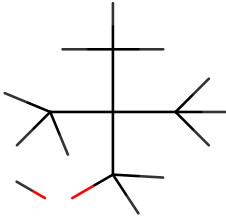
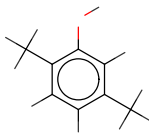
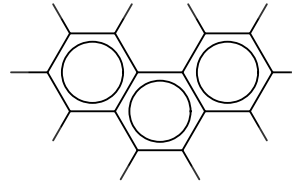
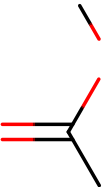
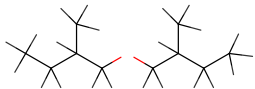
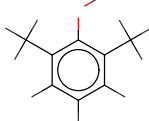
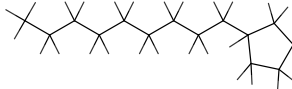
	этанамина		(2E) пент-2-ен		3-метилгексан		2,2,3,3-тетраметилгексан
	диметиламина		2-метилбут-1-ен		2,2-диметилпентан		2,2,5,5-тетраметилгексан
	2-аминоэтан-1-ол		2-метилбут-2-ен		2,3-диметилпентан		декан-1-ол
	этан-1,2-диамина		3-метилбут-1-ен		2,4-диметилпентан		1-метилнафталин

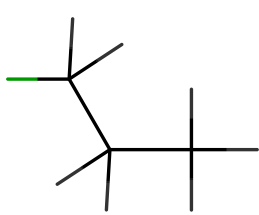
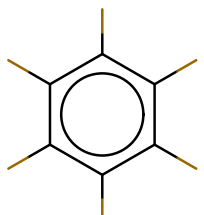
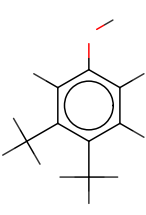

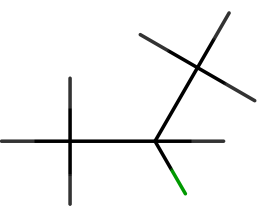
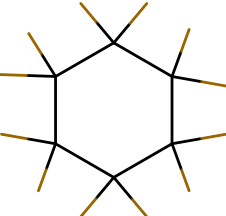
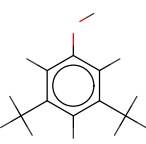
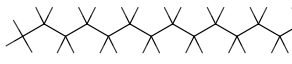
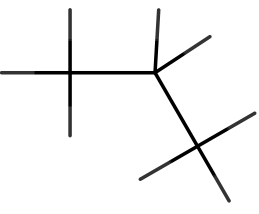

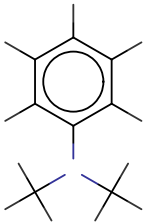
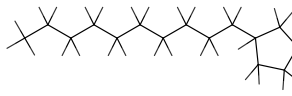
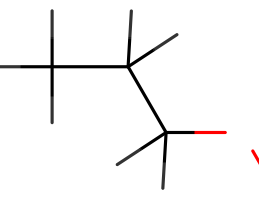
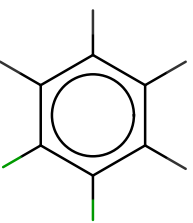
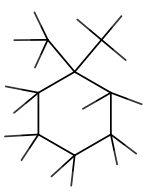
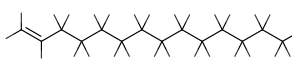
	проп-2- еннитрил		пентаналь		3,3- диметилпента н		2- метилнафта лин
	пропа-1,2- диен		пентан-2-он		3-этилпентан		бутилбензоат
	проп-1-ина		3- метилбутан- 2-он		2,2,3- триметилбута н		гексилциклоп ентан
	проп-2-енал		пентан-3-он		гепт-1-ол		ундец-1-ен

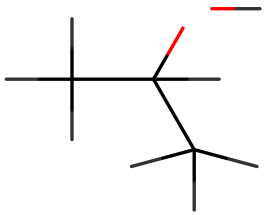
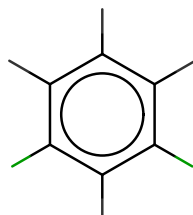
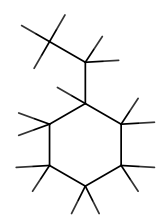

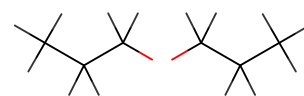
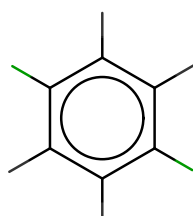
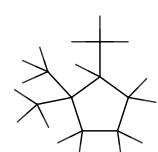
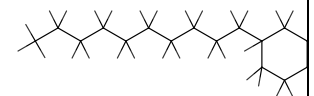
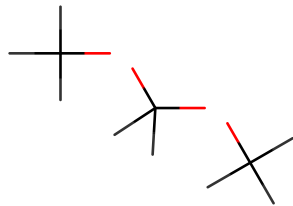
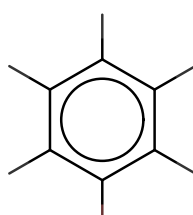
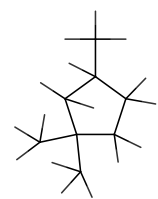

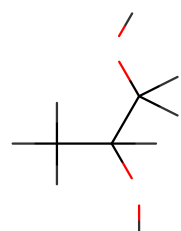
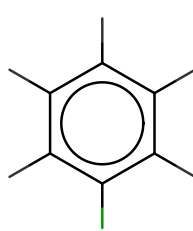
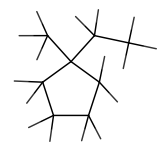

	проп-2-еновой кислоты		пентановая кислота		1,3-дигидро- 2- бензофуран- 1,3-дион		ундекан
	проп-2-еновой кислоты		2- метилпропи л формиат		этилбензол		фенилбензол
	3-хлорпроп-1- ен		пентаноат		1-фенилэтан- 1-он		феноксибензо л
	1,2,3- трихлорпропа н		этил пропионат		метил бензоат		гептилциклоп ентан

	пропаннитрил		пентаноат		1,2-ксилол		додец-1-ен
	циклопропана		метил-2- метилпропан оат		1,3-ксилол		додекан
	проп-1-ен		пиперидин		1,4-ксилол		1- (гексилокси) гексан
	1,2- дихлорпропан		пентан		этилбензол		додекан-1-ол

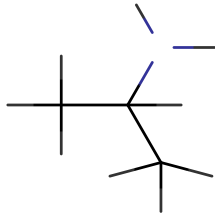
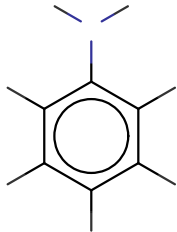

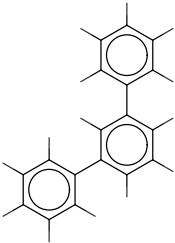
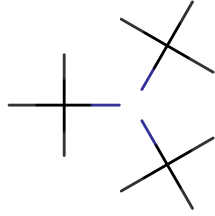
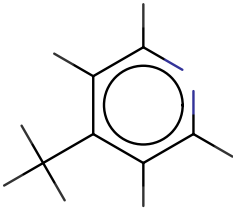
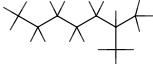
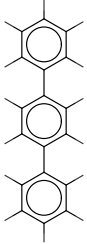
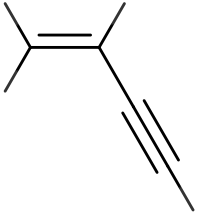

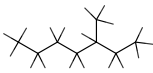

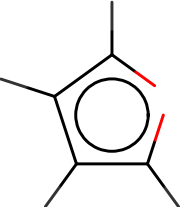
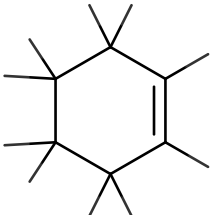
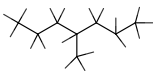
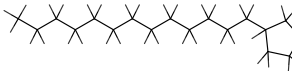
	пропан-2-он		2-метилбутан		2-этилфенол		трибутиламин
	проп-2-ен-1-ол		2,2-диметилпропан		3-этилфенол		бензилбензол
	проп-2-инал		пентан-1-ол		4-этилфенол		октилциклопентан
	2-метилоксиран		2-метилбутан-1-ол		этоксibenзол		тридек-1-ен

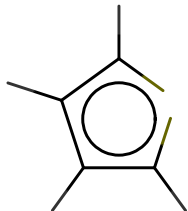
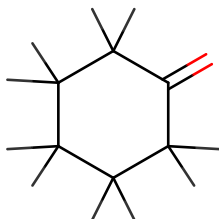
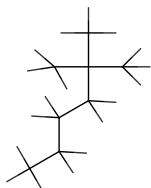

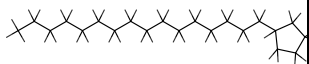



	3- (проп-2-ен-1-илокси) проп-1-ен		3- метилбутан-1-ол		2,3- диметилфенол		тридекан
	пропановая кислота		2- метилбутан-2-ол		2,4- диметилфенол		антрацен
	пропановой		2,2- диметилпропан-1-ол		2,5- диметилфенол		фенантрен
	пропаноат		2-метил-1-(2-метилбутокс и) бутан		2,6- диметилфенол		нонилциклопентан

	1-хлорпропан		гексафторбензол		3,4-диметилфенол		тетрадек-1-ен
	2-хлорпропан		додекафторциклогексан		3,5-диметилфенол		тетрадекан
	пропан		тетрадекафторгексан		N, N-диметиланилин		децилциклопентан
	пропан-1-ол		1,2-дихлорбензол		1,1-диметилциклогексан		пентадек-1-ен

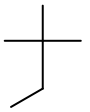
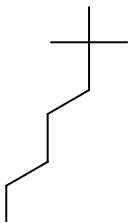
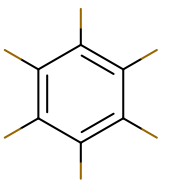
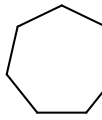

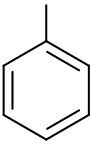
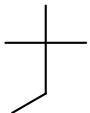
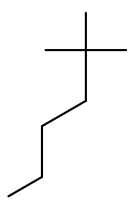
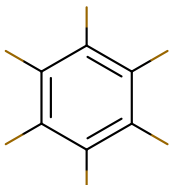
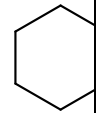
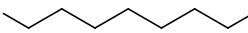

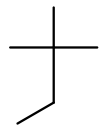
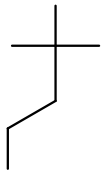
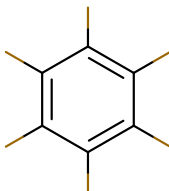

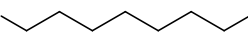

	пропан-2-ол		1,3- дихлорбензо л		этилциклогек сан		пентадекан
	1- пропоксипроп ан		1,4- дихлорбензо л		1,1,2- триметилцик лопентан		децилциклоге ксан
	диметоксимет ан		бромбензол		1,1,3- триметилцик лопентан		гексадек-1-ен
	пропан-1,2- диол		хлорбензол		1-этил-1- метилциклоп ентан		гексадекан

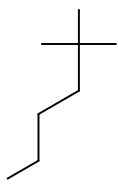
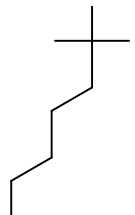
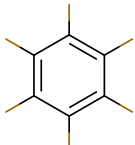
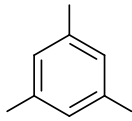
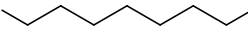

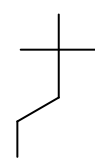
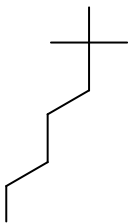
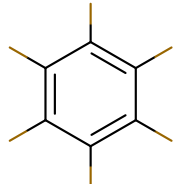
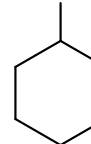
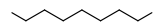
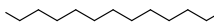
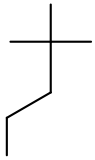
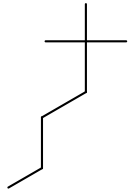


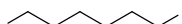
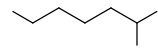
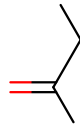
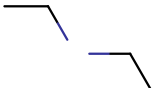

