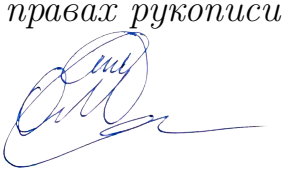
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ
НАУКИ ИНСТИТУТ ПРОБЛЕМ УПРАВЛЕНИЯ
ИМ. В.А. ТРАПЕЗНИКОВА РОССИЙСКОЙ АКАДЕМИИ НАУК

На правах рукописи



Милосердов Олег Александрович

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ПОЛИМЕРНЫХ ЦЕПЕЙ В ЗАДАЧАХ ПРЕДСКАЗАНИЯ ТРАНСПОРТНЫХ ХАРАКТЕРИСТИК СТЕКЛООБРАЗНЫХ ПОЛИМЕРОВ

Специальность 1.2.2 — Математическое моделирование, численные методы и комплексы программ (технические науки).

Научно-квалификационная работа (диссертация) на соискание ученой степени кандидата технических наук

Научный руководитель:
д.ф.-м.н., Профессор РАН
Губко Михаил Владимирович

Москва — 2022

Содержание

Введение

Многие научные и технологические задачи, стоящие перед обществом в XXI веке, от медицины и здравоохранения до производства и хранения энергии, имеют как минимум одну общую потребность – это новые материалы, способные обеспечить возрастающие с каждым годом потребности человечества [95].

Развитие информационных технологий позволяет вывести на новый уровень получение новых материалов, обладающих заданными свойствами. В основе *хемоинформатики* – науки на стыке химии и компьютерных наук – лежит представление о химическом пространстве как множестве всевозможных химических объектов, в частности, химических соединений или состоящих из них материалов [64, 78]. Понятно, что из-за обширности этого множества человечество пока прикоснулось лишь к малой его части. Одна из крупнейших молекулярных баз данных [88] включает в себя более 166 миллиардов молекул, которые содержат не более 17 тяжелых атомов. При этом количество структур маленьких молекул, которые потенциально могут использоваться в фармакологии, оценивается примерно в 10^{60} [107]. Поэтому исследование этого пространства может показаться невозможным.

Молекулярное моделирование, один из разделов хемоинформатики, предлагает способ исследования химического пространства без осуществления реальных экспериментов, основанный на разработке теоретических и вычислительных методов для моделирования и изучения поведения молекул, от небольших систем, характерных для неорганической химии, до больших полимерных или биологических систем. Области применения молекулярного моделирования включают вычислительную химию, разработку лекарств, вычислительную биологию и материаловедение.

Использование методов машинного обучения в последние годы играет важную роль в развитии методов молекулярного моделирования. Одним из привлекательных подходов, основанных на данных, является автоматическое улучше-

ние качества моделирования с поступлением новой информации. В контексте науки о материалах методы машинного обучения часто используются для прогнозирования физико-химических свойств материалов путем построения функции, сопоставляющей «молекулярный материал», то есть измеримые свойства молекул, со значением выбранного физического или химического свойства, такого как температура кипения, плавления, стеклообразования, плотность, проницаемость или биологическая активность.

Методы молекулярного моделирования обычно требуют больших вычислительных мощностей и значительного времени вычислений. Тем не менее, они позволяют получить существенный экономический эффект, который складывается из стоимости экспериментов (разработки методики синтеза вещества и изменений его желаемых свойств), замененных вычислениями по модели и, что зачастую более важно, экономии времени за счет ускорения описанных выше циклов и повышения эффективности каждого цикла за счет все более точного моделирования потенциальной пригодности материала.

Мембранное газоразделение – сравнительно новое направление промышленного разделения газовых смесей, быстро развивающееся благодаря ряду технологических преимуществ по сравнению с традиционными методами разделения газов (ректификацией, абсорбцией и адсорбцией) [20, 23, 111]. К достоинствам мембранного газоразделения относятся низкое энергопотребление, отсутствие фазовых переходов, малая материалоемкость, гибкость управления и модульность структуры технологических установок. Наибольшее распространение получили промышленные процессы мембранного газоразделения для выделения азота из воздуха, удаления углекислого газа из «кислых» природных газов и извлечение водорода из различных потоков химии и нефтепереработки.

В основе мембранного газоразделения лежит квалифицированный выбор материала мембран и возможность приготовления мембран, которые были бы и высокопроизводительными (то есть, пропускающими через себя большие объемы газовой смеси) и селективными (то есть, обеспечивающими разделение смеси

на компоненты с минимальными долями примесей). В подавляющем большинстве процессов мембранного газоразделения используются полимеры. Синтез или выбор полимера и изучение его свойств и поведения в процессе газоразделения является предметом мембранного материаловедения. Большей частью существующие процессы мембранного газоразделения основаны на использовании мембранных материалов, созданных десятилетия назад. Благодаря активности синтетической полимерной химии каждый год появляются все новые полимеры, отличающиеся улучшенными газоразделительными (т.н., транспортными) параметрами, и важнейшей задачей мембранного материаловедения является поиск новых высокоэффективных полимеров для решения разнообразных задач мембранной технологии. В частности, крайне важен скрининг и анализ имеющихся данных по связи структуры и характеристик полимерных материалов для направленного поиска подобных материалов и создания на их основе новых газоразделительных мембран.

Актуальность разработки надежного теоретического метода расчета и прогнозирования транспортных характеристик полимерных материалов обусловлена, в частности, сложностями экспериментального определения этих характеристик. Для проведения экспериментов требуются реагенты, специальные условия для синтеза, а также самое ценное – время ученых синтетиков. Расчетные методы квантовой химии и молекулярной динамики имеют ряд ограничений, связанных с подвижностью многих макромолекул. Также критичным является время расчета, которое часто оказывается очень велико, когда речь идет о скрининге и переборе большого числа различных молекул. Альтернативой этим подходам являются методы QSPR (Quantitative Structure-Property Relationship), которые успешно применяются для прогнозирования свойств химических соединений. В методах QSPR используются регрессионные или классификационные модели. Регрессионные модели связывают набор независимых переменных или предикторов (X) с зависимой переменной (Y), в то время как классификационные модели QSPR связывают переменные-предикторы с категориальным значени-

ем переменной Y. Таким образом, в методах QSPR предикторами являются различные структурные характеристики молекул, в то время как предсказываемой переменной является некоторое физическое или химическое свойство данной молекулы.

В настоящей диссертации описываются развитие и результаты применения на практике нового метода прогнозирования транспортных характеристик стеклообразных полимеров, т.н. «Предсказания на основе Поверхности Коротких Полимерных Цепей» (ППКПЦ) (Short Polymer Chain Surface Based Prediction), в основе которого лежит моделирование конформаций молекулы полимера размером порядка нескольких сотен атомов. В качестве примера применимости приложений метода предсказываются коэффициент растворимости при бесконечном разбавлении S и константы равновесия закона Генри в модели двойной сорбции k_D. Подход заключается в расчете площадей поверхности «обкатки» шаровой модели молекулы с помощью алгоритма Ли-Ричардса [67] для большой базы конформаций отдельных полимерных цепей и в построении на их основе кривых зависимости площадей «обкатки» от радиуса «обкатки». Коэффициенты линейной аппроксимации полученных зависимостей или эти зависимости целиком используются как объясняющие переменные во множественной линейной регрессии или в задаче классификации. Значимые переменные и их веса в регрессии находятся на основе экспериментальных измерений из Базы данных физико-химических свойств полимеров Лаборатории мембранного газоразделения Института нефтехимического синтеза Российской академии наук [15].

Конформации полимерных цепей являются результатом молекулярно механического моделирования, процедура которого реализована в среде RDKit (Python). Разработанный метод ППКПЦ реализован на свободно распространяемом ПО, максимально автоматизирован, имеет возможность распараллеливания на кластере и приемлемое время расчета одного полимера. Также метод обеспечивает стабильность получаемых результатов и их воспроизводимость, и

применим для специфических полимеров, используемых в мембранном газоразделении. Визуализация метода ППКПЦ представлена на рисунке 1.

С помощью разработанного метода ППКПЦ был предсказан коэффициент растворимости газов при бесконечном разбавлении S и константа растворимости закона Генри k_D модели двойной сорбции для стеклообразных полимеров. Для S предложено два типа регрессий: универсальная и частные регрессии по газам. Для универсальной регрессиий коэффициент детерминации на тестовой выборке составил $R^2 = 0.72$, а средняя относительная ошибка $MPE = 104\%$, что лучше чем у конкурентов на момент выполнения исследования. Частные регрессии по газам показали более высокие результаты, например значения средней относительной ошибки на тестовой выборке для наиболее популярных газов, таких как CO_2, N_2, O_2, CH_4, H_2 варьируется от 31 до 65 %.

Для полимеров из работы [14] с использованием частных газовых регрессий был предсказан коэффициент растворимости S. Коэффициенты растворимости, предсказанные с помощью метода, сравнивались с коэффициентами растворимости, рассчитанными методами MAC (modified method of atomic contributions) [113] и BC (Bond contribution method) [96]. Ошибки методов MAC / BC и ППКПЦ находятся в пределах одного порядка величины (за исключением двух точек, полученных методом BC).

Несмотря на меньший объем выборки, отсутствие нормировки объясняющих переменных на экспериментальную плотность образца, а также на отсутствие нормализации значений k_D по температуре, полученная универсальная регрессия для предсказания константы закона Генри k_D показала неплохой результат. Коэффициент детерминации регрессии составил 0.81 на тестовой выборке (при корреляции 0.9), что соответствует средней относительной ошибке предсказания k_D 47% на тестовой выборке.

Таким образом, разработанный метод ППКПЦ был применен для предсказания важнейших транспортных характеристик полимеров, используемых в газоразделительных мембранах, а также было проведено сравнение результатов

предсказания разработанного метода с методами МАС / ВС, являющимися методами групповых вкладов (см. подробнее раздел 1.3.3 ниже).

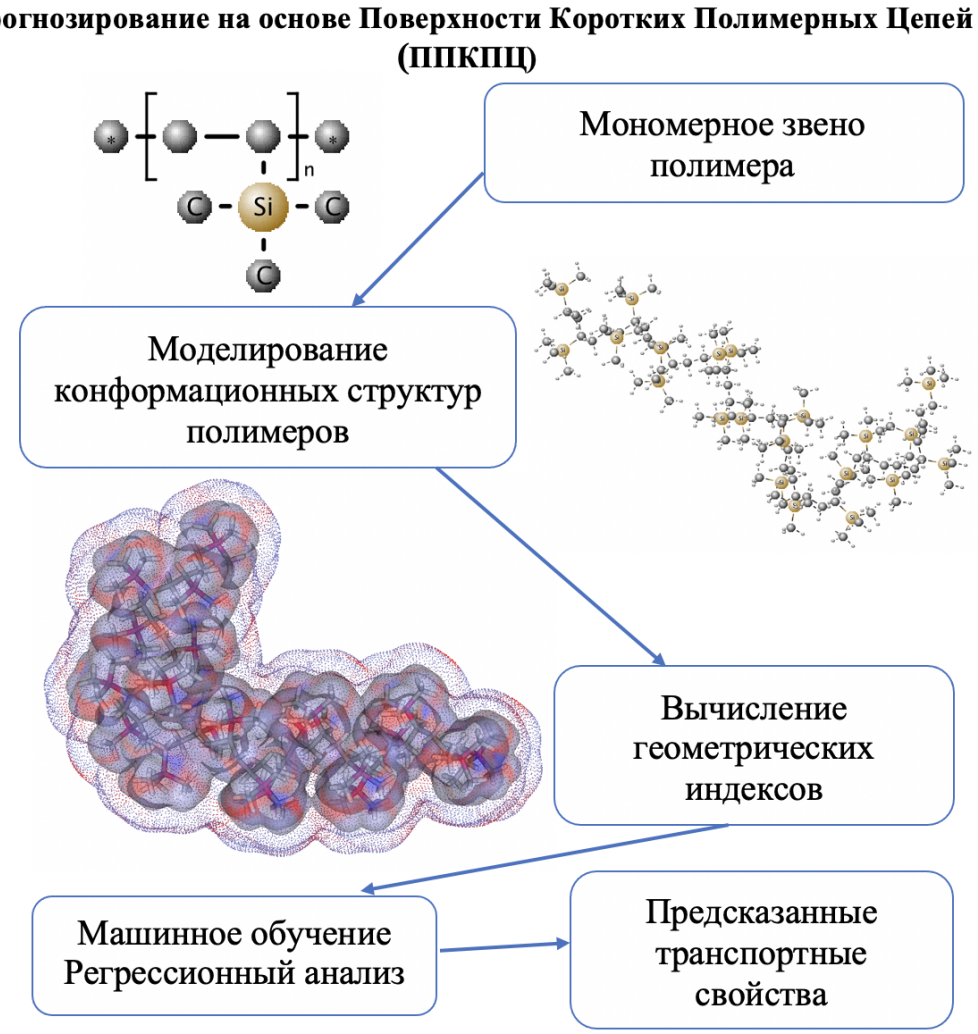


Рисунок 1 – Алгоритм метода ППКПЦ.