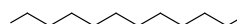
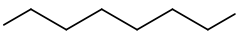

	пропан-1,3-диол		фторбензол		пропилциклопентан		додецилциклопентан
	пропан-1,2,3-триола		йодбензол		пропан-2-илциклопентан		гептадекан-1-ол
	(Метилсульфанил) этан		бензол		окт-1-ен		гептадекан
	пропан-1-амин		фенол		(2E) -окт-2-ен		1,2-дифенилбензол

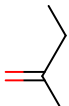

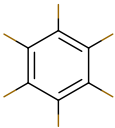

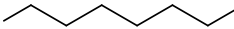

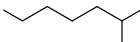
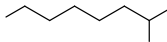
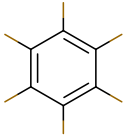

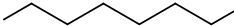

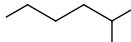
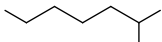
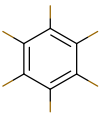



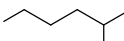
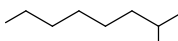
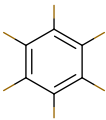



	пропан-2-амин		анилин		октан		1,3-дифенилбензол
	триметиламин		4-метилпиридин		2-метилгептан		1,4-дифенилбензол
	бут-1-ен-3-ин		гекса-1,5-диен		3-метилгептан		октадец-1-ен
	фуран		циклогексен		4-метилгептан		тридецилциклопентан


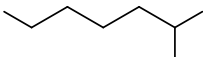
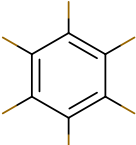
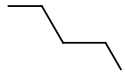



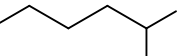
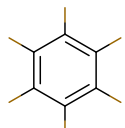




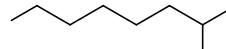
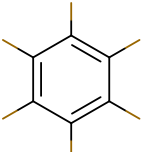
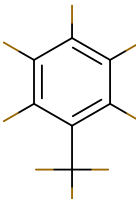

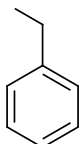

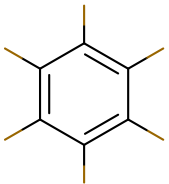
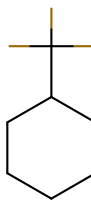
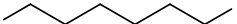

	тиофен		циклогексан он		2,2- диметилгекса н		октадекан
	гексадецилци клопентан		эйкозан		нонадекан		октадекан-1- ол


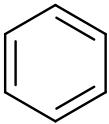
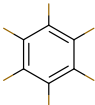
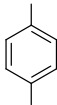
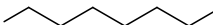



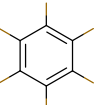

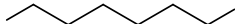




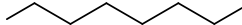


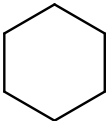

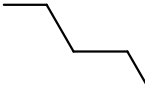

1.2. Смеси обучающей выборки для критических свойств


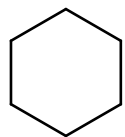
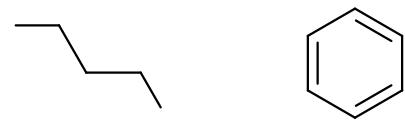



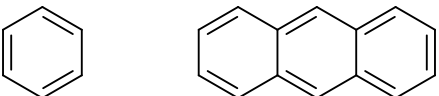

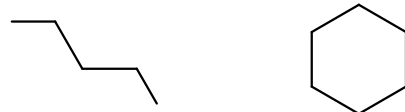

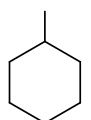

Структура	Имя	Структура	Имя	Структура	Имя
 	2,2- диметилб утан + 2,2- диметилг ептан	 	гексафторбензо л + циклогептан	 	гексан + толуол
 	2,2- диметилб утан + 2,2- диметилг ексан	 	гексафторбензо л + циклогексан	 	нонан + бензол
 	2,2- диметилб утан + 2,2- диметилп ентан	 	гексафторбензо л + циклопентан	 	нонан + циклогексан



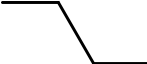
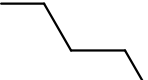

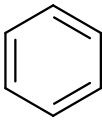

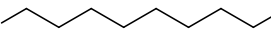
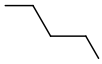
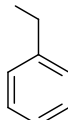
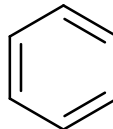




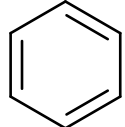
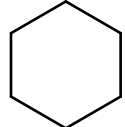



 	2,2- диметилг ексан + 2,2- диметилг ептан	 	гексафторбензо л + 1,3,5- триметилбензол	 	нонан + циклопентан
 	2,2- диметилп ентан + 2,2- диметилг ептан	 	гексафторбензо л + метилциклогекс ан	 	нонан + тридекан
 	2,2- диметилп ентан + 2,2- диметилг ексан	 	гексафторбензо л + декан	 	Октан+ 2- метилгептан
 	бутанола- 2-он + диэтилам ин	 	гексафторбензо л + додекан	 	Октан + уксусная кислота

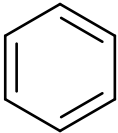
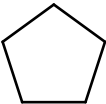

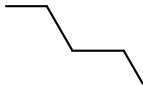

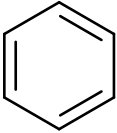


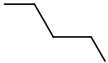


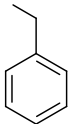
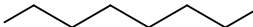
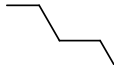

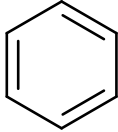

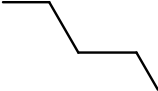

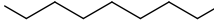
 	бутанола-2-он + гексан	 	гексафторбензо л + гептан	 	Октан + пропан-2-он
 	2- метилгепт ан + 2- метилокт ана	 	гексафторбензо л + гексан	 	Октан + бензол
 	2- метилгекс ан + 2- метилгепт ан	 	гексафторбензо л + нонан	 	Октан + бутанола-1-ол
 	2- метилгекс ан + 2- метилокт ан	 	гексафторбензо л + октан	 	Октан + бутан-2-ол

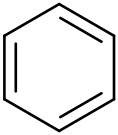



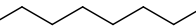

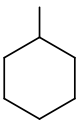

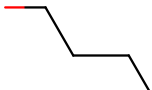

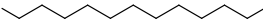
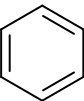
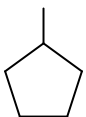



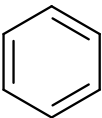
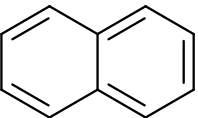
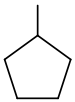
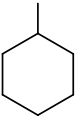
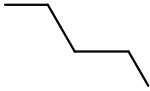

 	2-метилпентан + 2-метилгептан	 	гексафторбензол + пентан	 	Октан + циклогексана
 	2-метилпентан + 2-метилгексан	 	гексафторбензол + тетрадекафторгексан	 	Октан + циклопентан
 	2-метилпентан + 2-метилоктан	 	гексафторбензол + 1,2,3,4,5-пентафторфенил-6-(трифторметил)бензол	 	Октан + этилбензол
	2-метилпентан + вода	 	гексафторбензол + (трифторметил)циклогексан	 	Октан + пентан-2-ол

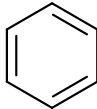



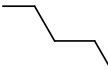
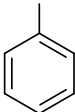

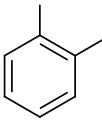


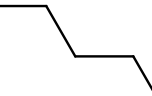






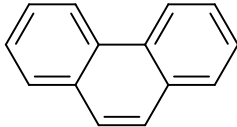
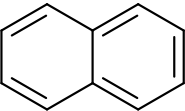
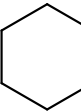


		пропан-2-он + бензол	 	гексафторбензол + 1,4-ксилол	 	Октан + пропан-1-ол
		пропан-2-он + тетрахлорметан	 	гексафторбензол + толуол	 	Октан + пропан-2-ол
		пропан-2-он + трихлорметан		метан + 2-метилбутан	 	Октан + пиридин
		пропан-2-он + циклогексан		метан + 2-метилпропан	 	пентан + пропан-2-он

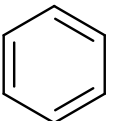
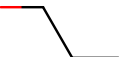
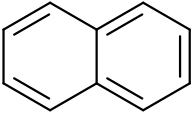



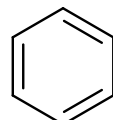

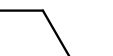




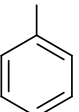

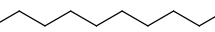

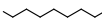
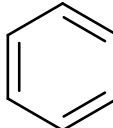

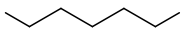

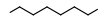
	пропан-2-он + этоксизетан		метан + циклогексан		пентан + бензол
	пропан-2-он + вода		метан + этан		пентан + бутан-1-ол
	бензол + антрацен		метан + этан		пентан + циклогексан
	бензол + бутан-1-ол		метан + метилциклогексан		пентан + циклопентан

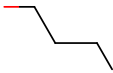





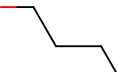
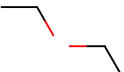

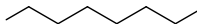


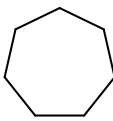
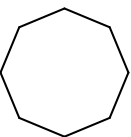
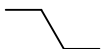
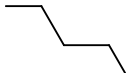
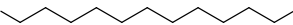

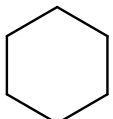
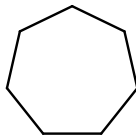
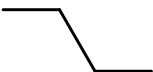
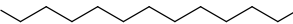

 	бензол + бутан-2- ол		метан + бутан	 	пентан + этанол
 	бензол + тетрахлор метан		метан + декан	 	пентан + этилбензол
 	бензол + трихлорм етан		метан + 2,2- диметилпропан	 	пентан + декан
 	бензол + циклогекс ан		метан + гептан	 	пентан + додекан

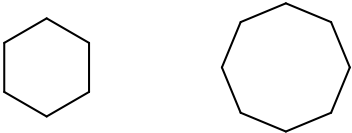
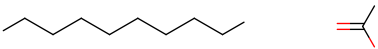
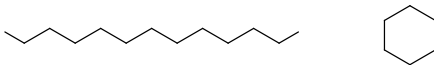

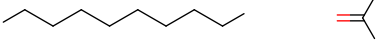
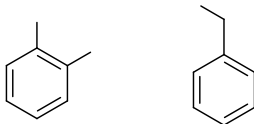
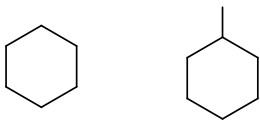
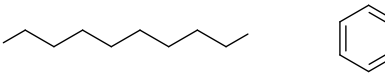


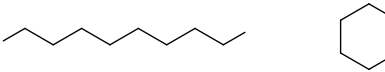
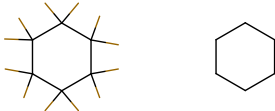
		бензол + циклопеп- тан		метан + гексан			пентан + 2,2- диметилпропа- н
		бензол + этанол		метан + нонан			пентан + гептан
		бензол + этилбензо- л		метан + октан			пентан + гексан
		бензол + этен		метан + пентан			пентан + нонан

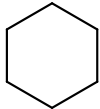

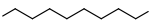
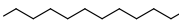
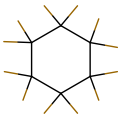

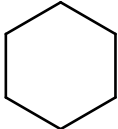


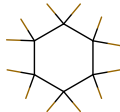


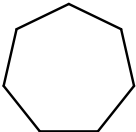




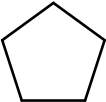
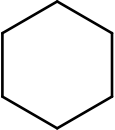


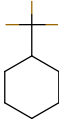

 	бензол + метанол		метан + пропан	 	пентан + октан
 	бензол + метилциклогексан	 	метанол + бутан-1-ол	 	пентан + тридекан
 	бензол + метилциклопентан		метанол + вода	 	пентан + ундекан
 	бензол + нафталин	 	метилциклопентан + метилциклогексан	 	пентан + пропан-2-ол