Актуальность темы исследования

Мембранное газоразделение является одним из быстро развивающихся направлений мембранной науки и технологий. Полимерные газоразделительные мембраны имеют широкий круг применения в различных отраслях промышленности:

- концентрирование водорода из отходящих газов каталитического риформинга, сбросных газов нефтехимии, для последующего применения в

процессах гидрирования для получения ценных химических продуктов, очистки нефти и т. д.;

- получение азота для создания инертной среды и обеспечения пожаро- и взрывобезопасности при хранении опасных веществ, нефтепродуктов, сжиженных углеводородов, тушении пожаров в шахтах, обеспечении условий для длительного хранения пищевых продуктов;
- обогащение воздуха кислородом для обеспечения медицинских нужд и технологических процессов в металлургии.

Однако большинство используемых в современных технологических процессах материалов полимерных мембран разработаны в 80х годах XX века и по своим характеристикам не в полной мере соответствуют современным задачам мембранного газоразделения. Поэтому требуются новые материалы для создания мембран, отличающихся улучшенными транспортными характеристиками. Проведение экспериментов и синтез кажущихся перспективными полимеров требует больших средств и времени, поэтому большое значение имеют математические модели, способные предсказывать характеристики интересующих нас полимеров по их структуре. Эти модели позволяют получить структурную формулу полимера с оптимальным набором транспортных характеристик, тем самым экономятся финансовые и временные ресурсы, которые были бы затрачены на поиск, подбор и синтез полимеров с худшими характеристиками. Существующие модели и методы обладают рядом недостатков, которые не позволяют использовать их для дизайна новых материалов, что обуславливает актуальность разработки улучшенных методов предсказания транспортных характеристик полимерных материалов.

Объектом исследования являются транспортные характеристики пористых материалов на основе различных классов химических веществ.

Предметом исследования – математические модели, алгоритмы и комплексы программ автоматизированного поиска перспективных мембранных ма-

териалов на основе количественного прогнозирования транспортных характеристик материалов по структурной формуле их молекул.

Цели и задачи

Целью исследования является разработка математических моделей молекул стеклообразных полимеров и численных методов расчета характеристик молекулярной поверхности их полимерных цепей для создания алгоритмов количественного прогнозирования транспортных характеристик газоразделительных мембран на основе этих полимеров.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Провести анализ существующих математических моделей стеклообразных полимеров и методов предсказания их транспортных характеристик.
2. Разработать математические модели поверхности молекул аморфных полимеров.
3. Предложить численные методы предсказания транспортных характеристик аморфных полимеров на основе площади поверхности их молекул.
4. Разработать комплекс программ, основанный на библиотеках с открытым исходным кодом, позволяющий провести как моделирование от мономерного звена полимера до его конформаций, так и рассчитать необходимые переменные для предсказания характеристик полимерных газоразделительных мембран.
5. Продемонстрировать эффективность разработанных методов и алгоритмов для задач анализа и синтеза полимерных материалов с заданными свойствами в интересах мембранной технологии.

Область исследования

Диссертационная работа соответствует специальности 1.2.2 «Математическое моделирование, численные методы и комплексы программ (технические науки)» по следующим пунктам:

1. Разработка новых математических методов моделирования объектов и явлений;

2. Развитие качественных и приближенных аналитических методов исследования математических моделей;
3. Реализация эффективных численных методов и алгоритмов в виде комплексов проблемно-ориентированных программ для проведения вычислительного эксперимента;
4. Комплексные исследования научных и технических проблем с применением современной технологии математического моделирования и вычислительного эксперимента; натурного эксперимента на основе его математической модели;
5. Разработка систем компьютерного и имитационного моделирования.

Методы исследования

В диссертационной работе применяются методы молекулярно-механического моделирования, машинного обучения, математической статистики. Используются статистические критерии, регрессионный анализ, регрессии с пошаговым отбором переменных, стратегии кросс-валидации, методы кластеризации. При реализации метода ППКПЦ используется среда Python и пакеты RDKit, pandas, multiprocessing, numpy, scipy, sklearn, statsmodels.

Научная новизна

1. На основе подхода QSPR (предсказания физико-химических свойств веществ по их структурным химическим формулам) предложен отличающийся от аналогов метод и алгоритм моделирования геометрии коротких отрезков молекул полимерных материалов, вычисления для них числовых геометрических индексов, а также предсказания транспортных характеристик материалов методами машинного обучения и регрессионного анализа.
2. Для предсказания физико-химических (в первую очередь, транспортных) характеристик полимеров предложено новое семейство геометрических

молекулярных дескрипторов, основанных на анализе кривых зависимости площади доступной поверхности молекул от радиуса «обкатки».

3. Предложенный общий алгоритм предсказания транспортных характеристик с использованием разработанных новых геометрических индексов впервые реализован в виде комплекса программ, автоматизирующего все процессы моделирования транспортных характеристик от загрузки исходных наборов до расчета прогнозируемых значений транспортных характеристик новых полимерных материалов по их структурной формуле.

Теоретическая и практическая значимость работы

Теоретическая значимость разработанного метода состоит в том что он позволяет решать задачу предсказания транспортных характеристик полимерных материалов. В совокупности с комплексом программ, разработанным на языке Python, он позволяет вырабатывать рекомендации по синтезу перспективных соединений даже в условиях работы на персональном компьютере.

Практическая значимость новых алгоритмов и разработанных программных средств продемонстрирована на задачах:

а) предсказания коэффициента растворимости (при бесконечном разбавлении) легких газов в стеклообразных полимерах различных классов, используемых в мембранной технологии (в частности, впервые построена универсальная регрессия, позволяющая предсказывать значение коэффициента растворимости при бесконечном разбавлении большинства легких газов в стеклообразных полимерах самых разных химических классов),

б) прогнозирования коэффициента растворимости и селективности растворимости новых полимеров,

в) предсказания константы Генри стеклообразных полимеров различных химических классов,

г) кластеризации полимерных материалов различных классов на основе геометрии их молекул для их типизации в интересах мембранной технологии и для исследования их физико-химических свойств.

Основные результаты, выносимые на защиту

1. Методы математического моделирования конформаций полимерных цепей и их геометрических свойств;
2. Регрессионные модели, позволяющие предсказать транспортные характеристики полимерных мембран;
3. Метод кластеризации конформационных структур аморфных полимеров, позволяющий провести параллели между транспортными характеристиками полимерных мембран и геометрической формой их полимерных цепей, а также показывающий эффективность разработанных методов и алгоритмов;
4. Комплекс программ, позволяющий полностью автоматизировать процесс получения молекул полимеров большого размера, а также процесс расчета необходимых индексов для предсказания транспортных характеристик полимеров.

Степень обоснованности и достоверности полученных результатов

Представленные в работе результаты решения поставленных задач являются достоверными и обоснованными, поскольку они используют корректные статистические методы в процессе проведения вычислений, а также по причине использования представительных валидационных выборок для проверки качества полученных регрессий. Разрабатываемые алгоритмы и методы прошли апробацию публикациями в международных журналах, а о результатах исследования доложено на нескольких конференциях. Для подтверждения работоспособности регрессионных моделей были предсказаны характеристики еще не синтезированных полимеров, результаты предсказания сравнивались с методом групповых вкладов [14,96,113] , а затем и с экспериментально измеренными характеристиками новых полимерных материалов.

Реализация и внедрение результатов исследования

В среде Python разработан комплекс программ, основанный на библиотеках с открытым исходным кодом, позволяющий:

- провести моделирование конформационных структур полимеров, используемых в мембранном газоразделении,
- рассчитать необходимые переменные для предсказания характеристик полимерных газоразделительных мембран,
- построить собственные регрессии для большого набора данных,
- предсказать транспортные характеристики полимерных газоразделительных мембран на основе предложенных регрессий,
- провести кластерный анализ конформационных структур аморфных полимеров на основе их геометрии для дальнейшего анализа связей между транспортными характеристиками полимерных мембран и кластерами.

Разработанный комплекс программ прогнозирования характеристик полимерных материалов использовался в ходе научных исследований Лабораторией мембранного газоразделения ИНХС РАН для прогнозирования коэффициента растворимости новых перспективных полимерных материалов.

Апробация результатов

По тематике диссертационной работы были сделаны доклады на следующих российских и международных конференциях:

- 57-я научная конференция МФТИ: Радиотехника и кибернетика, 24–29 ноября 2014 года;
- XII Всероссийская школа-конференция молодых ученых УБС 2015: Управление техническими системами и технологическими процессами, 7-11 сентября 2015 года;
- 58-я научная конференция МФТИ: Радиотехника и кибернетика, 23–28 ноября 2015 года;

- XIII Всероссийская научная конференция (с международным участием) Мембраны-2016, 10-14 октября 2016 года;
- 60-я научная конференция МФТИ: Радиотехника и кибернетика, 20–25 ноября 2017 года;
- XV Всероссийская школа-конференция молодых ученых УБС 2018: Управление техническими системами и технологическими процессами, 10-13 сентября 2018 года;
- 61-я научноя конференция МФТИ: ФРТК Секция интегрированных киберсистем, 19–25 ноября 2018 года;
- XVI Всероссийская школа-конференция молодых ученых УБС 2019: Управление техническими системами и технологическими процессами, 10-13 сентября 2019 года;
- XIV ВСЕРОССИЙСКАЯ НАУЧНАЯ КОНФЕРЕНЦИЯ с международным участием «МЕМБРАНЫ-2019» с 21 по 25 октября 2019 года;
- 32nd International Course and Conference on the Interfaces among Mathematics, Chemistry and Computer Sciences: Mathematics, Chemistry, Computing (7-11 June, 2021).

Также основные положения диссертации докладывались и обсуждались на заседаниях научных семинаров «Теория управления организационными системами» в Институте проблем управления имени В.А. Трапезникова Российской академии наук (ИПУ РАН) и «Применение хроматографии в нефтехимии и аналитике» Института нефтехимического синтеза им. А.В. Топчиева Российской академии наук (ИНХС РАН).

Публикации

По теме диссертации опубликовано 13 работ, среди которых 5 публикаций в рецензируемых научных изданиях из Web of Science/Scopus [14, 43–45, 74], в том числе, одна публикация за единоличным авторством соискателя [74], 8 публикаций в сборниках трудов и тезисов конференций [3–9, 75].

Личный вклад соискателя

Все исследования, изложенные в диссертационной работе, выполнены лично соискателем в процессе научной деятельности. Во всех работах, выполненных в соавторстве, автор внес значительный вклад в разработку представленных методов и алгоритмов, а также в проведение численных экспериментов и создание комплекса программ.

[3], [5], [7], [9], [44], [75] – разработка методов предсказания коэффициента растворимости, создание программного комплекса, проведение численных экспериментов.

[6], [7], [8]– разработка метода кластеризации аморфных полимеров, программная реализация расчетов, проведение численных экспериментов.

[14] – предсказание транспортных характеристик полимеров, сравнение методов предсказания транспортных характеристик, проведение численных экспериментов.

[45] – проведение численных экспериментов.

[4] – разработка методов предсказания константы Генри, проведение численных экспериментов.

Структура и объем работы

Диссертационная работа состоит из введения, четырех глав, заключения, списка литературы и двух приложений. Работа изложена на 159 страницах, содержит 8 таблиц и 31 иллюстрацию. Библиография содержит 114 наименований.

В Главе 1 диссертации приводятся несколько подходов к предсказанию транспортных характеристик полимеров, описывается постановка задачи дизайна материалов с экстремальными транспортными характеристиками. Также описываются методы предсказания транспортных характеристик полимерных материалов, в том числе, методы эмпирических групповых вкладов, методов

QSPR-моделирования, методы молекулярной динамики и механики, а также методы Монте-Карло и машинного обучения. Приводится описание способов представления молекул, а также перечисляются популярные силовые поля для проведения молекулярно-механического и молекулярно динамического моделирования.

В Главе 2 описывается разработанный метод предсказания транспортных характеристик аморфных полимеров на основе площади поверхности коротких полимерных цепей. В разделе 2.1 приводится метод математического моделирования транспортных характеристик полимеров на основе молекулярно-механического моделирования, описываются геометрические индексы на основе площади поверхности короткой полимерной цепи, а также излагается подход к построению регрессионной модели для предсказания транспортных характеристик полимерных мембран на основе представленных геометрических индексов. В разделе 2.2 приводятся разработанные алгоритмы предсказания транспортных характеристик полимерных мембран. Сначала подробно описан алгоритм получения конформаций молекул полимеров, затем алгоритм вычисления геометрических индексов, и после – подход к обучению регрессионных моделей.

В Главе 3 диссертации приводится описание разработанного комплекса программ для предсказания транспортных характеристик полимерных газоразделительных мембран. Описываются составляющие комплекса программ: блок интерфейсов данных, блок построения конформаций молекул, блок вычисления геометрических индексов и параметров молекул, блок регрессионного анализа. Затем приводятся сценарии использования комплекса программ: пользовательский и исследовательский.

В Главе 4 диссертации рассказывается о применении разработанной методики для дизайна материалов с заданными свойствами. Приводится информация о базе данных, в которой собраны более 6000 записей о взаимодействии элементов системы «газ-полимер». Дается обоснование достаточной длины цепочки и количества генерируемых конформаций. Затем подробно описывается предска-

зание одного из наиболее важных транспортных характеристик – коэффициента растворимости S, а также процесс обучения регрессионной модели, способной предсказать значение константы Генри k_D. В конце главы предлагается метод кластеризации полимеров на основе их геометрии, а также исследуются транспортные характеристики выявленных кластеров (групп полимеров) для выявления связей между геометрией полимерных цепочек и транспортными характеристиками веществ.

1 Задачи предсказания транспортных характеристик полимерных материалов

1.1 Транспортные характеристики полимерных материалов

Концепция растворения и диффузии, описывающая явления транспорта газов через непористые мембраны, впервые была сформулирована в середине XIX веке в работах Митчелла [76], Грэма [46] и Вроблевски [110]. Митчелл и Грэм показали, что газы способны проникать через непористые каучуковые пленки и связали это с процессом растворения и диффузии газов в полимерных материалах. Грэм пришел к выводу, что процесс проникновения состоит из двух стадий: сорбция газа каучуком и диффузия молекул сорбированного газа. Вроблевски на основе экспериментов выяснил, что поток через мембрану прямо пропорционален перепаду давления и обратно пропорционален толщине мембраны, то есть:

$$N = P(\Delta p/l). \tag{1.1}$$

Таким образом был введен коэффициент проницаемости P, который определяет перенос газа как внутреннее физическое свойства пары полимер/газ.

В 1855 году Адольф Фик предложил законы, описывающие процесс диффузии газов. В установившемся режиме поток газа в любой точке внутри полимерной пленке, разделяющей две области, заполненные газом под разными давлениями, определяется первым законом Фика:

$$N = \left(\frac{-D_{loc}}{1-w}\right)\left(\frac{dC}{dx}\right), \tag{1.2}$$

где N – поток газа относительно фиксированных координат, C – концентрация газа, x – расстояние поперек пленки, w – массовая доля газа в полимере, D_{loc} – коэффициент бинарной взаимной диффузии газа в полимере. В результате интегрирования данного выражения по x получается выражение

$$N = \frac{C_2 - C_1}{l} D, \tag{1.3}$$

где C_1 и C_2 – концентрации газа на нижней и верхней стороне мембраны соответственно, D – средний эффективный коэффициент диффузии, который определяется как:

$$D = \frac{1}{C_2 - C_1} \int_{C_1}^{C_2} \frac{D_{loc}}{1 - w} dC, \tag{1.4}$$

Таким образом, выражение для проницаемости газа в полимере можно переписать в виде

$$P = \frac{Nl}{p_2 - p_1} = \left(\frac{C_2 - C_1}{p_2 - p_1} \right) D. \tag{1.5}$$

Если учесть, что $C_2 >> C_1$ и $p_2 >> p_1$, то проницаемость можно представить как

$$P = \frac{C_2}{p_2} D, \tag{1.6}$$

Отношение концентрации газа, растворенного в полимере, при равновесии, к давлению газа (или парциальным давлением в случае смесей) в смежной газовой фазе называется коэффициентом растворимости и описывается формулой:

$$S = \frac{C}{p}, \tag{1.7}$$

Таким образом, основное уравнение, связывающее три ключевые транспортных характеристики, представляется уравнением:

$$P = S \cdot D \tag{1.8}$$

Получается, что проницаемость P зависит от коэффициента растворимости S – термодинамической характеристики, показывающей количество молекул газа, сорбированных в полимере и на нем, и коэффициента диффузии D – кинетической характеристики, показывающей подвижность молекул газа в процессе диффузии через полимер. То есть, проницаемость, которая представляет собой нормированный по давлению и толщине поток газа через полимерную пленку, зависит от произведения числа молекул газа, растворенных в полимере, и скорости их миграции через полимерную матрицу.

Другой ключевой характеристикой мембранного газоразделения является селективность. Идеальная селективность описывается формулой

$$\alpha_{AB} = \frac{P_A}{P_B} \tag{1.9}$$

где P_A и P_B коэффициенты проницаемости газов A и B, причем $P_A > P_B$. Селективность α_{AB} может быть представлена в виде:

$$\alpha_{AB} = \left(\frac{D_A}{D_B}\right)\left(\frac{S_A}{S_B}\right) = \alpha^D_{AB}\alpha^S_{AB}, \tag{1.10}$$

причем анализ α^D_{AB} и α^S_{AB}, как правило, очень полезен при разработке полимерных мембран.

Согласно модели двойной сорбции [21], общее количество C газа, растворенного в объеме аморфного стеклообразного полимера в зависимости от давления p, описывается комбинацией закона Генри и сорбции ленгмюровского типа по формуле

$$C = k_D \cdot p + \frac{C'_H \cdot b \cdot p}{1 + b \cdot p} \tag{1.11}$$

где k_D – константа равновесия закона Генри, b – константа равновесия лэнгмюровской сорбции, C'_H – лэнгмюровская сорбционная емкость. Коэффициент

растворимости при бесконечном разбавлении S вычисляется как отношение C/p при $p \to 0$, то есть из изотермы сорбции, S можно представить в виде суммы двух слагаемых:

$$S = k_D + C_H^{'} \cdot b. \tag{1.12}$$

При создании полимерных газоразделительных мембран ученые вынуждены мириться с компромиссом между газопроницаемостью и селективностью, что было показано Робсоном в 1991 году [91] с помощью диаграммы, позднее получившей его имя. Диаграмма Робсона – это график, который строится для некоторого множества полимеров и пары газов, в координатах $\log\alpha_{ij}$ от $\log P_i$. «Верхняя граница» определяется зависимостью $P_i = k\alpha_{ij}^n$, где P_i – проницаемость быстро проникающего компонента, α_{ij} (P_i/P_j) – фактор разделения (идеальная селективность), k является коэффициентом, а n является наклоном указанной зависимости в двойных логарифмических координатах. Ниже этой линии на графике зависимости $\log\alpha_{ij}$ от $\log P_i$ находятся практически все точки экспериментальных данных. Одной из основных целей текущих исследований полимерных мембран является разработка мембран, превосходящих верхнюю границу (рисунок 2) диаграммы Робсона.

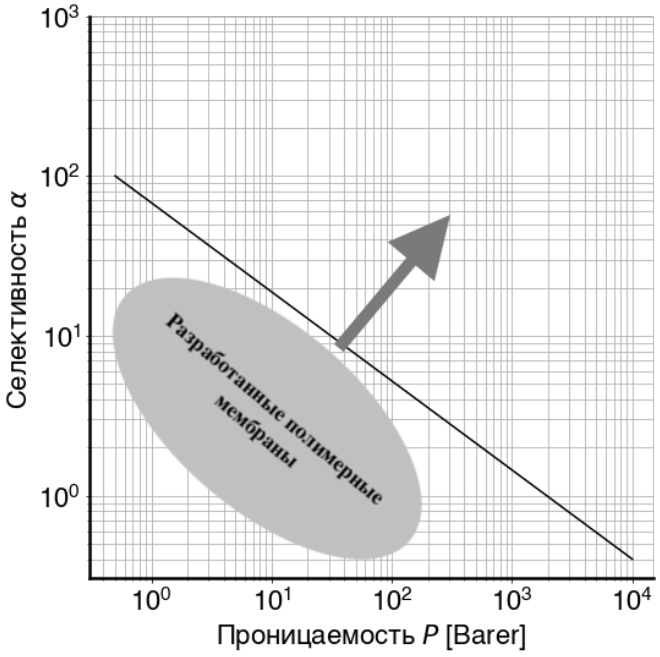


Рисунок 2 – Диаграмма Робсона.

Примеры диаграмм Робсона по различным парам газов представлены на рисунке 3. В 2008 году в работе [92] эмпирическое соотношение верхней границы для мембранного разделения газов, первоначально опубликованное в 1991 г., было пересмотрено с учетом множества новых данных, поэтому диаграммы построены по Базе данных «Газоразделительные параметры стеклообразных полимеров» с добавлением границы Робсона из работы [92]. Значения экспериментальных данных для пары «газ-полимер», взятые из разных источников, были усреднены при одинаковых заявленных условиях проведения экспериментов.

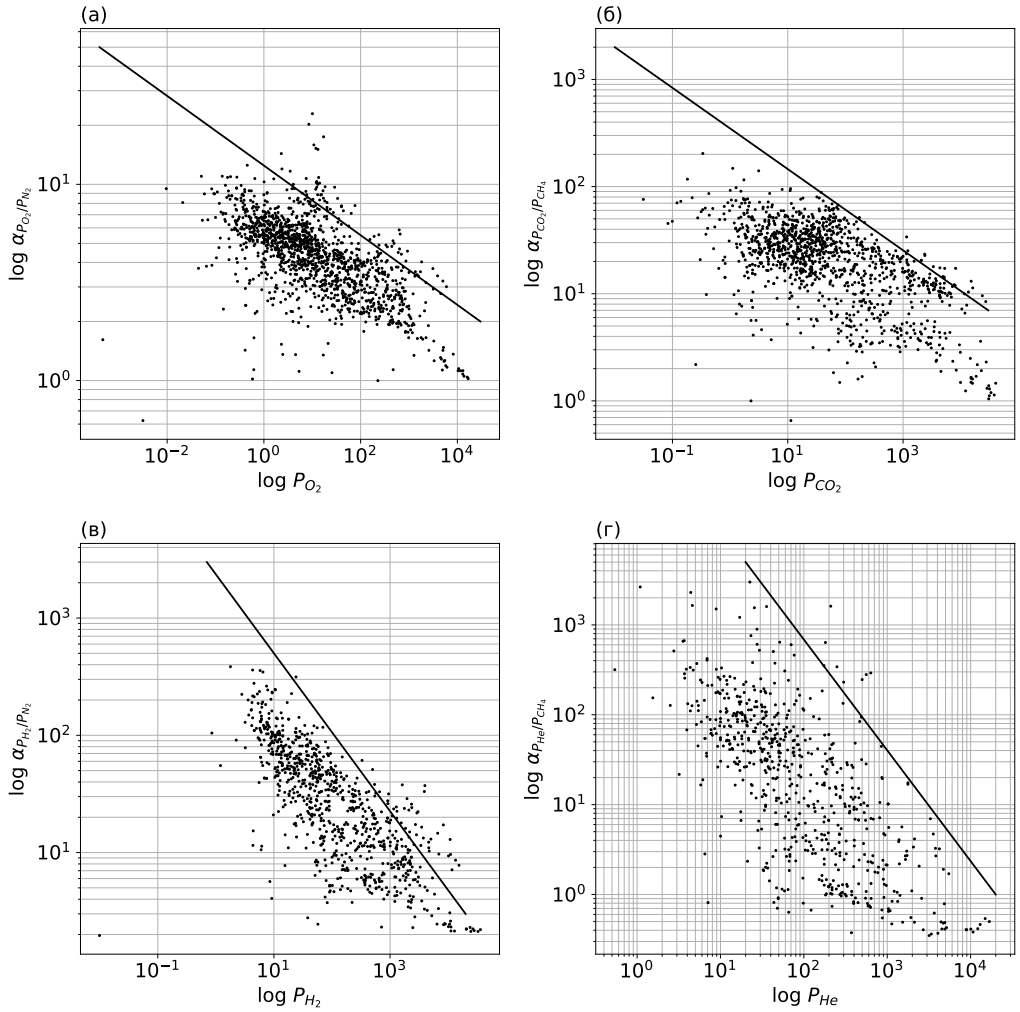


Рисунок 3 – Диаграммы Робсона для проницаемости: (а) O_2/N_2; (б) CO_2/CH_4; (в) H_2/N_2; (д) He/CH_4. Диаграммы построены согласно усредненным экспериментальным данным, полученным из базы данных ИНХС.

1.2 Постановка задачи дизайна материалов с экстремальными транспортными характеристиками

Решение задач управления свойствами полимерных мембран требуется во многих областях промышленности, биомедицины, хранения энергии и множества других отраслей. Важнейшая задача мембранного материаловедения – поиск новых высокоэффективных полимеров для решения разнообразных задач мембранной технологии. Для мембранного газоразделения решающую роль играют транспортные характеристики полимерного материала – проницаемость P_i и селективность P_i/P_j для пары разделяемых газов i и j. Возможность хотя бы приблизительно предсказать транспортные характеристики гипотетического полимера еще до его синтеза существенно сокращает усилия и экономит время, позволяя отбросить заведомо неудачные варианты и сконцентрироваться на наиболее перспективных соединениях.

Традиционно важнейшим фактором, влияющим на транспортные характеристики аморфного полимера, считается его химическое строение. Химическая природа мономерных звеньев и способы их объединения в полимерную цепь определяют взаимное расположение цепей и геометрию свободного объема внутри образца. Именно структура свободного объема полимерной матрицы определяет коэффициент проницаемости P_i молекул газа i через мембрану.

Согласно формуле (1.8), коэффициент проницаемости P можно записать в виде произведения коэффициента растворимости S и коэффициента диффузии D. Коэффициент S описывает движущую силу процесса переноса молекул газа, коэффициент D соответствует кинетической компоненте процесса. Во многих случаях именно коэффициент растворимости S, а также селективность растворимости S_i/S_i по паре газов i и j определяют транспортные характеристики и ценность полимера для мембранной технологии. Это обусловливает актуальность задачи предсказания S для вновь создаваемых полимерных материалов.

Одной из основных задач диссертационной работы является разработка ма-

тематических моделей и алгоритмов предсказания свойств и поиска веществ с заданными физико-химическими свойствами и применение их для выработки рекомендаций по синтезу перспективных соединений. Глобальную задачу, можно разбить на две задачи – прямую и обратную.

- «Прямая» задача – определение физико-химических характеристик полимеров по их химическому строению.
- «Обратная» задача – компьютерный синтез полимеров с заданным комплексом физико-химических свойств.