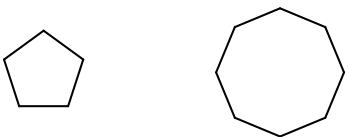
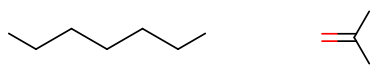
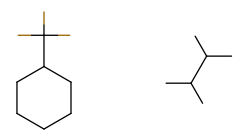
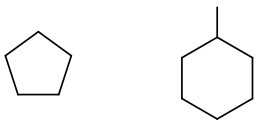
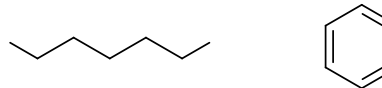
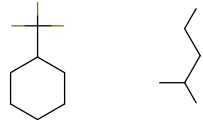

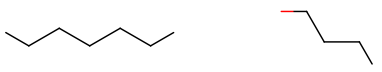
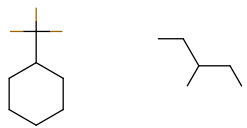

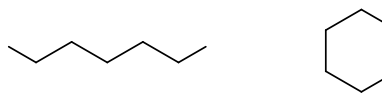
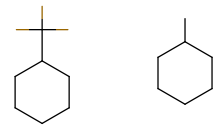
 	бензол + 2,2- диметилпропан	 	хлор (фтор) метан + 1,1- дихлор-1,2,2,2- тетрафторэтан	 	пентан + толуол
 	бензол + 1,2- ксилол	 	2-хлор-1,1,1,2- тетрафторэтан + октафторциклоб утан		пентан + вода
 	бензол + пентан-2- ол	 	хлортрифтормет ан + трифторметансу льфонат		гексадекафто ргептан + этан
 	бензол + фенантр ен	 	нафталин + циклогексан	 	гексадекафто ргептан + бутана


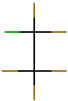


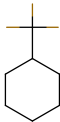

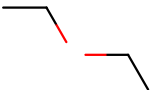
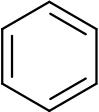

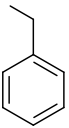

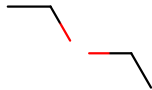





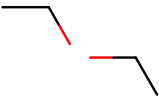

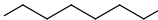


 	бензол + пропан-1- ол	 	нафталин + этен	 	гексадекафто ргептан + гептан
 	бензол + пропан-2- ол	 	бутан + пропан- 2-он	 	гексадекафто ргептан + гексан
 	бензол + толуол	 	бутан + декан	 	гексадекафто ргептан + нонан
	бензол + вода	 	бутан + гептан	 	гексадекафто ргептан + октан



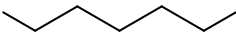









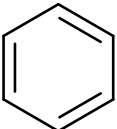



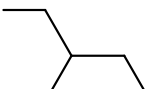

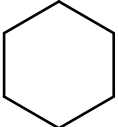




 	бутанола-1-ол + бутан-2-ол	 	бутан + гексан	 	гексадекафторгептан + пентан
 	бутанола-1-ол + этоксигептан	 	бутан + октан	 	гексадекафторгептан + пропан
 	циклогептан + циклооктан	 	бутан + пентан	 	тридекан + пропан-2-он
 	циклогексан + циклогептан		бутан + вода	 	тридекан + бензол




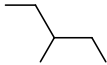
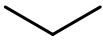


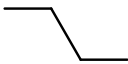



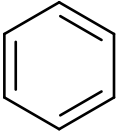





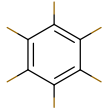

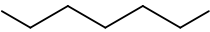



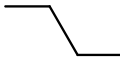
	циклогекс ан + циклоокт ан		декан + уксусная кислота		тридекан + циклогексан
	циклогекс ан + этанол		декан + пропан- 2-он		1,2-ксилол + этилбензол
	циклогекс ан + метилцик логексан		декан + бензол		додекафторци клогексан + циклогептан
	циклогекс ан + метилцик лопентан		декан + циклогексан		додекафторци клогексан + циклогексана





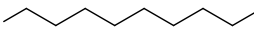

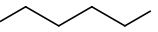
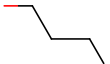










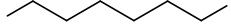
 	циклогексан + 2,2-диметилпропан	 	декан + додекан	 	додекафторциклогексан + циклооктано
	циклогексан + вода	 	декан + пиридин	 	додекафторциклогексан + циклопентан
 	циклопентан + циклогептан	 	додекан + уксусная кислота	 	тетрафторметан + трифторметансульфонат
 	циклопентан + циклогексан	 	додекан + пиридин	 	(Трифторметил) циклогексан + 2,2-диметилбутан

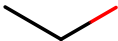



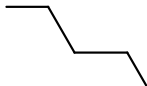
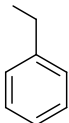


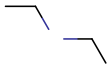



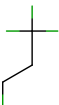

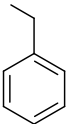



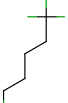

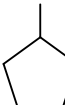

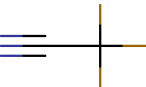
	циклопентан + циклооктан		гептан + пропан-2-он		(Трифторметил) циклогексан + 2,3-диметилбутан
	циклопентан + метилциклогексан		гептан + бензол		(Трифторметил) циклогексан + 2-метилпентан
	циклопентан + метилциклопентан		гептан + бутан-1-ол		(Трифторметил) циклогексан + 3-метилпентан
	дихлордифторметан + 1,1-дифторэтан		гептан + циклогексан		(Трифторметил) циклогексан + метилциклогексан

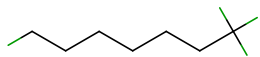




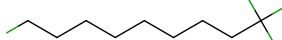


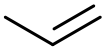




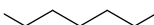

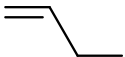



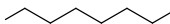
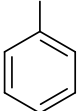

 	дихлорметан + 1-хлор-1,1,2,2,2-пentaфторэтан	 	гептан + циклопентан	 	(Трифторметил) циклогексан + гексан
 	этоксидэтан + бензол	 	гептан + этилбензол		пропан-2-ол + вода
 	этоксидэтан + трихлорметан	 	гептан + этен	 	пропан + 2,2-диметилбутан
	этоксидэтан + вода	 	гептан + октан	 	пропан + 2,3-диметилбутан





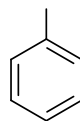





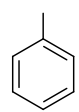
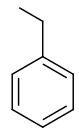



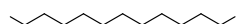
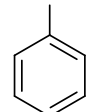

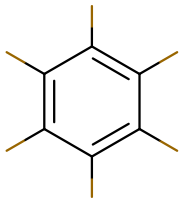
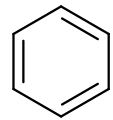
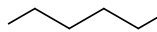

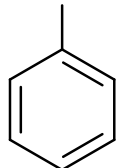
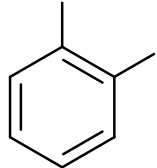
		этан + пропан-2- он		гептан + вода			пропан + 2- метилбутан
		этан + ацетилен		 гексадекан + бензол			пропан + 2- метилпентан
		этан + бензол		 гексан + 2,2- диметилбутан			пропан + 3- метилпентан
		этан + циклогекс ан		 гексан + 2,3- диметилбутан			пропан + пропан-2-он

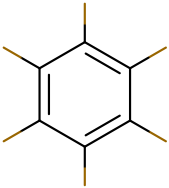



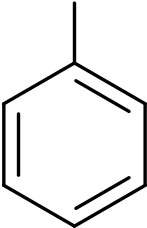
 	этан + этен	 	гексан + 3- метилпентан	 	пропан + ацетилен
 	этан + бутан	 	гексан + уксусная кислота	 	пропан + бензол
 	этан + декан	 	гексан + пропан-2-он	 	пропан + гексафторбен- зол
 	этан + гептан	 	гексан + ацетилен	 	пропан + бутан

— 	этан + гексан	 	гексан + бензол	 	пропан + декан
— 	этан + пропан	 	гексан + бутан- 1-ол	 	пропан + гептан
— 	этан + проп-1-ен	 	гексан + трихлорметан	 	пропан + гексан
—	этан + вода	 	гексан + циклогексан	 	пропан + октан

	этанол + вода	 	гексан + циклопентан	 	пропан + пентан
 	этилбензо л + этанол	 	гексан + диэтиламин	 	пропан + тетрадекафто ргексан
 	этен + 1,1,1,3- тетрахлор опропан	 	гексан + этилбензол	 	пропан + гексафторопр опан-2-он
 	этен + 1,1,1,5- тетрахлор опентан	 	гексан + метилциклопент ан	 	пропан + трифторацето нитрил

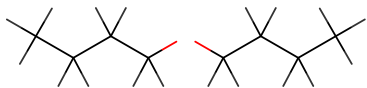
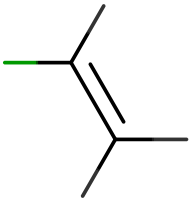
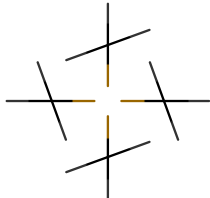
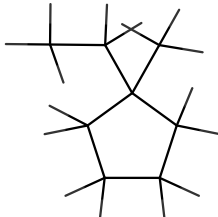
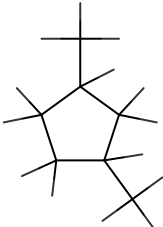
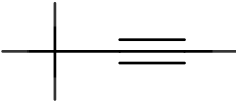
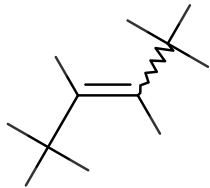
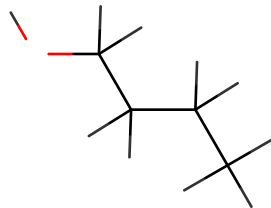
	этен + 1,1,1,7- тетрахлор огеэтан	 	гексан + декан	 	пропан + октафторпроп ан
	этен + 1,1,1,9- тетрахлор ононан	 	гексан + 2,2- диметилпропан	 	проп-1-ен + ацетилен
 	этен + ацетилен	 	гексан + гептан	 	проп-1-ен + бут-1-ен
 	этен + трихлорм етан	 	гексан + октан	 	толуол + циклогексан

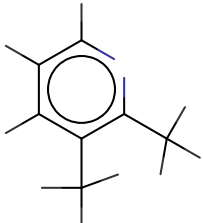
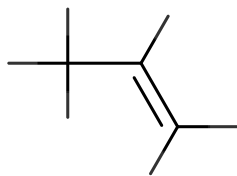
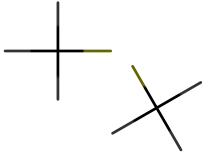
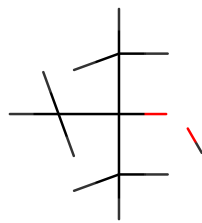
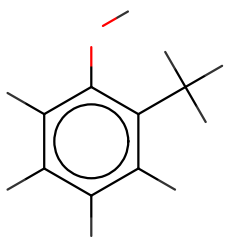
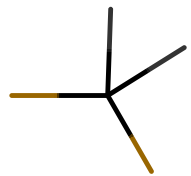


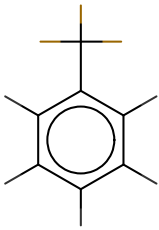
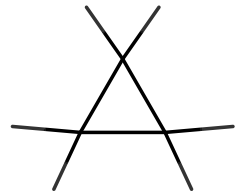
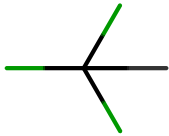
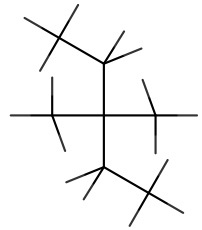

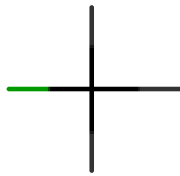
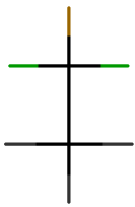
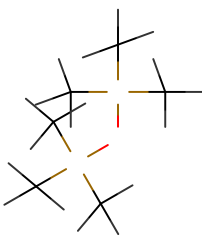
 	этен + этанол	 	гексан + тетрадекафторгексан	 	толуол + этанол
 	этен + хлорэтилен	 	гексан + тетрадекан	 	толуол + этилбензол
 	этен + проп-1-ен	 	гексан + тридекан	 	толуол + 2,2- диметилпропан
 	гексафторбензол + бензол	 	гексан + пропан-1-ол	 	толуол + 1,2- ксилол