Обратная задача (задача молекулярного дизайна) – это компонент более сложного процесса открытия новых материалов. Временной масштаб для внедрения новых технологий, от открытия в лаборатории до коммерческого продукта, исторически составляет от 15 до 20 лет [68]. Процесс обычно включает следующие этапы:

- создание новой или улучшенной концепции материала и моделирование ее потенциальной пригодности;
- синтез материала;
- использование материала в устройстве или системе;
- характеризация и измерение желаемых свойств.

Этот цикл включет в себя обратную связь для повторения, улучшения и уточнения будущих циклов открытий. Каждый шаг может занимать до нескольких лет. Для решения обратной задачи необходимо научиться решать прямую задачу.

Рассмотрим множество Ω допустимых молекул (например, представленных их структурными формулами или химическими графами), каждая из которых наделена $k+1$ релевантными физическими или химическими свойствами (например, нормальной плотностью, коэффициентом преломления, индексом удерживания, коэффициентом растворимости и т. д.), и пусть $P_i(G), i = 0, ..., k$, чис-

ловое значение i-го свойства молекулы $G \in \Omega$ (например, коэффициент растворимости). Типичной проблемой молекулярного дизайна является следующая задача оптимизации:

$$P_0(G) \to \min_{G \in \Omega} \left(\max_{G \in \Omega} \right) \tag{1.13}$$

при условиях

$$P_i^{\min} \leqslant P_i(G) \leqslant P_i^{\max}, i = 1, \ldots, k. \tag{1.14}$$

Когда функции $P_i(\cdot)$ известны из эксперимента лишь частично, они заменяются предсказанными величинами, связывающими структуру молекулы $G \in \Omega$ с предсказанным значением $\tilde{P}_i(G)$ i-го физического или химического свойства $(i = 0, \ldots, k)$ посредством числовых характеристик (известных как молекулярные дескрипторы), которые могут быть рассчитаны на основе молекулярной структуры. Типичное количественное соотношение структура-свойство (QSPR) включает несколько молекулярных индексов и может быть представлено в виде

$$\tilde{P}_i(G) = \tilde{P}_i \left(I_1(G), \ldots, I_m(G) \right), i = 0, \ldots, k,$$

где $I_1(G), \ldots, I_m(G)$ – значения молекулярных индексов для молекулярной структуры G. Оно зачастую может быть представлено в виде линейной регрессии, по сути, простой взвешенной суммы индексов:

$$\tilde{P}_i(G) = \alpha_{1,i} I_1(G) + \ldots + \alpha_{m,i} I_m(G), i = 0, \ldots, k.$$

Однако модели могут иметь более сложный вид нелинейных регрессий или даже нейронных сетей, если природа и механизмы предсказываемой величины требует более сложного описания.

За последние десятилетия был предложен и изучен ряд топологических, геометрических и квантово-механических молекулярных индексов [13,14,24,25].

Исчерпывающий перебор всех возможных молекул (подход грубой силы) может быть использован для решения этой проблемы только тогда, когда допустимый набор относительно невелик; для больших наборов математическая химия предлагает множество ограниченных методов поиска.

Как мы видим, решение прямой задачи является одним из этапов решения обратной задачи. Регрессии, связывающие между собой физико-химические свойства и характеристики молекулярных структур исследуемого вещества, входят в целевую функцию ограничений оптимизационной задачи.

Ниже представлены методы предсказания транспортных характеристик полимерных материалов.

1.3 Методы предсказания транспортных характеристик полимерных материалов

Для решения задачи дизайна материалов требуется количественное понимание взаимосвязей между структурой, свойствами и эксплуатационными характеристиками материалов. Развитие именно этого понимания и составляет основную цель моделирования материалов. Моделирование стало незаменимым инструментом в разработке новых материалов наряду с экспериментальными методами. Молекулярное моделирование – довольно молодая область науки и насчитывает менее 50 лет своей истории. Сегодня, благодаря современным вычислительным мощностям и возможностям, моделирование позволяет не только получить информацию о механизмах, влияющих на свойства материалов, но и количественно предсказывать свойства сложных систем материалов.

Полимерные кристаллиты в подавляющем большинстве случаев практически непроницаемы для пенетрантов с малой молекулярной массой, поэтому усилия по моделированию сосредоточены в основном на свойствах проницаемости аморфных полимеров как выше, так и ниже температуры стеклования, T_g. Надежное предсказание сорбционного равновесия и скоростей диффузии зависит

от наличия хорошей компьютерной модели полимерной матрицы. Вычислительная генерация атомистических конфигураций, которые точно представляют реальную полимерную матрицу, в целом является сложной задачей, и от качества решения этой задачи во многом зависит качество предсказательной модели. За генерацию атомистических конфигураций, в том числе, отвечают силовые поля, которых мы коснемся в следующем разделе.

1.3.1 Методы цифрового представления молекулярных структур

Химическое вещество представляет собой совокупность атомов, связанных вместе в пространстве. Структура химического вещества делает его уникальным и придает ему физические и биологические характеристики. Поэтому, прежде, чем переходить к методам предсказания транспортных характеристик полимерных материалов, необходимо подробнее ознакомиться с некоторыми методами, используемыми для компьютерного представления молекулярных структур.

Химические структуры обычно представляются в виде молекулярных графов. Граф – это абстрактная структура, содержащая узлы, соединенные ребрами. В молекулярном графе узлы соответствуют атомам, а ребра – связям. Атомы водорода часто опускаются, однако, часто при моделировании методами молекулярной механики и динамики, необходимо учитывать вклады атомов водорода. Узлы и ребра могут содержать связанные с ними свойства. Например, атомный номер или тип атома может быть связан с каждой вершиной, а порядок связи – с каждым ребром. Эти свойства атома и связи важны при выполнении операций с молекулярным графом. Граф представляет только топологию молекулы, то есть способ соединения вершин (или атомов). Подграф – это подмножество вершин и ребер графа; таким образом, график бензола является подграфом графа аспирина. Дерево – это особый тип графа, в котором есть только один путь, соединяющий каждую пару вершин, то есть внутри графа нет циклов или колец. Вершины являются либо узлами ветвления, либо листо-

выми узлами. Структура ациклических молекул представляется деревьями. В полном графе существует ребро между всеми парами вершин. Такие графы довольно редки в химии, однако могут быть полезны при работе с трехмерными структурами.

Как и почти во всех задачах хемоинформатики, молекулярное представление является ключевым аспектом, который следует учитывать при перечислении химических соединений. Вероятно, наиболее известным описанием соединений является двумерное (2D) графическое представление. В настоящее время существует множество программ, помогающих рисовать химические структуры и облегчающих хранение и преобразование между стандартными форматами файлов. Некоторые из этих программ имеют бесплатные академические версии, такие как MarvinSketch [69] и ACD/ChemSketch [102], а другие являются коммерческими, например ChemDraw [27], Schredinger [98] и MOE [28], и это лишь некоторые из них.

Трехмерные (3D) структуры также широко используются, особенно сейчас, когда были разработаны многочисленные компьютерные программы для их расчета и визуализации. Эти представления обеспечивают мощный и интуитивно понятный инструмент для понимания многих аспектов химии. Однако они имеют ограничения, особенно когда речь идет о повседневных задачах хемоинформатики, требующих хранения и работы с огромным количеством соединений. В этих приложениях молекулярная информация обычно представлена линейной записью. Ниже описаны некоторые из наиболее часто используемых линейных нотаций для перечисления химических структур: SMILES, InChi и InChikeys.

Линейные нотации представляют собой короткие и удобочитаемые описания молекулярных графов. Ярким примером является широко используемая упрощенная система ввода молекулярных данных The Simplified Molecular-Input Line-Entry System (SMILES), которая фиксирует структуру молекул в форме однозначной текстовой строки с использованием буквенно-цифровых символов.

Они позволяют эффективно хранить и быстро обрабатывать большое количество молекул. В нотации SMILES используются следующие основные правила кодирования молекул [108]:

1. Атомы кодируются их атомными символами. Атомы водорода опускаются.
2. Соседние атомы располагаются рядом друг с другом, а связи характеризуются как одинарные (-), двойные (=), тройные (#) или ароматические (:). Одинарные и ароматические связи обычно не учитываются.
3. Вложения в круглых скобках указывают ответвления в молекулярной структуре.
4. Для линейного представления циклических структур связь разрывается в каждом кольце, а за соединительными атомами кольца следует одна и та же цифра в текстовом представлении.
5. Атомы в ароматических кольцах обозначаются строчными буквами. В некоторых случаях могут быть проблемы с восприятием ароматичности.

Несмотря на то, что строки SMILES однозначно описывают химические структуры, они не уникальны, поскольку для одного и того же молекулярного графа существует несколько действительных представлений SMILES. Этот факт является одним из основных минусов нотации SMILES.

InChI – это международный химический идентификатор, разработанный под эгидой IUPAC, Международного союза теоретической и прикладной химии при основном вкладе NIST (Национальный институт стандартов и технологий США) и InChI Trust [58]. Цель InChI состоит в том, чтобы установить уникальную маркировку для каждого соединения и облегчить связывание различных компиляций данных. Это обозначение разрешает многие химические неясности, не затронутые SMILES, особенно в отношении стереоцентров, таутомеров и других проблем модели валентности. Однако в большинстве случаев людям трудно читать и интерпретировать InChI. InChI включают в себя различные

слои и подслои информации, разделенные косой чертой (/). Каждая строка InChI начинается с номера версии InChI, за которым следует основной слой. Этот основной слой содержит подслои для эмпирической формулы, соединений атомов и положений атомов водорода. Идентичность каждого атома и его ковалентно связанных партнеров предоставляет всю информацию, необходимую для основного слоя. За основным слоем могут следовать дополнительные слои, например, для заряда, изотопного состава, таутомерии и стереохимии.

InChIKey представляет собой сжатое цифровое представление InChI фиксированной длины (27 символов), разработанное для упрощения поиска химических структур в Интернете. Первый блок из 14 символов для InChIKey кодирует молекулярное строение ядра, как описано формулой, связностью, водородными позициями и зарядовыми подслоями основного слоя InChI. Другие структурные особенности, дополняющие основные данные, а именно точное положение подвижных атомов водорода, стереохимических, изотопных и металлических лигандов, в зависимости от того, что применимо, кодируются вторым блоком InChIKey.

Строковые нотации типа SMILES и InChi описывают типы атомов, их последовательность и связи между ними, но не включают 2D- или 3D-координаты, которые необходимы для моделирования. Для хранения молекулярного графа в памяти компьютера чаще всего используются различные виды таблицы связей (connection table). Простейший тип таблицы связей состоит из двух разделов: списка атомных номеров атомов в молекуле и списка связей, заданных как пары связанных атомов. Более подробные формы таблицы соединений включают дополнительную информацию, такую как состояние гибридизации каждого атома и порядок связи. Также в таблицах содержится информация о координатах (xy) или (xyz) атомов. Многие разработчики программных обеспечений используют свои форматы хранения молекулярных структур, однако, наиболее распространенными являются CML (язык химической разметки), SDF (формат структурных данных), PDB (банк данных белков), и формат файла XYZ

в котором хранятся трехмерные координаты. Используя различные форматы химических файлов, молекулы могут быть представлены и сохранены с соответствующими расширениями, такими как «.pdb», «.sdf», «.xyz» и т. д. Ниже подробнее остановимся на SDF и PDB форматах, так как именно между этими форматами встал вопрос выбора при проведении исследований, описываемых в данной работе.

Формат структурных данных (SDF) [30], также известный как SD-файлы, представляет собой распространенный химический формат файлов, используемый для записи нескольких химических структуры и связанных полей данных. SDF был разработан и опубликован компанией Molecular Design Limited (MDL) и стал наиболее широко используемым стандартом для импорта и экспорта информации о химических веществах. Файлы формата SDF фактически представляют собой формат molfile (MDL Molfile), в котором несколько записей разделяются строками, состоящими из четырех знаков доллара ($$$$). Общий формат файла SDF состоит из блоков информации с одним форматом составной записи. Основные блоки это: блок информации об атомах, блок информации о связях, блок свойств и блок дополнительной информации. Пример файла с поли[о-(триметилгермил)фенил]ацетилен представлен на рисунке 4.

Формат файла Protein Data Bank (PDB) [22] представляет собой текстовый формат файла, описывающий трехмерные структуры молекул. Формат PDB обеспечивает описание и аннотацию белков, структуры нуклеиновых кислот, включая координаты атомов, наблюдаемые ротамеры боковых цепей, назначение вторичной структуры, а также связность атомов. Структуры, на которые нанесены другие молекулы, такие как вода, ионы, нуклеиновые кислоты, лиганды и т. д., также могут быть описаны в формате pdb. Пример файла с поли[о-(триметилгермил)фенил]ацетилен представлен на рисунке 5.

```
  Mrv2001 06022213302D

 12 12  0  0  0  0            999 V2000
    2.1450   -1.6499    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    2.1450   -2.4749    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    1.4299   -1.2336    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    1.4299   -2.8835    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.7150   -1.6499    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.7150   -2.4749    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    1.4299   -0.4086    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.7150    0.0000    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    2.8521   -1.2336    0.0000 Ge  0  0  0  0  0  4  0  0  0  0  0  0
    2.4435   -0.5185    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    3.5670   -0.8250    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    3.2685   -1.9485    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
  1  2  2  0  0  0  0
  1  3  1  0  0  0  0
  1  9  1  0  0  0  0
  2  4  1  0  0  0  0
  3  5  2  0  0  0  0
  3  7  1  0  0  0  0
  4  6  2  0  0  0  0
  5  6  1  0  0  0  0
  7  8  2  0  0  0  0
  9 10  1  0  0  0  0
  9 11  1  0  0  0  0
  9 12  1  0  0  0  0
M  STY  1   1 SRU
M  SCN  1   1 HT
M  SAL   1 12   7   8   3   5   6   4   2   1   9  12  11  10
M  SDI   1  4    1.8433    0.0048    1.8433   -0.6504
M  SDI   1  4    0.3126   -0.6547    0.3126    0.0006
M  SMT   1 n
M  END
>  <Class>
polyacetyenes

>  <ClassID>
14

>  <SubClassID>
z
```

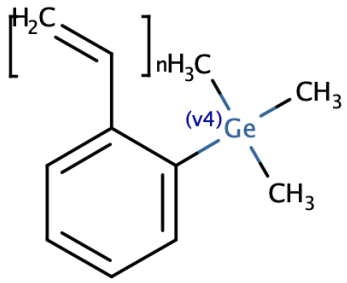


Рисунок 4 – Пример SDF файл полимера поли[о-(триметилгермил)фенил]ацетилен с добавленным 2D изображением.

```
HEADER    PROTEIN                                 02-JUN-22   NONE
TITLE     NULL
COMPND    NULL
SOURCE    NULL
KEYWDS    NULL
EXPDTA    NULL
AUTHOR    Marvin
REVDAT   1   02-JUN-22         0
HETATM    1  C     n     1       4.004  -3.080   0.000  0.00  0.00           C+0
HETATM    2  C     n     1       4.004  -4.620   0.000  0.00  0.00           C+0
HETATM    3  C     n     1       2.669  -2.303   0.000  0.00  0.00           C+0
HETATM    4  C     n     1       2.669  -5.383   0.000  0.00  0.00           C+0
HETATM    5  C     n     1       1.335  -3.080   0.000  0.00  0.00           C+0
HETATM    6  C     n     1       1.335  -4.620   0.000  0.00  0.00           C+0
HETATM    7  C     n     1       2.669  -0.763   0.000  0.00  0.00           C+0
HETATM    8  C     n     1       1.335   0.000   0.000  0.00  0.00           C+0
HETATM    9 Ge     n     1       5.324  -2.303   0.000  0.00  0.00          Ge+0
HETATM   10  C     n     1       4.561  -0.968   0.000  0.00  0.00           C+0
HETATM   11  C     n     1       6.658  -1.540   0.000  0.00  0.00           C+0
HETATM   12  C     n     1       6.101  -3.637   0.000  0.00  0.00           C+0
MASTER        0    0    0    0    0    0    0    0   12    0    0    0
END
```

Рисунок 5 – Пример PDB файл полимера поли[о-(триметилгермил)фенил]ацетилен.

Минусами PDB файлов является отсутствие возможности хранения нескольких молекулярных структур в одном файле, а также отсутствие полей для для записи дополнительной пользовательской информации. Поэтому для хранения базы данных, а также отдельных полимеров в данной работе был выбран формат SDF.

1.3.2 Методы QSAR/QSPR моделирования

Количественные отношения структура-активность/свойство (QSAR/QSPR) являются конечным результатом процесса, который начинается с подходящего описания молекулярных структур и заканчивается некоторыми выводами, гипотезами и предсказаниями поведения молекул в окружающей среде, биологических и физико-химических системах в условиях окружающей среды. Так как в данной работе в качестве предсказываемых величин рассматриваются именно транспортные характеристики газоразделительных полимерных мембран, то далее мы будем рассматривать именно QSPR.

Модели и методы QSPR основаны на предположении, что структура молекулы (например, ее геометрические, стерические и электронные свойства) должна содержать особенности, ответственные за ее физические, химические и биологические свойства, и на способности улавливать эти особенности в одной или нескольких формах. С помощью моделей QSAR/QSPR свойство нового разработанного или неиспытанного химического вещества может быть выведено из молекулярной структуры аналогичных соединений, свойство которых уже было оценено или получено экспериментально.

Как отмечается в [106], разработка моделей QSPR – достаточно сложный процесс. Вначале определяется цель исследования и вырабатывается понимание, насколько общей должна быть модель и какой предсказательной силой обладать. Это влечет за собой выбор набора молекул, к которым применяется процедура моделирования. Окончательное решение по определению набора молекул в основном зависит от предполагаемого использования модели и наличия экспериментальных данных. Причем часто, именно набор экспериментальных данных определяет необходимую сложность модели и инструменты, которые будут использованы для моделирования. Помимо объема экспериментальных данных, итоговая модель будет зависеть еще и от качества и точности этих данных. Это, очевидно, подразумевает получение надежных экспериментальных данных. Данные о химических веществах могут быть получены экспериментально или взяты из литературы. В обоих случаях следует тщательно оценивать точность: ограничивающим фактором при разработке моделей QSPR является наличие качественных экспериментальных данных, поскольку точность оцениваемого моделью свойства не может превышать степень точности входных данных. Более того, когда данные собираются из литературы, чтобы избежать дополнительной изменчивости данных из-за разных источников информации, данные следует брать только из одного источника или из сопоставимых источников.

Следующей, после определения количества и источника данных, фазой про-

цесса QSPR является определение надежного химического пространства или, другими словами, выбор тех структурных особенностей, которые считаются наиболее ответственными за моделирование целевого свойства. Это подразумевает выбор подходящих молекулярных индексов, но в большинстве случаев нет априорных сведений о том, какие молекулярные дескрипторы являются лучшими. Тенденция заключается в использовании огромного количества индексов, которые, как мы надеемся, включают в себя переменные-кандидаты для моделирования, а затем применяется метод выбора переменных. Могут быть приняты две основные стратегии: использование алгоритмов для выбора оптимального подмножества (подмножеств) индексов и использование методов (например, PCA или PLS), способных сжать большой объем доступной химической информации в несколько основных переменных.

Большинство стратегий QSPR, направленных на построение моделей, основаны на методах регрессии, классификации и использовании нейронных сетей в зависимости от изучаемой проблемы. Также например, для оценки распределения соединений в химическом пространстве в основном используются кластерные методы.

Существует множество различных молекулярных индексов, пожалуй большая часть из них описана в фундаментальной работе [106]. Определение индексов, согласно Роберто Тодескини и Вивиане Консонни, следующее: «Молекулярный дескриптор – это конечный результат логической и математической процедуры, которая преобразует химическую информацию, закодированную в символическом представлении молекулы, в полезное число или результат какого-то стандартизированного эксперимента».

Молекулярные дескрипторы группируются в соответствии с их размерностью как 0D, 1D, 2D, 3D и 4D [106]. 0D-дескрипторы не зависят от молекулярной связности и конформаций и относятся к количеству атомов и типов связей. 1D-дескрипторы содержат информацию о количестве фрагментов, и их расчет не зависит от информации о структуре молекулы. 2D-дескрипторы, на-

зываемые инвариантами графов или топологическими дескрипторами, получаются из молекулярных графов и конформационно независимы. Напротив, 3D-дескрипторы зависят от геометрических координат атомов молекул (квантовая химия, молекулярная поверхность, объемы и т. д.). 4D-дескрипторы представляют собой энергетические дескрипторы, полученные путем вычисления энергии взаимодействия между соединением и молекулярными зондами.

1.3.3 Метод групповых вкладов

Подход QSPR к предсказанию транспортных характеристик полимерных материалов, основанный на использовании топологических индексов и метода эмпирических групповых вкладов, был предложен в [96, 113]. Методы групповых вкладов широко используются для прогнозирования свойств органических соединений, лекарственных веществ. В методе эмпирических групповых вкладов каждому уникальному подграфу молекулярного графа («группе») присваивается такой статистический вес, чтобы взвешенная сумма числа вхождений в полимерное звено всевозможных молекулярных групп наилучшим образом коррелировала со значением интересующего физико-химического показателя на некоторой выборке экспериментальных данных. Суммирование вкладов с учетом количества каждого из фрагментов в структурной формуле мономерного звена позволяет сделать численную оценку требуемого физико-химического свойства. Способ разбиения структуры полимера на фрагменты является ключевым моментом, от которого в значительной степени зависит точность прогнозирования физико-химических свойств.

Также существуют различные модификации методов групповых вкладов. Например, в методе модифицированных атомных вкладов MAC (modified method of atomic contributions) [113] и методе вклада связей BC (Bond contribution method) [96] все структуры представляются в виде молекулярных графов. Затем, структурная формула повторяющегося звена разбивается на набор фрагментов (атомов или связей с ближайшими соседями), и каждому

фрагменту присваивается приращение (переменная, представляющая количественный вклад фрагмента) рассматриваемого физического свойства. Сумма этих приращений, умноженная на нормировочный коэффициент (введенный для учета разницы в молекулярном весе или длине полимерной цепи), представляет собой прогнозируемое значение рассчитываемого физического свойства. Система таких уравнений для набора выбранных полимеров затем решается для получения ряда значений приращения и ряда прогнозируемых значений свойств. Полученные приращения затем используются для расчета целевого свойства полимера. Отличие метода MAC от BC заключается в способе разбиения повторяющихся единиц на фрагменты. В обоих случаях повторяющееся звено полимера расщепляется на атомы (MAC) или связи (BC) основной цепи и боковых групп; однако метод BC также включает «смешанные» связи, соединяющие основную цепь с боковыми группами.

Из недостатков методов групповых вкладов можно выделить два основных. Во-первых, относительно слабая предсказательная сила (способность к обобщению за пределы обучающей выборки), что является следствием сложности и нелинейности связи между топологией полимерного звена и структурой свободного объема полимерного образца. Во-вторых, использование метода главных компонент для сокращения числа объясняющих переменных, не дает возможности объяснить природу полученных зависимостей. В то же время, MAC – это единственный из известных из литературы методов, который напрямую может быть сравнен с подходом, предлагаемым в настоящей диссертационной работе. В частности, подробно описываемый в Главе 2 метод ППКПЦ ниже в разделе 4.4 сравнивается с методом MAC в задаче поиска высокопроницаемых полимеров.

1.3.4 Эмпирические силовые поля

Потенциальное эмпирическое силовое поле обеспечивает математическое описание потенциальной энергии $V(r_P)$ как функции атомной конфигурации

r_P полимера. Здесь r_P можно представить как набор декартовых координат всех атомов или, в более общем смысле, центров взаимодействия, составляющих полимер. Хорошее силовое поле должно воспроизводить структурные, термодинамические и динамические свойства полимера в условиях термодинамического равновесия. Как правило, силовые поля для выбранной функции энергии требуют учета экспериментальных данных из областей физики и химии, или даже квантово-механических расчетов. Силовые поля представляют собой межатомные потенциалы и используют ту же концепцию, что и силовые поля в классической физике, с той разницей, что параметры силового поля в химии описывают энергетический ландшафт, из которого силы, действующие на каждую частицу, выводятся как градиент потенциальной энергии.

Различают различные классы силовых полей. Одним из классических и широко применимых является UFF (Universal Force Field [87]) (универсальное силовое поле) – общее силовое поле с параметрами для большей части элементов таблицы Менделеева вплоть до актиноидов включительно, разработанное в Университете штата Колорадо. UFF относится к классическим силовым полям и подходит не для всех молекул. Это силовое поле хорошо работает с неорганическими материалами и металлоорганическими материалами, однако с органическими соединениями его использовать не рекомендуют.

Силовое поле Дрейдинга, разработанное в 1990 году [70], используется для предсказания структуры и динамики органических, биологических и неорганических молекул основной группы. Идея заключается в использовании общих силовых констант и геометрических параметров, основанных на простых соображениях гибридизации, а не отдельных силовых констант и геометрических параметров, которые зависят от конкретной комбинации атомов, участвующих в связи, угле или скручивании. Таким образом, все расстояния связей выводятся из атомных радиусов, и существует только одна силовая постоянная для связей, углов и инверсий и только шесть различных значений для торсионных барьеров. Параметры определены для всех возможных комбинаций атомов.

Силовые поля MMFF94 и MMFF94(s) (Merck Molecular Force Field) [49–55] представляет собой семейство химических силовых полей, разработанных исследовательскими лабораториями Merck. Они основаны на силовом поле MM3. MMFF не оптимизирован для моделирования белков или малых молекул, но неплохо работает для широкого спектра расчетов органической химии.

Как правило, силовые поля специализируются на различных классах молекул. Например, из классических силовых полей, AMBER (Assisted Model Building and Energy Refinement) и GROMOS (GROningen MOlecular Simulation) хорошо работают для белков и ДНК, а IFF (Interface Force Field) – первое силовое поле, работающее с металлами.

Помимо классических силовых полей существуют поля, полученные методами машинного обучения, в частности SchNet – нейронная сеть, использующая сверточные слои с непрерывной фильтрацией для прогнозирования химических свойств и поверхностей потенциальной энергии, или PhysNet – это функция энергии на основе нейронной сети для прогнозирования энергий, сил и (флуктуирующих) частичных зарядов.