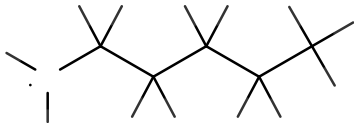
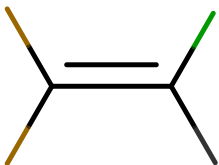
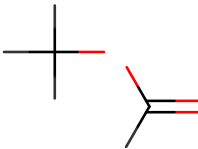
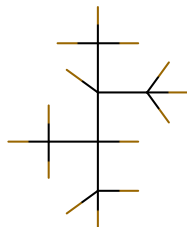
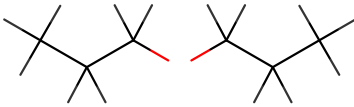
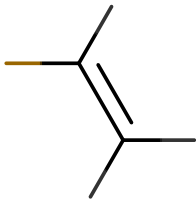
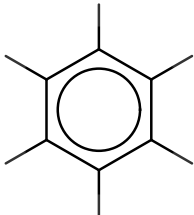
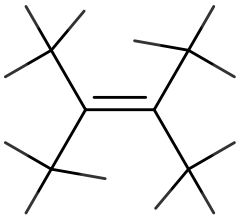
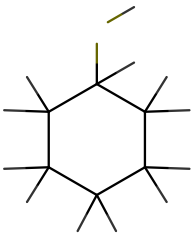
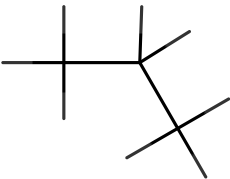
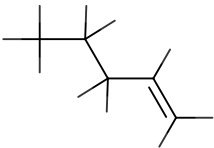
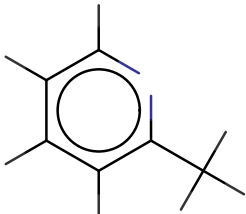
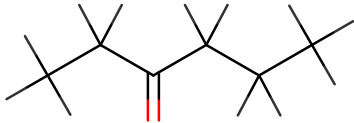

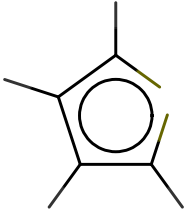
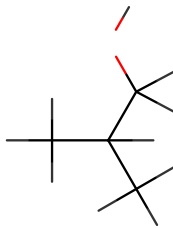
		гексафтор бензол + тетрахлор метан			гексан + пропан-2-ол		толуол + вода
---	---	---	--	---	-------------------------	---	---------------

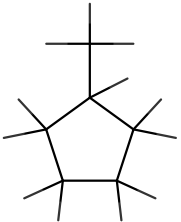
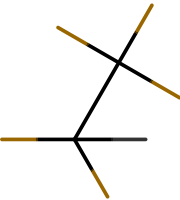
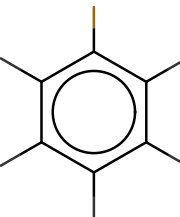
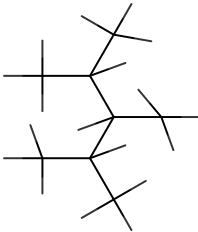
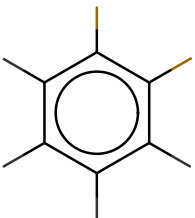
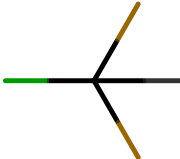
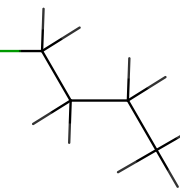
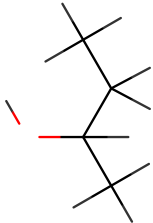
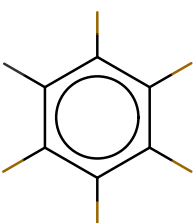
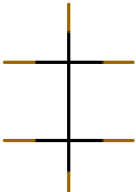
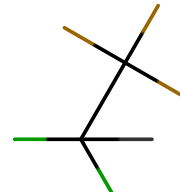

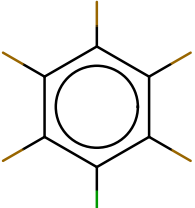
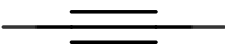
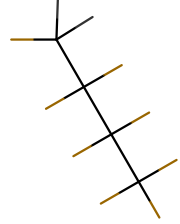
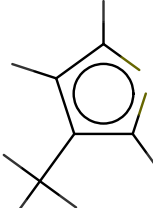
ПРИЛОЖЕНИЕ 2

2.1. Молекулы обучающей выборки для второго вириального коэффициента индивидуальных веществ

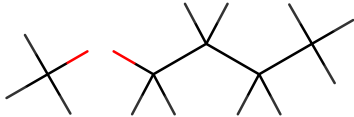
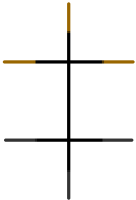
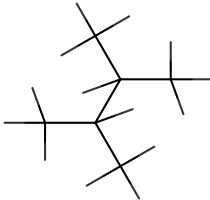
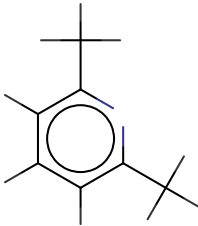
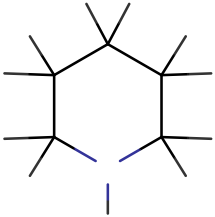
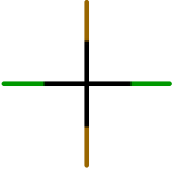
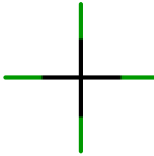
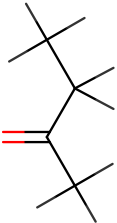
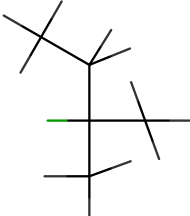
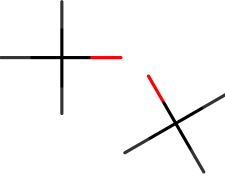
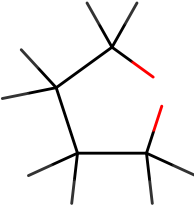
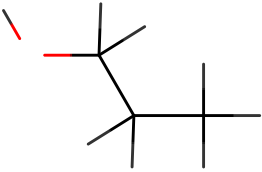
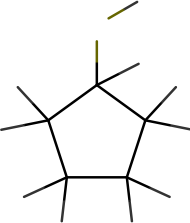
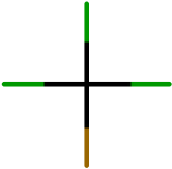
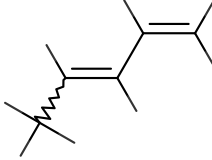
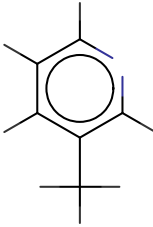
Структура	Имя	Структура	Имя	Структура	Имя	Структура	Имя
	дибутиловый эфир		хлорэтилен		тетраметилсилан		1-этил-1-метилциклопентан
	цис-1,3-диметилциклопентан		1-пропин		(Z) -2-бутен		1-бутанол

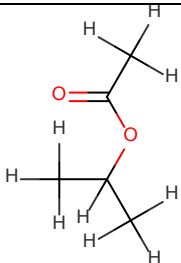
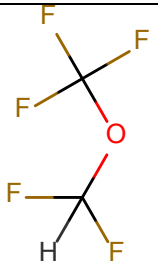
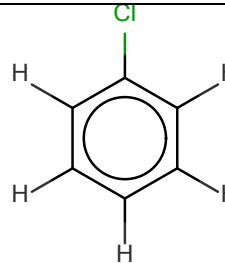
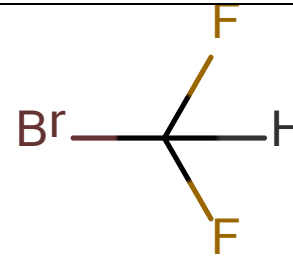
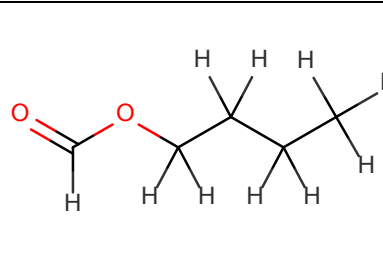
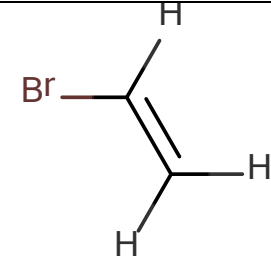
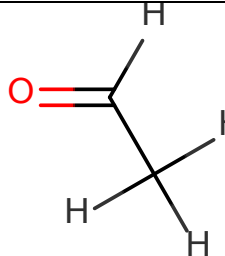
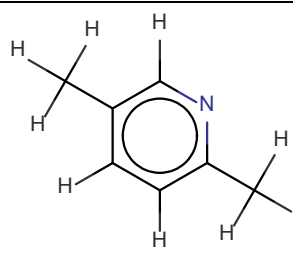
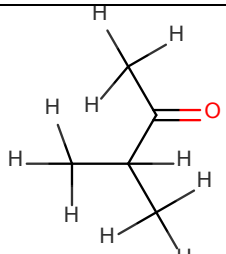
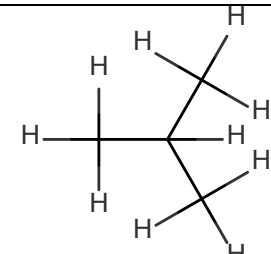
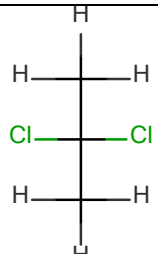
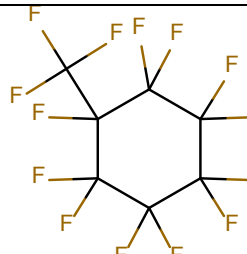
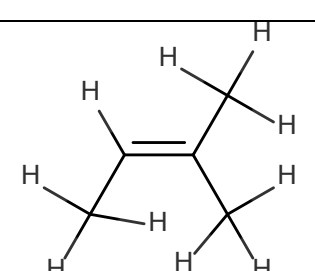
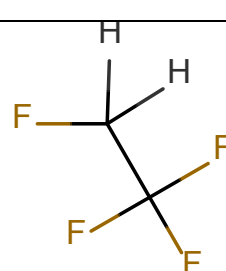
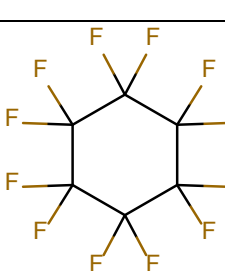
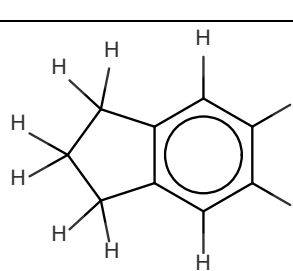
	2,3- диметилпир идин		пропен		2-тиапропан		2- Метил- 2- пропан ол
	2- метилфенол		Дифтормета н		1-гексанол		1- гептен
	(Трифторме тил) бензол		Циклопропа н		Трихлормет ан		3,3- дими лпента н
	Бутиловый эфир		Хлорметан		1,1-дихлор- 1-фторэтан		гекса тилдис илокса н

	Метилпентан		1-хлор-2,2-дифторэтен		Метил метаноат		2,3-бис (трифторметил) - перфторбутан
	Дипропиловый эфир		фторэтен		Бензол		2,3-диметил-2-бутен
	Циклогексантиол		Пропан		1-пентен		2-метилпиридин
	3-гексанон		этандинитрил		тиофен		2-Метил-1-пропанол