Силовые поля активно используются в методах молекулярной механике и молекулярной динамике при моделировании методом Монте-Карло. В литературе известно несколько подходов к предсказанию транспортных характеристик полимеров. Наиболее хорошо разработанные и популярные основаны на атомистическом компьютерном моделировании (см. обзоры в [38, 39]). В частности, для предсказания коэффициентов проницаемости, диффузии и растворимости газов в полимерах использовались методы молекулярной динамики и большого канонического ансамбля [59, 81]. В диссертационной работе используется силовое поле MMFF94, также были опробованы силовые поля Дрейдинга и UFF.

1.3.5 Методы молекулярной механики

В отличие от малых молекул, структура которых может быть минимизирована с помощью методов квантовой механики (с использованием таких программ, как Spartan), большие молекулярные структуры (такие как ДНК, РНК, белки или другие органические полимеры) должны быть минимизированы с помощью молекулярной механики, основанной на законах Ньютона. Атомы рассматриваются как объекты, обладающие массой, а связи – как пружины с соответствующими силовыми константами. Силовое поле, содержащее все соответствующие параметры для данного атома и типы связи, используется для решения уравнений энергии, которые суммируют все энергии по всем атомам и связям в молекуле. Эти энергии включают взаимодействия между связанными атомами (растяжение, изгиб, кручение, виляние) и между несвязанными атомами (электростатические и ван-дер-ваальсовы). Для расчетов минимизации положения атомов внутри молекулы должны быть систематически или случайным образом перемещены, а энергия пересчитана с целью найти более низкую энергию и, следовательно, более стабильную молекулу. Вычисления минимизации не могут исследовать все конформационное пространство и не могут легко переместить структуру от локального минимума к глобальному минимуму, если два из них разделены большим энергетическим барьером. Минимизация энергии обычно осуществляется в отсутствие растворителя.

Распространенными силовыми полями, используемыми для больших молекул, являются CHARMM, AMBER и GROMOS. Параметры для конкретного типа атома в данной связи включают атомную массу, радиус Ван-дер-Ваальса, частичный заряд для атомов (из квантовой механики) и длину связи (из данных электронной дифракции), углы и силовые константы для связей (смоделированы как пружины). Эти параметры получены из экспериментов и теоретических (обычно квантово-механических) расчетов малых органических молекул. Затем решается уравнение потенциальной энергии, состоящее из условий растяжения

связи, изгиба угла и изменения угла кручения (связанные взаимодействия), а также электростатических и ван-дер-ваальсовых взаимодействий (несвязанных).

Преимуществом подходов минимизации энергии или «молекулярной механики» (ММ) является их скорость. Недостатком является то, что они не обеспечивают строгую выборку статистического механического ансамбля напрямую. Альтернативные методы, начиная со случайно расположенных невозмущенных цепей и затем вводя исключенные объемные взаимодействия, могут привести к искажению конформаций цепей на малых масштабах [19]. Тем не менее, конфигурации ММ представляют собой хорошие начальные предположения для более точных методов моделирования, учитывающих температурные флуктуации образца, и могут порождать полностью равновесные модели полимеров. В диссертационной работе методы ММ используются при генерации конформаций молекул полимеров (см. раздел 2.1.1).

1.3.6 Методы молекулярной динамики

В отличие от методов ММ, позволяющих получить статическую картину трехмерной конформации молекулы, методы молекулярной динамики используются для того чтобы описать термодинамическое движение молекулы и ее взаимодействие с другими молекулярными структурами. В простейшей форме метода молекулярной динамики [12] атомы перемещаются в соответствии с уравнениями движения Ньютона в эмпирическом силовом поле, моделирующим взаимодействия между молекулами газа и атомами полимерной цепи. Жесткие молекулы требуют использования уравнений Эйлера, возможно, выраженных в терминах кватернионов Гамильтона. Молекулы с внутренними степенями свободы, но также подверженные структурным ограничениям, могут требовать использования метода метода Лагранжа для включения геометрических ограничений в динамические уравнения. Нормальная равновесная молекулярная динамика соответствует микроканоническому ансамблю статистической меха-

ники, но в некоторых случаях требуются свойства при постоянной температуре (а иногда и давлении); существуют способы модификации уравнений движения для создания таких систем, но, конечно, отдельные траектории больше не представляют собой решение уравнений Ньютона. Уравнения движения можно решить только численно. Из-за характера межатомного взаимодействия атомные траектории неустойчивы, так как бесконечно малое возмущение будет нарастать с экспоненциальной скоростью. Поэтому нет смысла гнаться за точностью траекторий даже в пределах ограниченных промежутков времени. Часто бывает достаточно метода численного интегрирования сравнительно низкого порядка.

Молекулярная динамика местами противоречит теории относительности и квантовой механике. Например, специальная теория относительности запрещает передачу информации со скоростью, превышающей скорость света, а моделирование в методе молекулярной динамики предполагает использование сил, природа которых предполагает бесконечную скорость распространения. В отличие от квантовой механики, в основе которой лежит принцип неопределенности, молекулярная динамика требует и предоставляет полную информацию о положении и импульсе в любое время. На практике явления, изучаемые моделированием методом молекулярной динамики – это явления, в которых релятивистские эффекты не наблюдаются, а квантовые эффекты могут быть, при необходимости, учтены в качестве квазиклассических поправок. И, несмотря на вышеописанные несостыковки, молекулярная динамика остается крайне эффективным методом моделирования. Среди программных пакетов, эффективно реализующих параллельные стратегии молекулярной динамики, можно назвать, например, LAMMPS [105].

Моделирование полимерных материалов по-прежнему сталкивается с двумя серьезными проблемами, обе из которых являются объектами интенсивных исследований. Первая проблема заключается в том, что силовые поля, доступные для атомистического моделирования реальных полимеров, все еще имеют огра-

ниченную точность. Это в особенности верно для полимеров, содержащих сильно полярные, ассоциированные или негибкие фрагменты. Вторая и, возможно, более серьезная проблема заключается в том, что свойства реальных полимерных материалов определяются очень широким спектром масштабов длины и времени – от 0,1 нм до мм и от 10 фс до лет. Они на много порядков превышают самые длинные масштабы времени и длины, которые можно смоделировать с помощью обычных алгоритмов на доступных в настоящее время компьютерах. Моделирование атомистической молекулярной динамики может отслеживать временную эволюцию модельных систем размером порядка 10 нм в течение времени порядка 10–100 нс. Эта задача может быть решена за счет разработки многомасштабных или иерархических стратегий моделирования и симуляции, основанных на систематическом приближении молекулярного представления. Такая стратегия, как правило, состоит из нескольких взаимосвязанных уровней, при этом каждый уровень обращается к явлениям в определенном окне продолжительности и временных масштабов, получая входные данные с более мелких уровней и предоставляя входные данные для более грубых [104]. Методы молекулярной динамики используются в диссертационной работе В диссертационной работе методы молекулярной динамики использовались в первых версиях метода генерации конформаций молекул полимеров (см. раздел 3.1.2), где моделировалось тепловое движение атомов для решения проблемы нереалистичности конформационных структур молекул.

1.3.7 Методы Монте-Карло

Целью моделирования методом Монте-Карло (МК) является создание ансамбля репрезентативных конфигураций в конкретных термодинамических условиях для сложной макромолекулярной системы [35]. Эти конфигурации генерируются путем применения случайных возмущений к системе. Для правильной выборки репрезентативного пространства возмущения должны быть достаточно большими, энергетически возможными и высоковероятными. Мо-

делирование методом Монте-Карло не дает информации об эволюции во времени. Скорее, оно обеспечивает набор репрезентативных конфигураций и, следовательно, конформаций, из которых могут быть рассчитаны вероятности и соответствующие термодинамические наблюдаемые величины, такие как свободная энергия. Моделирование методом Монте-Карло важно не только само по себе, но и играет фундаментальную роль при разработке сложных и гибридных молекулярно-динамических алгоритмов.

Термодинамические свойства рассчитываются как средние по всем выбранным конфигурациям. В методе Монте-Карло каждая следующая конформация определяется при помощи случайных процессов, а не путем решения уравнений Ньютона, как в случае использования методов молекулярной динамики. Вместо того, чтобы оценивать силы, определяющих возрастающие атомные движения, в методе Монте-Карло моделируют относительно большие движения системы и определяют, действительно ли измененная структура энергически возможна при моделируемой температуре. Поскольку метод Монте-Карло сканирует конформационное пространство молекулы без построения настоящей временной траектории, он не может давать численной информации о численных временных зависимостях. Однако, метод намного лучше метода молекулярной динамики для расчета термодинамических характеристик молекул, например, для расчета спектра возможных конформаций и их энергий.

1.3.8 Методы большого канонического ансамбля

В отличие от метода молекулярной динамики, широко используемого при моделировании диффузии, методы большого канонического ансамбля показали свою эффективность при моделировании растворимости. В тесте Видома [38] коэффициент растворимости вычисляется на основе потенциальной энергии молекулы газа, помещенной в случайное место полимерной фазы.

Как метод молекулярной динамики, так и метод большого канонического ансамбля требуют в качестве входных данных компьютерную пространственную

модель полимерной матрицы, которая обычно получается с помощью методов молекулярной механики. Последние сводятся к минимизации потенциальной энергии системы выбором положений атомов одной или несколько полимерных цепочек, помещенных в куб размера порядка нескольких нанометров с периодическими граничными условиями (для ускорения релаксации положений атомов здесь также используется метод большого канонического ансамбля).

Также стоит отметить, что в большинстве работ (за исключением, пожалуй, статьи [71]), предполагается заранее известной плотность стеклообразной полимерной фазы, что делает проблематичным применение такого моделирования в рамках молекулярного дизайна, когда требуется моделировать гипотетические, ранее не синтезированные или не изученные полимеры.

Кроме того, детальное моделирование таких сложных объектов, как полимерная матрица, требует весьма существенных даже по современным меркам вычислительных мощностей, применения суперкомпьютеров, что ограничивает возможность массового применения метода молекулярной механики для быстрого моделирования десятков и сотен перспективных полимеров (в том числе не синтезированных ранее) с целью отбора материалов с экстремальными значениями заданных транспортных характеристик.

1.3.9 Теория переходного состояния Гусева-Сутера для полимерных матриц, испытывающих изотропное движение

Прямое применение метода молекулярной динамики к предсказанию, например диффузии, невозможно, поскольку требует моделирования на колоссальном временном отрезке. *Теория переходного состояния* [48] позволила преодолеть эти трудности, что обусловило популярность данного подхода в моделировании процессов диффузии. В теории переходного состояния и ее многочисленных вариантах [47, 80] используется иерархия взаимосвязанных моделей процесса диффузии, причем каждая модель действует на своем временном и пространственном уровне [38], от поправочных членов к эмпирическому сило-

вому полю до геометрического анализа границ «ям» потенциальной энергии (т.н., стабильных состояний) и марковской сетевой модели случайных переходов частицы газа между этими стабильными состояниями. В работе [48] Гусев и Сутер предположили, что за время пребывания системы «полимер-пенетрант» в сорбционном «состоянии» атомы полимера совершают гармонические колебания вокруг своих равновесных положений в матрице без пенетранта. Эти движения (маломамплитудные колебания длин связей и валентных углов, либрации торсионных углов) называются «упругими движениями» и отличаются от «структурной релаксации», включающей, например, торсионные переходы в главных цепях или боковых группах. По существу, используется допущение о двойном разделении шкалы времени: характерное время для упругих движений много меньше, чем время между пенетрантными прыжками, а последнее много меньше, чем времена, управляющие матричными релаксационными процессами.

В основе предложенного Гусевым и Сутером метода моделирования лежит концепция, согласно которой динамика малых молекул, растворенных в плотных полимерах, связана с упругим движением полимерных матриц, но может рассматриваться отдельно и независимо от их структурной релаксации. В то время как разделение между динамикой растворенного вещества и структурной релаксацией в плотных полимерах, как ожидается, будет более точным для стеклообразных систем, полученные результаты показали, что это приближение также может быть использовано для полимеров, значительно превышающих их температуру стеклования. Это указывает на то, что движение газа как в стеклообразных, так и в каучукоподобных полимерах может задействовать один и тот же механизм. Также можно учесть тепловое движение каждого атома полимера в отдельности, хотя целесообразное приближение однородных изотропных

тепловых колебаний является наиболее простым и дает удовлетворительные результаты.

1.3.10 Методы машинного обучения

Начиная с 2012 года после победы с большим отрывом нейронной сети AlexNet в конкурсе ImageNet Large Scale Visual Recognition Challenge глубокое обучение произвело революцию в различных задачах компьютерного зрения, обработки естественного языка и многих других. Позже, глубокие генеративные модели стали применять и для генерации и оптимизации молекул. Предполагается, что в будущем подобные системы будут использоваться для создания различных молекул, что значительно сократит ресурсы, затрачиваемые на последующий синтез и характеристику неперспективных структур в лабораториях. Применение нейронных сетей прошло этапы от строкового представления молекул SMILES к более сложным представлениям, таким как грамматика графов и трехмерные представления. Сейчас нейронные сети активно используются для различных задач компьютерной химии и молекулярного дизайна. Одним из наиболее серьезных достижений является создание алгоритма AlphaFold. AlphaFold – это система искусственного интеллекта, разработанная исследовательской группой DeepMind, которая предсказывает трехмерную структуру белка на основе его аминокислотной последовательности. В CASP14 AlphaFold был лучшим методом прогнозирования структуры белка с большим отрывом, обеспечивая прогнозы с высокой точностью. Хотя у системы все еще есть некоторые ограничения, результаты CASP показывают, что у AlphaFold есть отличный потенциал, чтобы помочь нам понять структуру белков и продвинуть биологические исследования. Однако, несмотря на высокую точность предсказания в задачах связанных с биополимерами, подобные решения требуют невероятно больших массивов экспериментальных данных, порядка сотен тысяч и миллионов структур. В мембранном газоразделении объем экспериментальной выборки ограничивается несколькими тысячами структур, что, к

сожалению, не позволяет использовать в полной мере все возможности нейронных сетей.