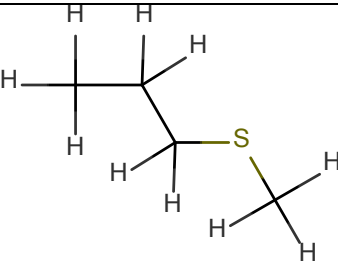
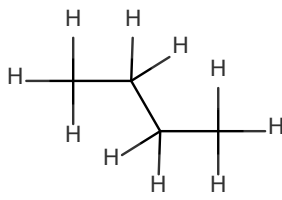
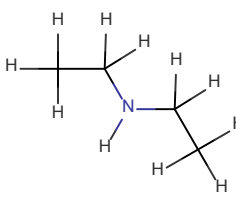
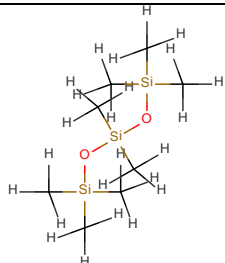
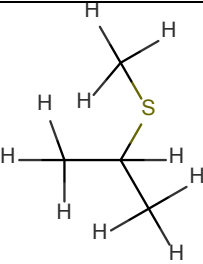
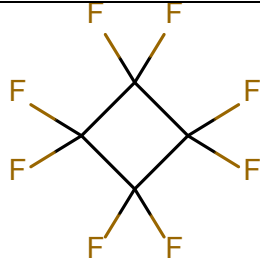
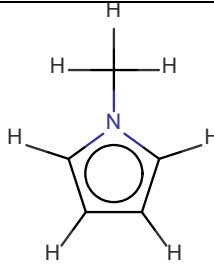
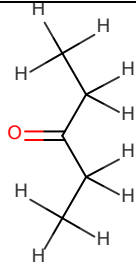
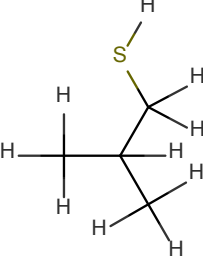
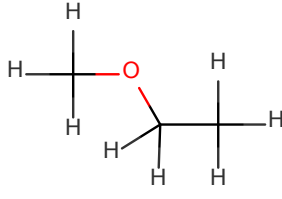
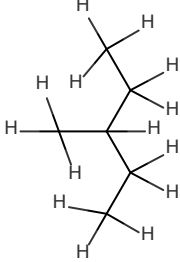
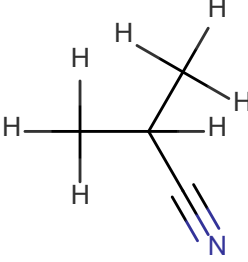
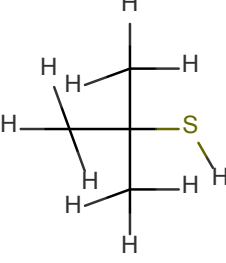
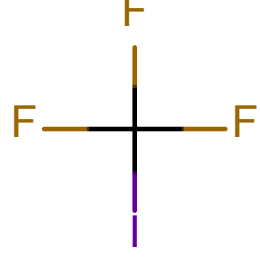
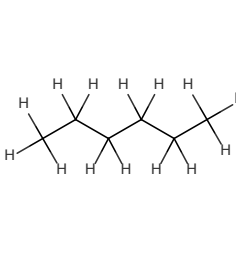
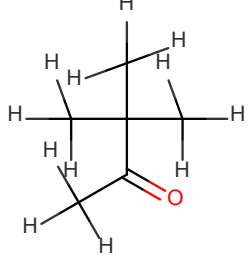
	метилциклопентан		1,1,1,2,2-пентафторэтан		фторбензол		2,3,4-триметилпентан
	1,2-дифторбензол		хлородифторметан		1-хлорбутан		(R)-2-бутанол
	1,2,3,4,5-пентафторбензол		гексафторэтан		1,1,1,2,2,2-гексафторэтан		октан
	хлорпентафторбензол		этин		1,1,1,2,2,3,3,4-октафторбутан		3-метилфенол

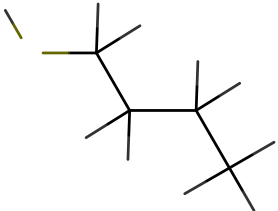
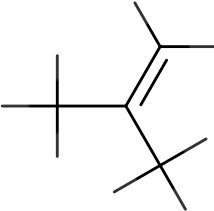
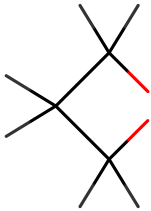
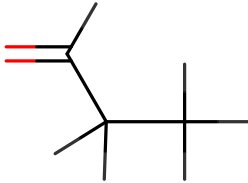
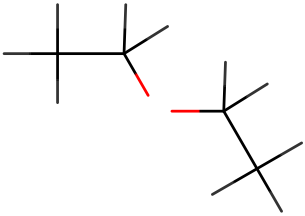
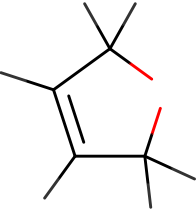
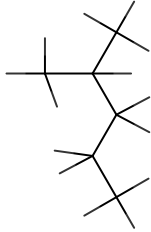

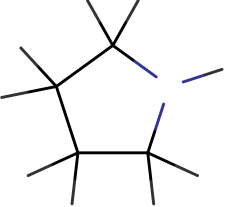
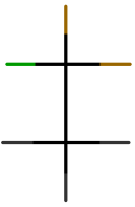
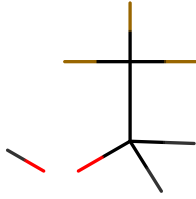
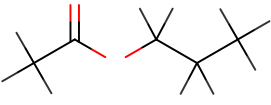
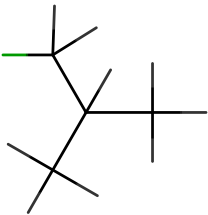
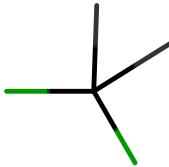
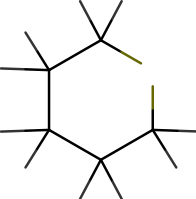
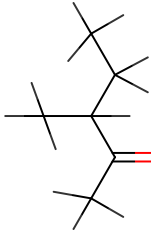
	1- пентанамин		метанамин		Циклогексан		метокс ибензол
	2-Метил-2- бутантиол		дихлорфтор метан		метанол		2,4- димети лпирид ин
	3,3-диметил- 2-тиабутан		1,2- пропадиен		2- бромпропа н		1,4- димети лбензол
	1- пентантиол		Хлор- 1,1,2,2,2- пентафторэт ан		бромэтан		2- Метил- 1- бутанти ол

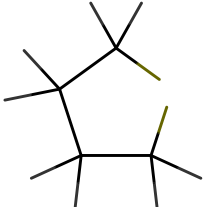
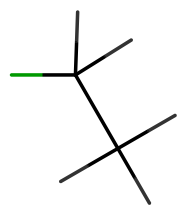
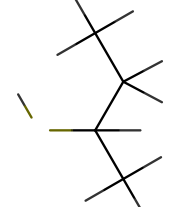
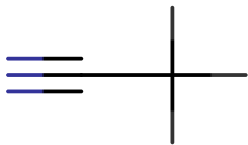
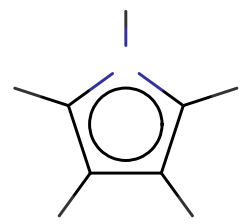
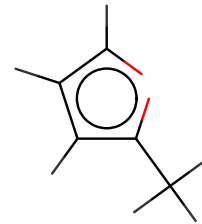
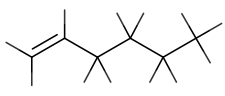
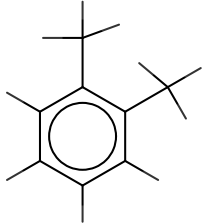
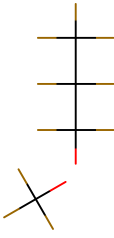
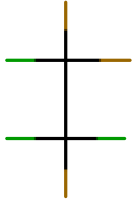
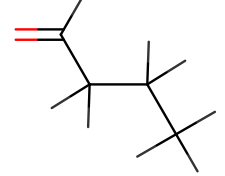
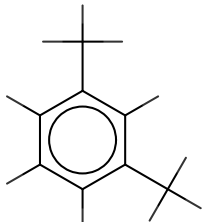
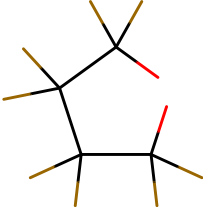
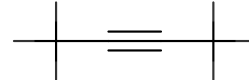
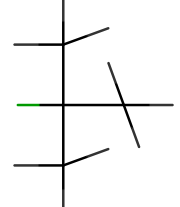
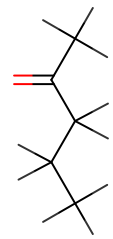
	Бутиловый эфир		1,1,1- трифторэтан		2,3- диметилбута н		2,6- димети лпирид ин
	пиперидин		дихлордифт орметан		тетрахлорме тан		2- бутано н
	2-хлор-2- метилбутан		диметилowy й эфир		Тетрагидроф уран		1- пропан ол
	циклопентан тиол		Трихлорфто рметан		(Z) -1,3- пентадиен		3- метилп иридин

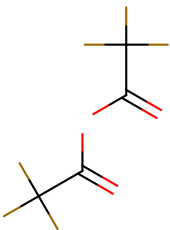
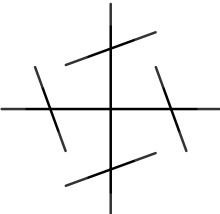

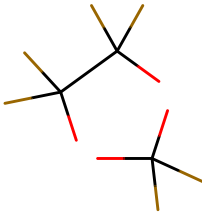
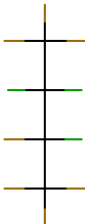
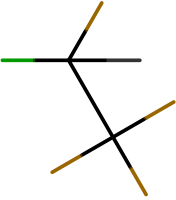
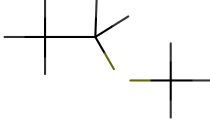
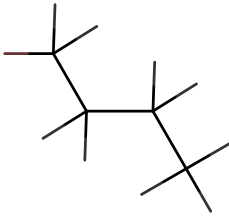
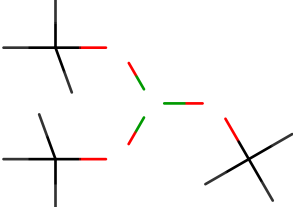
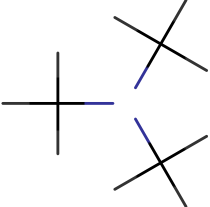
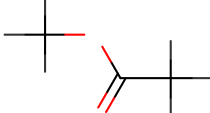
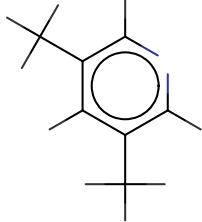

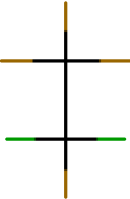
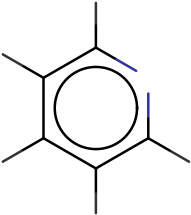
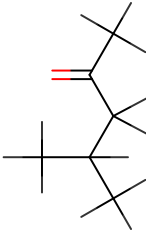
	1-метилэтил этаноеат		(Дифтормет окси) трифтормет ансульфонат		хлорбензол		бромди фторме тан
	бутил метаноеат		бромэтен		Ацетальдеги д		2,5- димети лпири дин
	3-Метил-2- бутанон		2- метилпропа н		2,2- дихлорпроп ан		(Трифт ормети л) ундека фторци клогекс ан
	2-Метил-2- бутен		1,1,1,2- тетрафторэт ан		додекафторц иклогексан		2,3- дигидр оинден