Несмотря на небольшой объем данных существуют работы, в которых нейронные сети используются для предсказания свойств полимерных мембран. Например, целью работы [57] было разработать количественную зависимость структура-свойство с использованием искусственной нейронной сети для улучшения прогнозирования коэффициентов газопроницаемости мембран для некоторых основных промышленных газов, таких как O_2, N_2, CO_2 и CH_4. Используя банк данных, основанный на 149 полимерах, для всех молекул авторы рассчитывают 21 дескриптор с использованием метода группового вклада Ямпольского, который разлагает структуры полимеров на их наименьшие группы и характеризует повторяющиеся звенья полимера. Дескрипторы подразделяются на две группы: атомы центральной и атомы боковой цепи. На вход нейросети авторы подают рассчитанные молекулярные дескрипторы, выходом же являются коэффициенты газопроницаемости. Для каждого газа строится отдельная нейронная сеть со своей архитектурой от одного до трех скрытых слоев. Нейросети показывают высокие коэффициенты корреляции (R) 0,999, 0,999, 0,984 и 0,999 и низкие среднеквадратические ошибки (RMS) 1,054, 2,635, 150 и 2,46 для N_2, O_2, CO_2 и CH_4 соответственно. К минусам данной работы можно отнести отсутствие физичности и интерпретируемости полученных моделей, а также крайне малый размер обучающей выборки, что может свидетельствовать о слабой обобщающей способности полученных моделей.

1.3.11 Другие методы предсказания транспортных характеристик полимерных материалов

Стоит отметить еще несколько работ, в которых рассматривались родственные проблемы [2, 13, 17, 24, 93, 94]. Монографию Дж. Бичерано [24], в которой декларировалось применение теории графов, но в действительности при вычислении конкретных параметров использовались подгоночные параметры, строго

говоря, можно было бы исключить из рассмотрения. Важный вклад в вычисление многочисленных физико-химических параметров полимеров был внесен А. Аскадским [2, 17]. Однако среди большого числа предсказанных параметров (даже коэффициентов проницаемости для некоторых газов) коэффициенты растворимости газов отсутствуют. Метод, родственный используемому в настоящей работе, был применен И. Роновой с сотрудниками [13, 93, 94]. Он включает обкатку полимерных цепей сферическими зондами с размерами, моделирующими размеры молекул газов-пенетрантов. Но в этих работах рассматривались только корреляции экспериментальных значений P и D с доступным свободным объемом у молекулярных цепей и никаких предсказаний не делалось. Такие корреляции были проанализированы в том числе и в работе [94], где они были построены и для коэффициентов растворимости, найденных как отношения экспериментальных параметров P/D.

1.3.12 Заключение по методам

Описанные выше методы предсказания характеристик полимерных материалов по их молекулярной структуре обладают рядом недостатков, которые не позволяют использовать их при решении обратной задачи (дизайна материалов). Расчетные методы квантовой химии и молекулярной динамики имеют ряд ограничений, связанных с подвижностью многих макромолекул. Методы квантовой химии и механики требуют больших вычислительных мощностей, а также значительного времени вычисления на один полимер. Методы молекулярной динамики, в том числе, реализованные в пакете LAMMPS, требуют ручной генерации входных данных на каждую молекулярную структуру. Другие пакетные решения с реализованными методами молекулярной динамики OpenMM [31], Hoomd-Blue [16, 42] были разработаны скорее для биополимеров и не имеют готовых силовых полей для работы со стеклообразными полимерами из мембранного газоразделения.

В некоторых работах, в том числе с использованием методов большого ка-

нонического ансамбля, предполагается заранее известной плотность стеклообразной полимерной фазы, что делает невозможным применение этих методов для широкого круга, как известных полимеров с еще не вычисленной плотностью, так и еще не синтезированных. Методы групповых и модифицированных атомных вкладов обладают низкой способностью к обобщению за пределы обучающей выборки, несмотря на то, что в пределах изучаемого множества полимерных структур могут показывать хорошую точность предсказания транспортных характеристик. Поэтому подобные методы, каждый раз при введении новых классов полимеров требуют добавления новых индексов, проведения анализа и построения новых моделей.

Итак, основные недостатки перечисленных методов это:

- низкая скорость вычисления;
- слабая предсказательная сила (способность к обобщению за пределы обучающей выборки);
- отсутствие возможности автоматизировать процесс расчетов;
- высокие требования к используемым вычислительным мощностям;
- высокие требования к размеру обучающей выборки, на порядки превышающие имеющиеся на данный момент экспериментальные данные.

В следующей главе решается задача разработки новых методов предсказания транспортных характеристик аморфных полимеров.

2 Метод предсказания транспортных характеристик аморфных полимеров на основе площади поверхности коротких полимерных цепей

В настоящем исследовании развивается новый подход к компьютерному моделированию полимерных материалов и предсказанию транспортных характеристик аморфных полимеров на его основе. Впервые представленный в работе [44], он лишен перечисленных в предыдущем разделе недостатков: не требует экспериментальных данных о полимерном материале; применим к широкому кругу полимеров, в том числе еще не синтезированных; предъявляет весьма скромные требования к вычислительным мощностям.

В целом описываемый ниже метод предсказания транспортных характеристик аморфных полимеров является попыткой соединить в себе достоинства метода эмпирических групповых вкладов и трудоемких, но аккуратных методов молекулярной динамики и большого канонического ансамбля. В основе предлагаемого метода лежит гипотеза о том, что коэффициент растворимости газа в полимере и связанные с этим характеристики в той или иной мере зависят от параметров поверхности контакта между молекулой полимера и молекулой газа.

Подход [44] основан на молекулярно-механическом моделировании относительно короткого отрезка полимерной цепи и использовании геометрических индексов, таких как Ван-дер-Ваальсов объем и площадь доступной поверхности молекулы для построения (на основе некоторого объема экспериментальных данных) многомерной регрессии для предсказания транспортных характеристик полимеров, в частности, коэффициента растворимости при бесконечном разбавлении и константы равновесия закона Генри (1.11) в модели двойной сорбции.

Предложенный метод (ППКПЦ) можно разбить на три основных блока: молекулярно-механическое моделирование, вычисление геометрических индексов и построение регрессий. Идея подобного разделения состоит в том, что все три шага можно делать независимо, как части решения различных задач. Результаты молекулярно-механического моделирования можно использовать для расчета разных наборов индексов, используя которые можно строить различные предсказательные модели на основе различных регрессий, нейронных сетей и других инструментов машинного обучения.

1. **Молекулярно-механическое моделирование**

   Производится молекулярно-механическе моделирование относительно короткого отрезка одной полимерной цепи.

2. **Вычисление геометрических индексов**

   С помощью алгоритма Ли-Ричардса вычисляются поверхностные и поверхностно-зарядные геометрические индексы.

3. **Построение регрессии**

   Для построения предсказательной модели используется множественная линейная регрессия с пошаговым отбором переменных. Агломеративный метод кластеризации используется для кластеризации конформаций полимеров и поиска зависимостей между транспортными характеристиками полимеров различных химических классов.

Первый этап – моделирование короткого отрезка полимерной цепи и использование площади доступной поверхности – идейно наиболее близок подходу [66]. Второй этап – использование статистических инструментов для обучения линейной регрессии – ближе к подходу [96,113]. В то же время, есть и многочисленные отличия, обуславливающие научную новизну подхода настоящей диссертационной работы: использование большого числа (а именно, девяти) различных геометрических индексов, идея о том, что важными объясняющими переменными могут стать параметры зависимости площади доступной поверхности (и

родственных ей индексов) от радиуса «обкатки», уточненный (по сравнению с работой [114]) общий вид зависимости транспортных характеристик от параметров газа-пенетранта и полимера, статистически корректная процедура отбора значимых переменных для линейной регрессии, позволяющая избежать эффекта переобучения.

2.1 Математическое моделирование транспортных характеристик полимеров

Модели геометрии молекул полимера лежат в основе предлагаемой методики прогнозирования транспортных характеристик. В этом разделе подробно описываются три блока, составляющие предложенный метод ППКПЦ. Вначале описывается процедура молекулярно-механического моделирования короткого участка полимерной цепи, затем, описываются предложенные геометрические индексы и, в завершение, предлагается подход к построению регрессионных моделей.

2.1.1 Молекулярно-механическое моделирование

Для определения физико-химических характеристик полимеров по их химическому строению необходимо провести моделирование отдельной полимерной цепи. Подход к моделированию полимерной цепи был впервые представлен в работе [44]. В отличие от коэффициента диффузии D и коэффициента проницаемости P (который зависит от D), значения S меньше зависят от взаимной «упаковки» полимерных цепей и свободного объема, поэтому можно ожидать разумных результатов от учета лишь геометрии одиночной полимерной цепи. Более того, ожидалось, что коэффициент растворимости будет зависеть в основном от локальной конформации (равновесное пространственное расположение атомов) полимерных цепей в пределах нескольких повторяющихся звеньев, что определяет доступность различных атомов в полимерной цепи для проникающей молекулы. Конформация такой «олигомерной» цепи, состоящей из

нескольких повторяющихся звеньев, может быть вычислена намного быстрее, чем типичная полимерная матрица при моделировании молекулярной динамики и механики (см. разделы 1.3.6 и 1.3.5).

Чтобы оценить значения индекса для каждого рассматриваемого полимера, следуем методике, которую можно рассматривать как упрощенную версию метода моделирования молекулярной структуры, описанного в классической статье [38]. Для каждого рассматриваемого полимера готовится «олигомерная цепь», состоящая из нескольких сотен атомов, затем ее геометрия оптимизируется в некотором силовом поле. Следуя этому алгоритму, каждый раз при оптимизации «олигомерной» цепи мы получаем конформацию молекулы полимера. Такой минималистичный подход к геометрии полимерной цепи не учитывает дальнодействующие внутрицепочечные взаимодействия атомов в структуре полимера (например, структура «клубка») и межцепочечные взаимодействия (упаковка цепей), уделяя при этом внимание форме коротких конформаций молекулы: полимерная цепь, например, спираль с шагом 3 или 4 мономера в случае поли (триметилсилилпропина) (ПТМСП) или беспорядочный набор из жестких «ломанных стержней» в полиимидах (см. пример на рисунке 9б). В то же время это требует довольно скромного времени вычислений, что является важным фактором и преимуществом, учитывая количество пар «газ-полимер» в экспериментальном наборе данных (для каждой пары необходимо вычислить геометрию полимера в качестве предварительного шага к регрессионному анализу).

Конформация молекулы – это пространственное расположение атомов и групп атомов, которое задается набором и последовательностью конфигурационных изомеров и их относительным взаимным расположением в цепи, обусловленным тепловым движением или внешними воздействиями на молекулу. Конформация является важным фактором формирования структуры свободного объема и во многом определяет транспортные характеристики полимерной мембраны (проницаемость, растворимость и диффузию), поэтому важно пра-

вильным образом провести ее моделирование. Вычислительная сложность моделирования молекул размером порядка нескольких сотен атомов, как правило, очень высокая, причем, при увеличении длинны «олигомерной» цепи время и сложность вычисления кратно возрастает, как показано в разделе 2.2.1. Также в результате теплового движения или иных внешних воздействий на молекулу для каждой конфигурации полимерной цепи обычно реализуется бесчисленное множество различных конформаций, следовательно, нельзя смоделировать только одну из них. Однако в работе [74] было показано, что для данного метода достаточно моделировать лишь небольшое число случайных конформаций полимерной цепи фиксированного размера. Подробнее остановимся на этом в разделе 3.2.

2.1.2 Геометрические индексы на основе площади поверхности короткой полимерной цепи

Геометрия полимерных конформаций должна быть преобразована в числовые дескрипторы, используемые в качестве объясняющих переменных в регрессиях для прогнозирования транспортных характеристик, в частности, коэффициента растворимости S, который определяет движущую силу мембранного процесса.

Использование площади поверхности молекулы газа в качестве предиктора растворимости было впервые предложено Ямпольским и др. [114] где коэффициент растворимости в полимере объяснен взаимодействиями между молекулами газа и «поверхностью» молекулы полимера, причем было выяснено, что это взаимодействие можно описать логлинейной зависимостью коэффициента растворимости от площади поверхности молекулы газа. Коэффициенты такой линейной зависимости различны для разных полимеров, и логично предположить, что они как-то зависят от аналогичных параметров молекулы полимера. Остается открытым вопрос, как коэффициенты линейной регрессии зависят от характеристик и структуры полимера. В настоящей диссертации статисти-

ческий подход [114] расширяется за счет рассмотрения площади поверхности полимера в дополнение к площади поверхности молекулы газа. Кроме того, для построения универсального выражения для коэффициента растворимости газов в полимерах предлагаются несколько производных индексов на основе поверхностей и зарядов молекул.