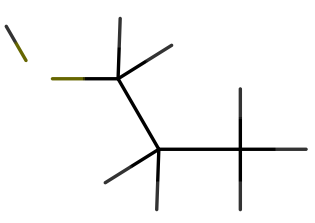
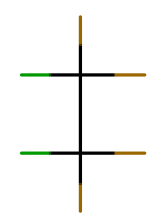
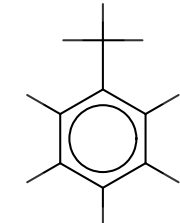
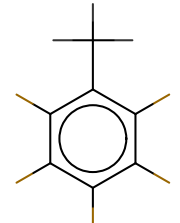
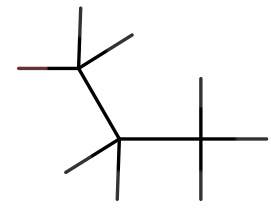
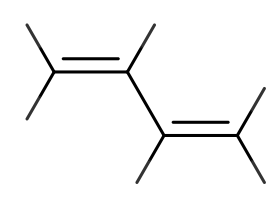
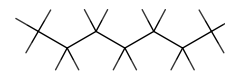
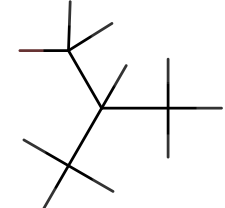
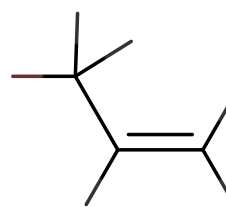
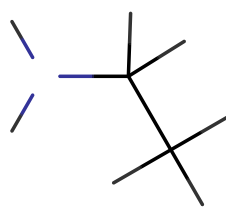
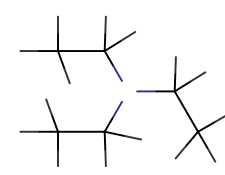
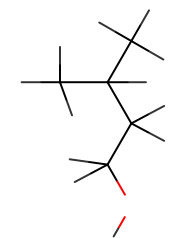
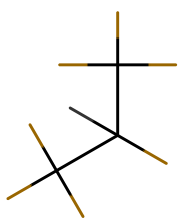
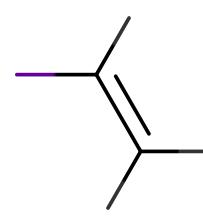
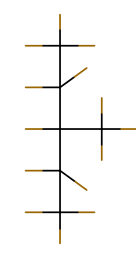
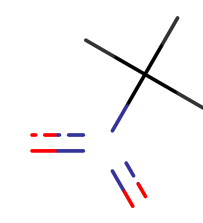
	2-Метил-1- бутен		Оксиран		1,4-диоксан		4- метилп иридин
	Циклопента н		1,1- дифторэтан		2-бромбутан		бензолт иол
	спиро [2,2] пентан		1,1,1,2,2,3,3- Гептафторп ропан		Этил метаноат		1-октен
	бутанамин		N- метилметан амин		1- этоксипропа н		этилбен зол


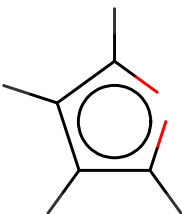
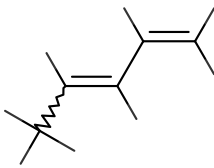
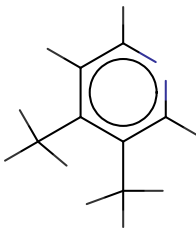
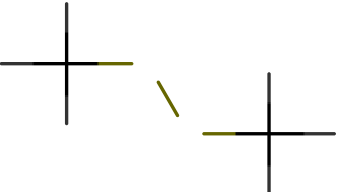
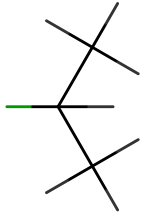
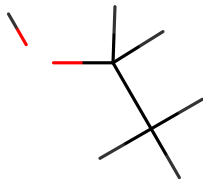
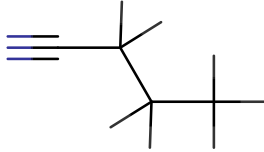
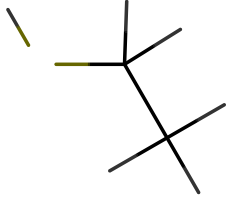
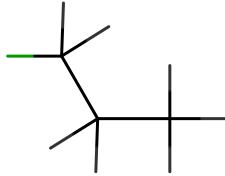
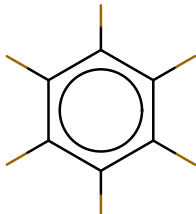
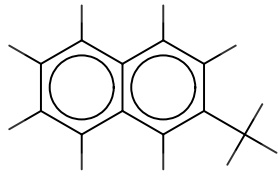
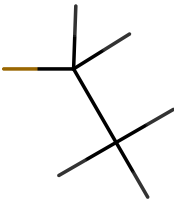
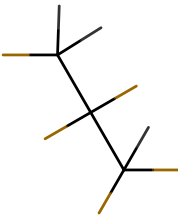
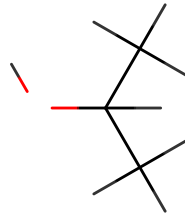
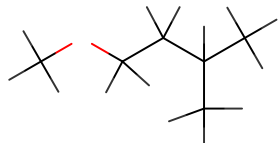
	2-тиапентан		Бутан		N-этилэтанами н		октамет илтрис илокса н
	3-Метил-2- тиабутан		октафторци клобутан		N-метилпирро л		3-пентан он
	2-Метил-1- пропантиола		метоксиэтан		3-метилпентан		2-метилп ропанн итрил
	2-Метил-2- пропантиол		трифторйод метан		гексан		3,3-димери л-2-бутано н

	1-бутантиол		2-метилпропен		оксетан		пропаналь
	Диэтиловый эфир		2,5-дигидрофуран		2-метилпентан		1Н-перфторгептан
	Пирролидин		1-хлор-1,1-дифторэтан		2,2,2-трифторэтанол		Пропил этаноат
	1-хлор-2-метилпропан		дихлорметан		тиациклогексан		3-Метил-2-пентанол

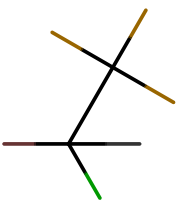
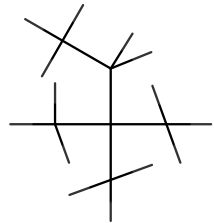
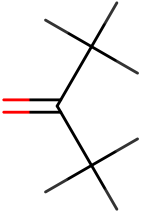
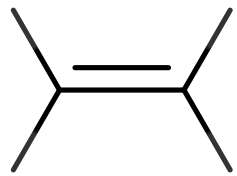
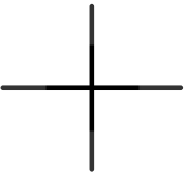
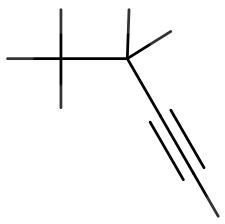
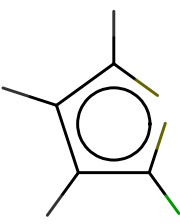
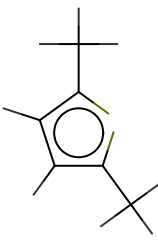
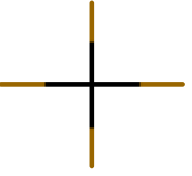
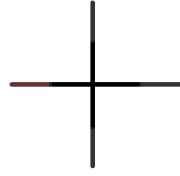
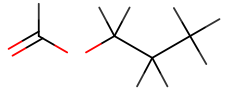
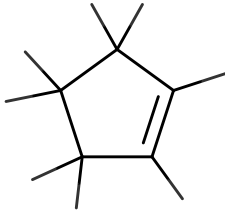
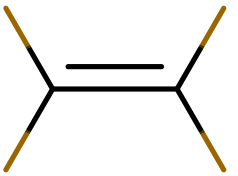
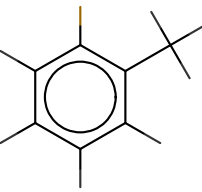
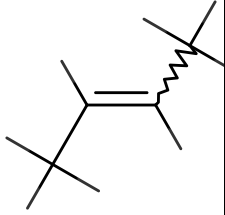
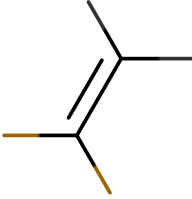
	тетрагидрот иофен		хлорэтан		2-бутантиол		Ацетон итрил
	1Н-пиррол		2- метилфуран		1-гексен		1,2- димети лбензол
	1,1,1,2,2,3,3- гептафтор-3- (трифтормет окси) пропан		1,1,2- трифтор- 1,2,2- трихлорэтан		бутаналь		1,3- димети лбензол
	Октафтороте трагидрофур ан		2-бутен		2-хлор-2- метилпропа н		2- пентан он

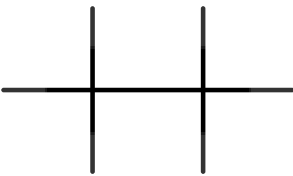
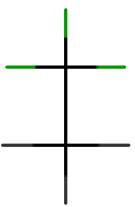
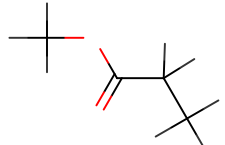
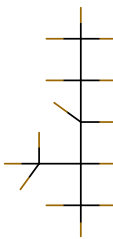
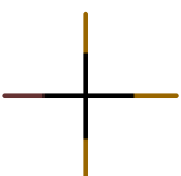
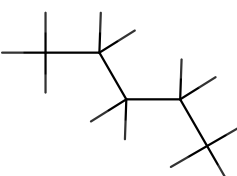
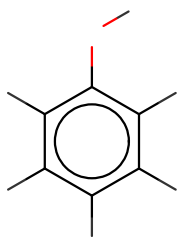
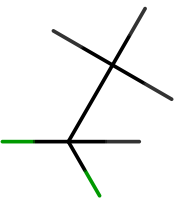
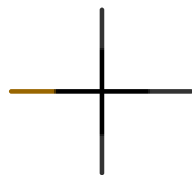
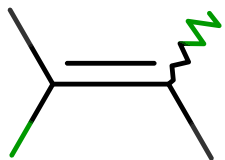
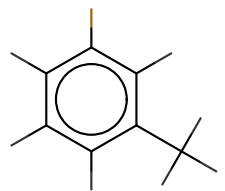
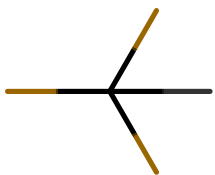
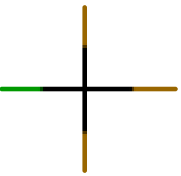
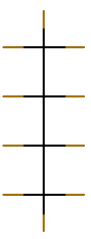

	трифторуксусный ангидрид		2,2-диметилпропан		додекафторпропан		2,2,4,4,5,5-гексафтор-1,3-диоксолан
	2,2,3-трихлоропентафторобутан		1-хлор-1,2,2,2-тетрафторэтан		2-тиабутан		1-бромбутан
	триметилборат		N, N-диметилметанамин		Метилэтанوات		3,5-диметилпиридин
	2-пропантиол		1,1-дихлор-1,2,2,2-тетрафторэтан		пиридин		4-Метил-2-пентанол

	1- пропантиол		1,2-дихлор- 1,1,2,2- тетрафторэт ан		Толуол		2,3,4,5, 6- пентаф тортолу ол
	1- бромпропа н		1,3- Бутадиен		Гептан		1-Бром- 2- метилп ропан
	3-Бром-1- пропен		этанамин		N, N- диэтилэтана мин		3- метил- 1- бутано л
	1,1,1,2,3,3,3- Гептафторпр опан		иодэтен		ундекафторо -3- (трифтормет ил) -пентан		нитром етан








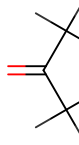
	октафторпропан		фуран		(E)-1,3-пентадиен		3,4-диметилпиридин
	2,3-дитиобутан		2-хлорпропан		Спирт этиловый		бутаннитрил
	этанол		1-хлорпропан		гексафторбензол		2-метилантрацен
	фторэтан		1,1,2,2,3-пentaфторпропан		2-пропанол		Метил 3-метилбутилэфир




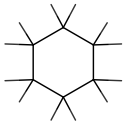

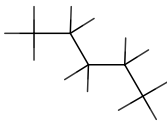
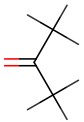
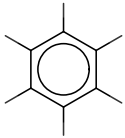



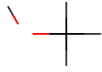
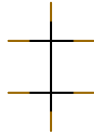
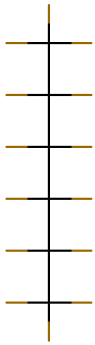
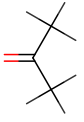
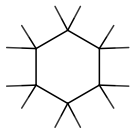



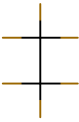


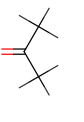

	1,1,2- трифторэтан		1,1,1,2,3,3- гексафторпропан		1-фтор-4- метилбензол		уксусная кислота
	1,1,2,2- тетрафторэтан		1,2- дихлорэтан		бис (1- метилэтил) эфир		Бензойная кислота
	(E) -1,2- дихлорэтен		1-бутен		2- метилтиофен		Этил этанойт
	1,2,2- трихлорэтен		иодметан		диэтил сульфид		2- метилбутан








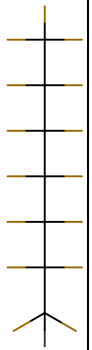






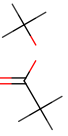
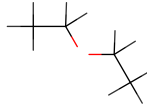





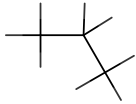
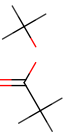
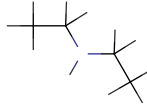
	1-Бром-1-хлор-2,2,2-трифторэтан		2,2-диметилбутан		пропанон		Этан
	метан		1-бутен		2-хлортиофен		2,5-диметилтиофен
	Тетрафторметан		бромметан		Пропил метаноат		циклопентен
	тетрафторэтен		1-фтор-2-метилбензол		(E) -2-бутен		1,1-дифторэтен




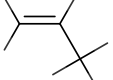
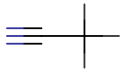


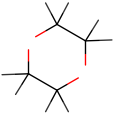

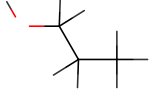

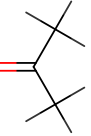
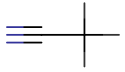
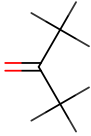
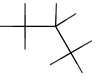
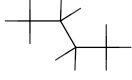

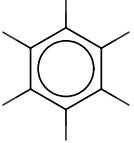

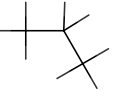
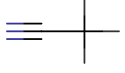
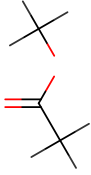

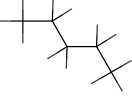
	Этан		1,1,1-трихлорэтан		Метил пропаноат		Ундека фторо-2-(трифторметил)-пентан
	трифторбромметан		пентан		Фенол		1,1-дихлорэтан
	фторметан		(Z)-1,2-дихлорэтен		1-Фтор-3-метилбензол		трифторметан
	хлортрифторметан		декафторбутан		Тетрадекафторгексан		

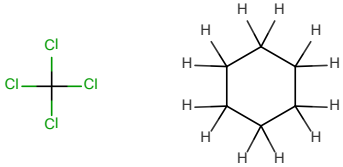
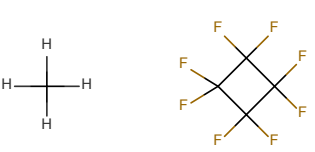
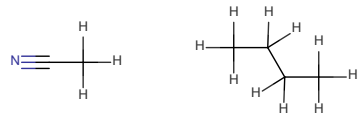
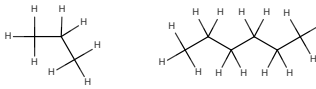

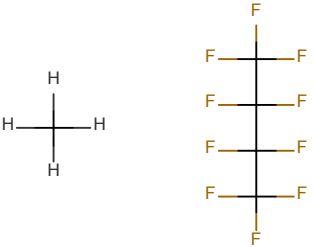
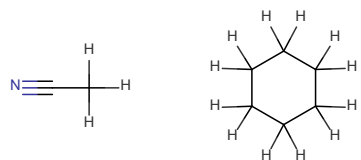
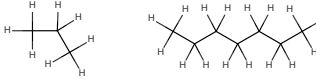

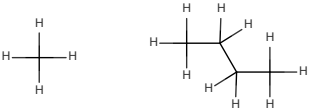
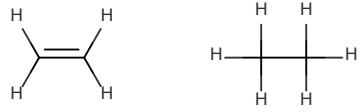
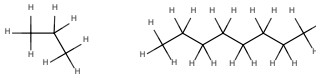
2.2. Смеси обучающей выборки для вторых вириальных кросс-коэффициентов


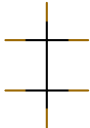

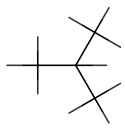





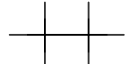

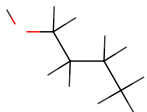

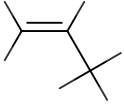

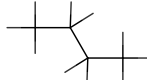

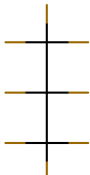

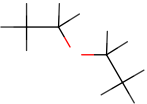

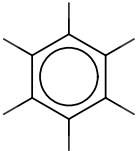


Структура	Имя	Структура	Имя	Структура	Имя	Структура	Имя
	хлортрифторметан + Трихлорфторметан		Нитрометан + пропанон		Гексафторэтан + бутан		Пропанон + бутан
	хлортрифторметан + хлордифторметан		Нитрометан + Бензол		Гексафторэтан + додекафторпропан		Пропанон + тетрадекафторгексан

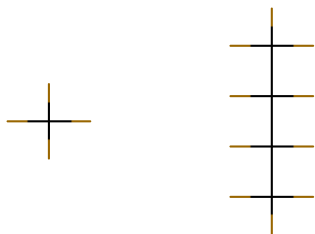
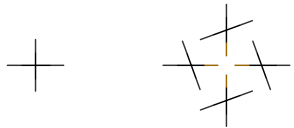
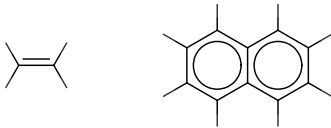
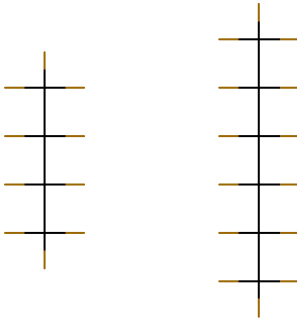

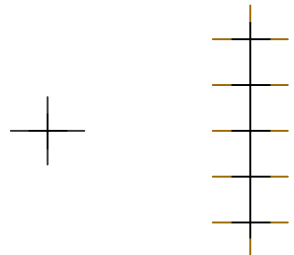
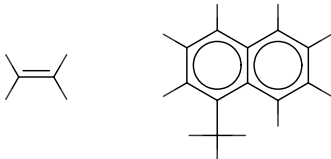
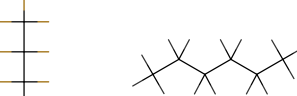
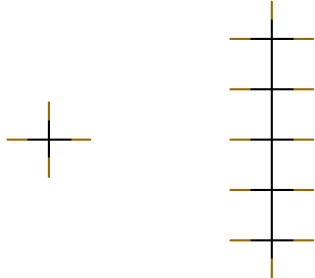

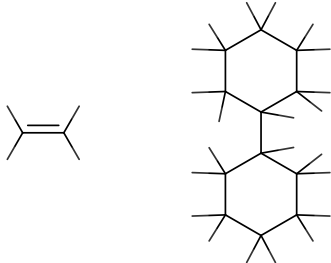
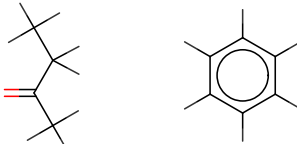
 	хлортри фтормет ан + трифтор метан	 	Нитроме тан + циклогек сан	 	Гексафто рэтан + пентан	 	Пропанон + Бензол
 	Дихлорд ифторме тан + Трихлор фтормет ан	 	Метан + Метанол	 	Гексафто рэтан + тетрадека фторгекс ан	 	Пропанон + циклогекс ан
 	Дихлорд ифторме тан + хлордиф тормета н	 	Метан + гексафто рэтан	 	Пентафт орэтан + 1,1,1,2- тетрафто рэтан	 	Пропанон + Гексан

 	Дихлордифторметан + трифторметан	 	Метан + этен	 	Pentafluoroethane + 1,1,1-трифторэтан	 	Пропанон + 1H-перфторпропан
 	Трихлорфторметан + трифторметан	 	Метан + Этан	 	1,1,1,2-тетрафторэтан + 1,1-дифторэтан	 	Метилэтаноат + Диэтиловый эфир
 	Тетрафторметан + Метан	 	Метан + Этанол	 	1,1,1,2-тетрафторэтан + Пропан	 	Метилэтаноат + Диэтиламин




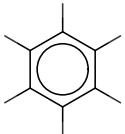


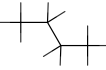
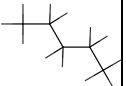





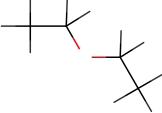
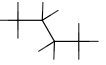




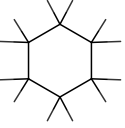
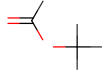



 	Тетрахл орметан + этен	 	Метан + пропен	 	Ацетони трил + Ацетальд егид	 	Пропан + 1,4- диоксан
 	Тетрахл орметан + 1- пропано л	 	Метан + пропанон	 	Ацетони трил + пропанон	 	Пропан + бутан
 	Тетрахл орметан + Бензол	 	Метан + Пропан	 	Ацетони трил + Метил этаннат	 	Пропан + пентан


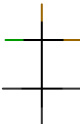



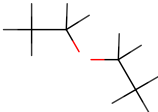
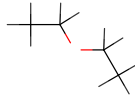
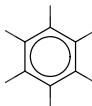



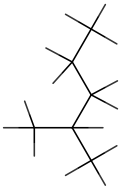

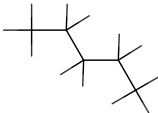
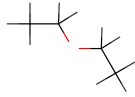
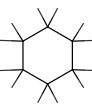



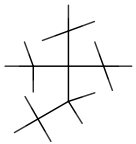

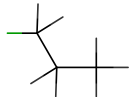
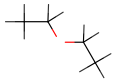

	Тетрахлорметан + циклогексан		Метан + октафторциклобутан		Ацетонитрил + бутан		Пропан + Гексан
	Тетрафторметан + трифторметан		Метан + декафторбутан		Ацетонитрил + циклогексан		Пропан + гептан
	Тетрафторметан + Метан		Метан + бутан		Этен + Этан		Пропан + октан




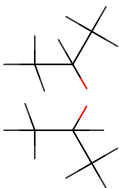

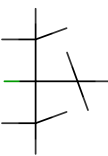
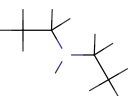


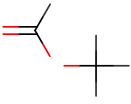

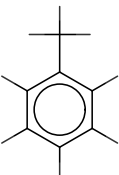
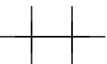
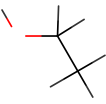

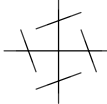

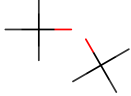



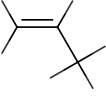
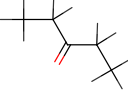

 	Тетрафт орметан + гексафт орэтан	 	Метан + 2- метилпро пан	 	Этен + Этанол	 	1- пропанол + гептан
 	Тетрафт орметан + Этан	 	Метан + 1- бутанол	 	Этен + пропен	 	Декафтор бутан + бутан
 	Тетрафт орметан + октафто рпропан	 	Метан + Диэтилов ый эфир	 	Этен + Бензол	 	Декафтор бутан + додекафто рпропан




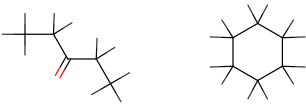
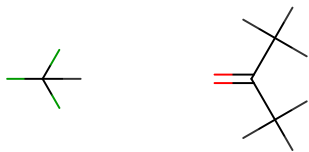

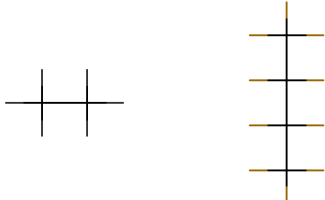
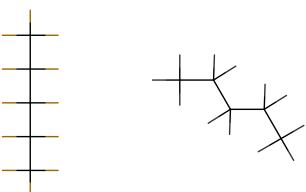
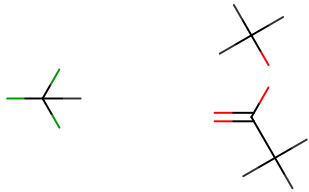

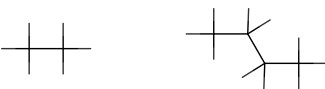
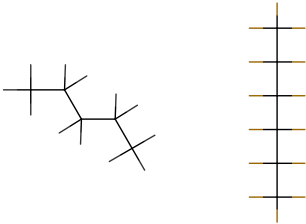
	Тетрафторметан + декафторбутан		Метан + тетраметилсилан		Этен + Нафталин		Декафторбутан + тетрадекафторгексан
	Тетрафторметан + бутан		Метан + додекафторпропан		Этен + 1-метилнафталин		Декафторбутан + Гексан
	Тетрафторметан + додекафторпропан		Метан + пентан		Этен + бициклогексил		Бутанон + Бензол




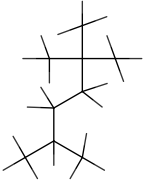

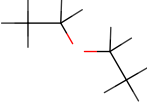
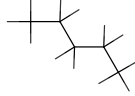
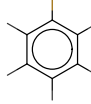

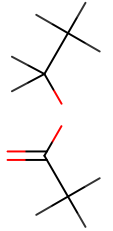

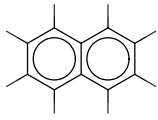

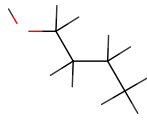
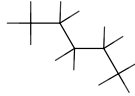
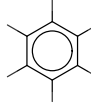

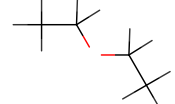

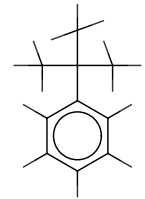

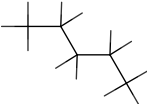
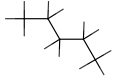

	Тетрафторметан + 2,2-диметилпропан		Метан + 2-метилбутан		Этен + дифенилметан		1,4-диоксан + Бензол
	Тетрафторметан + тетрадекафторгексан		Метан + 2,2-диметилпропан		Этен + Фенантрен		1,4-диоксан + Циклогексан
	Тетрафторметан + Гексан		Метан + тетрадекафторгексан		Этен + Антрацен		Бутан + 2-метилпропан


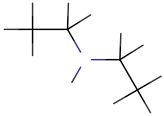



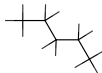


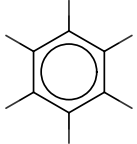
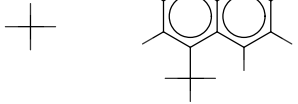


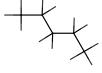


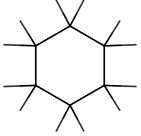
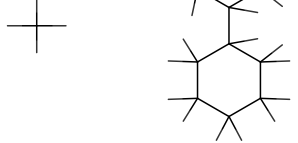

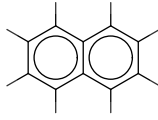
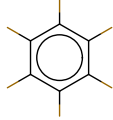
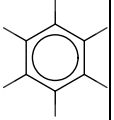
 	хлордифторметан + трифторметан	 	Метан + Бензол	 	Этен + гексадекан	 	Бутан + пентан
 	Хлордифторметан + 1,2-дихлор-1,1,2,2-тетрафторэтан	 	Метан + 2,5-гександион	 	Оксиран + Диэтиловый эфир	 	Бутан + Гексан
 	Хлордифторметан + 1,1-дихлор-2,2,2-трифторэтан	 	Метан + циклогексан	 	Метил метаноат + Пропил метаноат	 	Бутан + октан




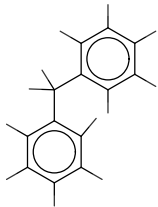

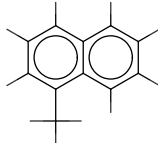
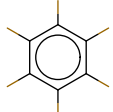
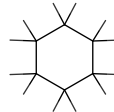

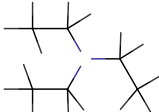

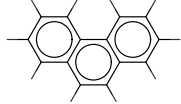

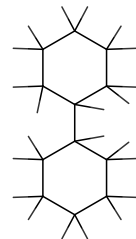



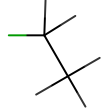

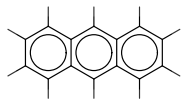

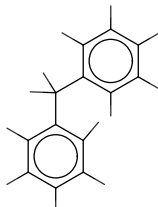
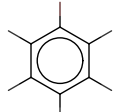
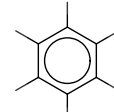
 	Хлорди фтормет ан + 1- хлор- 1,1- дифторэ тан	 	Метан + Гексан	 	Бромэтан + Диэтилов ый эфир	 	Диэтилов ый эфир + Бензол
 	Хлорди фтормет ан + 1,1- дифторэ тан	 	Метан + 2- метилпен тан	 	Бромэтан + пентан	 	Диэтилов ый эфир + циклогекс ан
 	Трихлор метан + Метан	 	Метан + 2,2- диметил бутан	 	Хлорэтан + 1- хлорпроп ан	 	Диэтилов ый эфир + Гексан


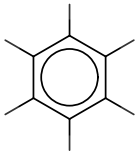


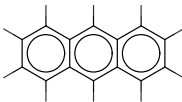
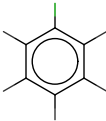
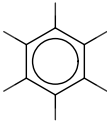

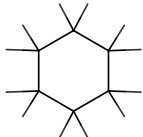



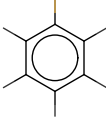
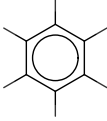

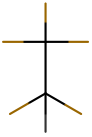
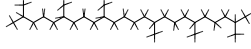


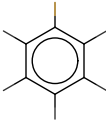
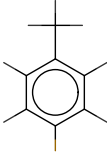
 	Трихлор метан + этен	 	Метан + диизопро пиловый эфир	 	Хлорэтан + 2-хлор- 2- метилпро пан	 	Диэтила ин + циклогекс ан
 	Трихлор метан + Метил метаноа т	 	Метан + Толуол	 	Этан + Этанол	 	Тетрамети лсилан + 2,2- диметилп ропан
 	Трихлор метан + диметил овый эфир	 	Метан + гептан	 	Этан + пропен	 	3- пентанон + Бензол





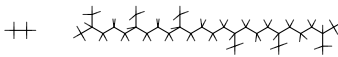
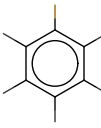
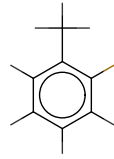




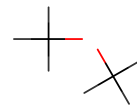
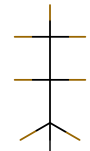
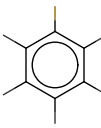
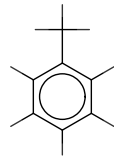



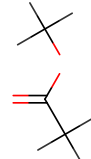


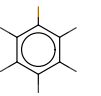

	Трихлор метан + Этанол		Метан + октан		Этан + Пропан		3- пентанон + циклогекс ан
	Трихлор метан + пропано н		Метан + 2,5,8,11- тетраокс адодекан		Этан + декафтор бутан		Додекафт орпропана + пентан
	Трихлор метан + Метил этаноеат		Метан + нонан		Этан + бутан		Пентан + тетрадека фторгекса н


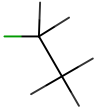

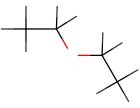

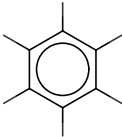



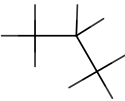

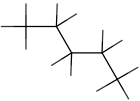

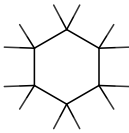
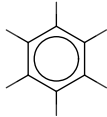
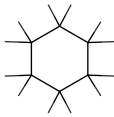

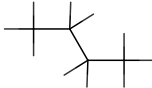

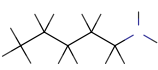




 	Трихлор метан + Пропил метаноа т	 	Метан + 2,2,5- триметил гексан	 	Этан + Диэтилов ый эфир	 	Пентан + фторбензо л
 	Трихлор метан + Этил этаноеат	 	Метан + Нафтали н	 	Этан + 1- бутанол	 	Пентан + Бензол
 	Трихлор метан + Диэтило вый эфир	 	Метан + (1,1- диметилэ тил) бензол	 	Этан + пентан	 	Пентан + Гексан

 	Трихлор метан + диэтила мин		Метан + Декан	 	Этан + Гексан	 	Пентан + Гептан
 	Трихлор метан + Бензол		Метан + 1- метилна фталин	 	Этан + Октан	 	Пентан + октан
 	Трихлор метан + циклоге ксан		Метан + бициклог ексил	 	Этан + Нафтали н	 	Гексафтор бензол + Бензол


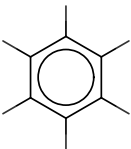

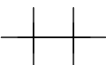
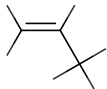
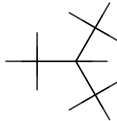
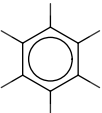
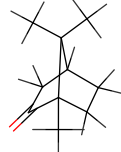

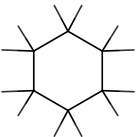
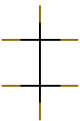






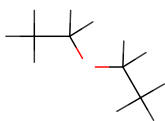
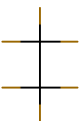
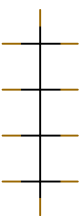
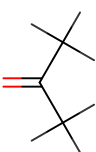
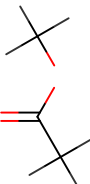


 	Трихлор метан + Гексан	 	Метан + дифенил метан	 	Этан + 1- метилна фталин	 	Гексафтор бензол + циклогекс ан
 	Трихлор метан + Триэтил амин	 	Метан + Фенантр ен	 	Этан + бициклог ексил	 	тетрадека фторгекса н + Гексан
 	Дихлор метан + хлорэта н	 	Метан + Антраце н	 	Этан + дифенил метан	 	Бромбенз ол + Бензол



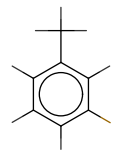
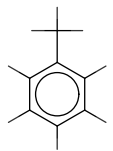
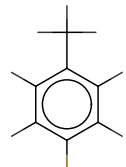
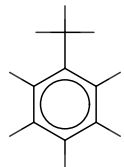
 	Дихлор метан + Бензол	+		Метан + гексадека н	 	Этан + Антраце н	 	Хлорбенз ол + Бензол
 	Дихлор метан + циклоге ксан	+		Метан + эйкозан	 	Этан + гексадека н	 	Фторбенз ол + Бензол
 	Дифтор метана + пентафт орэтан	+		Метан + 2,6,10,15, 19,23- гексамет илтетрак озан	 	Этан + эйкозан	 	Фторбенз ол + 4- фтортолуо л

 	Дифторметан + 1,1,1,2-тетрафторэтан	 	Метанол + этен		Этан + 2,6,10,15,19,23-гексаметильтетракозан	 	Фторбензол + 2-фтортолуол
 	Бромметан + Хлорметан	 	Метанол + Этан	 	Диметилэтер + 1H-перфторпропан	 	Фторбензол + Толуол
 	Бромметан + бромэтан	 	Метанол + Метилэтанойт	 	Этанол + пентиламин	 	Фторбензол + гептан

 	Бромметан + хлорэтан	 	Метанол + Диэтиловый эфир	 	Этанол + Бензол	 	Фторбензол + октан
 	Бромметан + Пропан	 	Метанол + пентан	 	Этанол + циклогексан	 	Бензол + циклогексан
 	Бромметан + бутан	 	Метанол + пентиламин	 	Этанол + Гексан	 	Бензол + Гексан

	Бромметан + пентан		Метанол + Бензол		Октафторпропан + Пропан		Бензол + Толуол
	Хлорметан + Этан		Метанол + циклогексан		Пропен + Пропан		Бензол + гептан
	Хлорметан + пропанон		Метанол + Гексан		Пропен + бутан		Бензол + октан

 	Хлорметан + Бензол	 	Гексафторэтане + Этан	 	Пропен + 2-метилпропан	 	Бензол + 1,7,7-триметилбикло[2.2.1]-2-гептанон
 	Хлорметан + циклогексан	 	Гексафторэтане + октафторпропан	 	Пропен + 1-гептен	 	Циклогексан + Гексан
 	Иодметан + Диэтиловый эфир	 	Гексафторэтане + декафторбутан	 	Пропанон + Метилэтанолат	 	Гексан + октан

 	Гептан + Октан	 	3- фтортолу ол + Толуол	 	4- фтортолу ол + Толуол		
---	-------------------	---	----------------------------------	---	----------------------------------	--	--