В настоящей работе коэффициенты растворимости предсказываются с помощью геометрических показателей, характеризующих площадь поверхности и поверхностный заряд конформаций олигомеров. Как уже упоминалось, использование площади поверхности молекулы газа в качестве предиктора растворимости было впервые предложено Ямпольским и др. [114] где коэффициент растворимости в полимере объяснен взаимодействиями между молекулами газа и «поверхностью» молекулы полимера. Основной предсказывающей переменной была площадь ван-дер-ваальсовой поверхности (Van-der-Waals surface Area VDWSA) молекул газа (площадь поверхности, образованная ван-дер-ваальсовыми сферами атомов молекулы газа, не покрытая ван-дер-ваальсовыми сферами других атомов молекулы). Коэффициенты растворимости различных газов в полимере продемонстрировали строгий логлинейный закон относительно VDWSA для каждого рассматриваемого полимера. Другой геометрический дескриптор, также проверенный Ямпольским и др. – это доступная площадь поверхности, и она также продемонстрировала хорошую корреляцию.

Площадь доступной поверхности (accessible surface area, ASA) молекулы – это площадь поверхности, описываемой центром шарообразного «зонда» заданного радиуса во всевозможных положениях его касания с ван-дер-ваальсовой поверхностью этой молекулы. Иллюстрация к построению ASA для простой молекулы из нескольких атомов приведена на рисунке 6.

ASA является геометрическим 3D индексом [106], то есть функцией, которая каждой конформации молекулы (пространственному расположению ее атомов) ставит в соответствие число, зависящее только от взаимного располо-

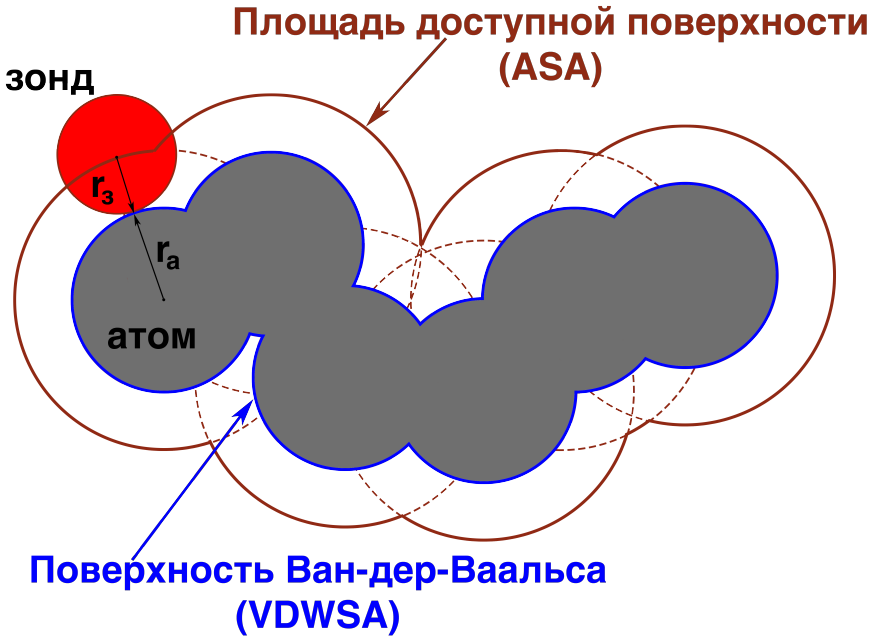


Рисунок 6 – Примеры расчета доступной площади поверхности молекулы ASA.

жения атомов, но не от положения молекулы в пространстве. По сравнению с похожим индексом – площадью поверхности контакта (площадью Конолли) – ASA молекулы имеет преимущество: именно площадь доступной поверхности пропорциональна вероятности контакта между молекулой (полимера) и хаотически движущимся «зондом», представляющим молекулу газа-пенетранта.

ASA для радиуса «зонда» 1.4-1.5 Å (соответствующего воде) широко используется в биохимии для изучения взаимодействия белковой молекулы с растворителями. В мембранной технологии газы-пенетранты отличаются по своему эффективному радиусу, поэтому для целей мембранного газоразделения важно знать ASA (и значения других геометрических индексов) молекул аморфных полимеров для различных радиусов «обкатки» R.

Индексы ASA^+, ASA^-, $PPSA_3$,$PNSA_3$ и $DPSA_3$ (см. таблицу 1), используемые в исследовании, относятся к семейству поверхностно-зарядных индексов (Charged Partial Surface Area descriptors (CPSA)) предложенных в работе [100]

в 1990 году. Подобные индексы составляют набор различных индексов, которые объединяют форму молекулы и информацию о распределении частичных зарядов на поверхности молекулы, а следовательно, кодируют признаки, ответственные за полярные взаимодействия между молекулами. Представление молекулы, используемое для получения индексов CPSA, рассматривает атомы молекулы как твердые сферы, определяемые радиусом Ван-дер-Ваальса. Доступная для растворителя площадь поверхности ASA используется в качестве площади молекулярной поверхности, в работе [100] она рассчитана с использованием сферы радиусом 1,5 Åдля аппроксимации поверхности контакта, образующейся при взаимодействии молекулы воды с рассматриваемой молекулой. Более того, контактная поверхность, на которой могут иметь место полярные взаимодействия, характеризуется специфическим электронным распределением, полученным путем картирования парциальных зарядов атомов на доступной для растворителя поверхности.

Индекс ASA^+ характеризует удельную площадь доступной поверхности, где контакт происходит в точке поверхности с частичным положительным ($q > 0$) зарядом (Å^2 моль/см^3), а индекс ASA^- показывает удельную площадь доступной поверхности, где контакт происходит в точке поверхности с частичным отрицательным ($q < 0$) зарядом (Å^2 моль/см^3). Частичные положительные и отрицательные заряды рассчитываются с использованием итерационной процедуры, представленной в [40]. Использование поверхностно-зарядных индексов ASA^+ и ASA^- нацелено на получение характеристик для оценки положительных и отрицательно заряженных поверхностей. Индекс ASA_H характеризует удельную площадь доступной гидрофобной (с низким, $|q| < 0.125$, уровнем частичного заряда) поверхности(Å^2 моль/см^3), а индекс ASA_P показывает удельную площадь доступной полярной (с высоким, $|q| > 0.125$, уровнем частичного заряда) поверхности (Å^2 моль/см^3). Индексы ASA_H и ASA_P чаще всего используются при работе со структурами белков, однако, оказались также полез-

ны и для предсказания характеристик полимеров, используемых в мембранном газоразделении.

Помимо геометрических индексов ASA, ASA^+ и ASA^-, есть несколько специальных индексов, а именно,$PPSA_3$ и $PNSA_3$ представляют собой «положительную / отрицательную площадь поверхности, взвешенную с атомным зарядом, [13, 106]», которые, как полагают, являются точными мерами реактивности из-за сил Ван-дер-Ваальса. Другими словами, $PPSA_3$ дает средний частичный заряд положительно заряженной молекулярной поверхности, $PNSA_3$ дает средний частичный заряд отрицательно заряженной поверхности, а $DPSA_3 = PPSA_3 + PNSA_3$ является чистым зарядом площади поверхности. Более строго индексы определяются как:

$DPSA_3 = \sum_i asa_i \cdot q_i$, где asa_i– вклад i-го атома в удельную площадь доступной поверхности молекулы, q_i – это частичный заряд i-го атома ($\text{Å}^2\mathbf{e}$ моль/см^3).

$PPSA_3 = \sum_i asa_i \cdot q_i$, где суммирование ограничено атомами с положительным частичным зарядом: $asa_i \cdot q_i > 0$, ($\text{Å}^2\mathbf{e}$ моль/см^3).

$PNSA_3 = \sum_i asa_i \cdot q_i$, где суммирование ограничено атомами с отрицательным частичным зарядом: $asa_i \cdot q_i < 0$, ($\text{Å}^2\mathbf{e}$ моль/см^3).

В таблице 1 приведен список геометрических индексов с краткими описаниями, которые были рассчитаны для молекул газа и полимеров и затем использованы для прогнозирования коэффициента растворимости. Геометрические индексы полимера нормированы на единицу объема с использованием экспериментальной плотности полимера там, где она была измерена. В биохимии также широко используются несколько родственных ASA геометрических индексов [106]. В мембранном газоразделении «обкатка» полимерных цепей сферическими зондами радиуса R, моделирующего размер молекулы газа-пенетранта, применялась И. Роновой с сотрудниками [13, 93, 94] для вычисления доступного свободного объема (free accessible volume, FAV) молекулярных цепей, однако, насколько нам известно, кривые зависимости $ASA(R)$ полимерных материа-

лов в задачах мембранной науки и технологии никогда не исследовались. В настоящей работе именно параметры этой кривой и аналогичных ей позволяют предсказать коэффициент растворимости газа и константу равновесия закона Генри.

Таким образом, необходимо получить значения каждого из предложенных геометрических индексов путем варьирования радиуса «обкатки». По вычисленным значениям геометрических индексов в зависимости от значения радиуса «обкатки» можно построить кривые зависимости (например, $ASA(R)$), которые и будут определять влияние молекулярной структуры полимера на транспортные параметры полимерной мембраны.

Рисунки 7 и 8 демонстрируют примеры зависимостей индексов от радиусов пенетрантов для трех полимеров различных химических классов:

- 11z01 поливинил хлорид из класса виниловые полимеры,
- 27z01 поликарбонат из поликарбонаты,
- 26z09 поли(фенолфталеинфталат) из полиэфиров.

Номера полимеров соответствуют полю PolymerID Базы данных «Газоразделительные параметры стеклообразных полимеров» (см. подробнее в разделе 3.2).

Поверхностные геометрические индексы рассчитаны для радиусов пенетрантов из набора [0, 0.05, 0.1, 0.15, 0.2, 0.3, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8, 2, 2.2, 2.4, 2.6, 2.8, 3], а затем аппроксимированы на диапазон значений радиусов с шагом 0.05 в пределах от 0 до 3 Å. Выбор диапазона и шага изменения в данном ряде радиусов пенетрантов осуществлялся согласно таблице эффективных радиусов по Теплякову и Мересу (таблица 2). В таблице также приведены значения максимальной площади проекции шаровой модели молекул газов.

Специальные индексы призваны отразить некоторые аспекты взаимодействия между частично заряженными атомами молекулы полимера и молекулой проникающего газа. Рассмотрим пример полиимидной цепи полимера BPADA-

Таблица 1 – Площадь доступной поверхности и родственные геометрические индексы.

Индекс	Описание индекса
ASA	Удельная (на один см^3 образца) площадь доступной поверхности (Å^2 моль/см^3)
ASA^+	Удельная площадь доступной поверхности, где контакт происходит в точке поверхности с частичным положительным ($q > 0$) зарядом (Å^2 моль/см^3)
ASA^-	Удельная площадь доступной поверхности, где контакт происходит в точке поверхности с частичным отрицательным ($q < 0$) зарядом (Å^2 моль/см^3)
ASA_H	Удельная площадь доступной гидрофобной (с низким, $\|q\| < 0.125$, уровнем частичного заряда) поверхности(Å^2 моль/см^3)
ASA_P	Удельная площадь доступной полярной (с высоким, $\|q\| > 0.125$, уровнем частичного заряда) поверхности (Å^2 моль/см^3)
$DPSA_3$	$DPSA_3 = \sum_i asa_i \cdot q_i$, где asa_i– вклад i-го атома в удельную площадь доступной поверхности молекулы, q_i – это частичный заряд i-го атома (Å^2**e** моль/см^3)
$PPSA_3$	$PPSA_3 = \sum_i asa_i \cdot q_i$, где asa_i– где суммирование ограничено атомами с положительным частичным зарядом: $asa_i \cdot q_i > 0$, (Å^2**e** моль/см^3)
$PNSA_3$	$PNSA_3 = \sum_i asa_i \cdot q_i$, где asa_i– где суммирование ограничено атомами с отрицательным частичным зарядом: $asa_i \cdot q_i < 0$, (Å^2**e** моль/см^3)
$VDWV$	Удельный (на один см^3) Ван-дер-Ваальсов объем макромолекулы (Å^3 моль/см^3): объем, очерчиваемый ван-дер-ваальсовой поверхностью, то есть доступной поверхностью при нулевом радиусе “обкатки”

Таблица 2 – Радиусы популярных газов-пенетрантов по Теплякову и Мересу, а также максимальные площади проекции молекулы.

Газ	R, Å	$MaxPA$, Å^2
He	0.89	6.158
H_2	1.07	5.610
Ne	1.15	7.451
O_2	1.45	10.205
Ar	1.485	11.104
CO_2	1.51	14.113
CO	1.52	14.303
N_2	1.52	10.482
CH_4	1.59	12.047
Kr	1.61	12.819
C_2H_2	1.69	14.044
Xe	1.76	14.657
C_2H_4	1.785	17.058
H_2S	1.80	13.539
SO_2	1.80	17.993
NF_3	1.81	16.747
C_2H_6	1.845	17.929
C_4H_6	1.92	27.638
C_3H_6	1.93	22.416
C_3H_8	2.05	23.667
C_4H_{10}	2.20	29.470
CF_4	2.35	17.649
NH_3	2.35	10.958
C_2F_6	2.55	25.449
SF_6	2.75	26.177

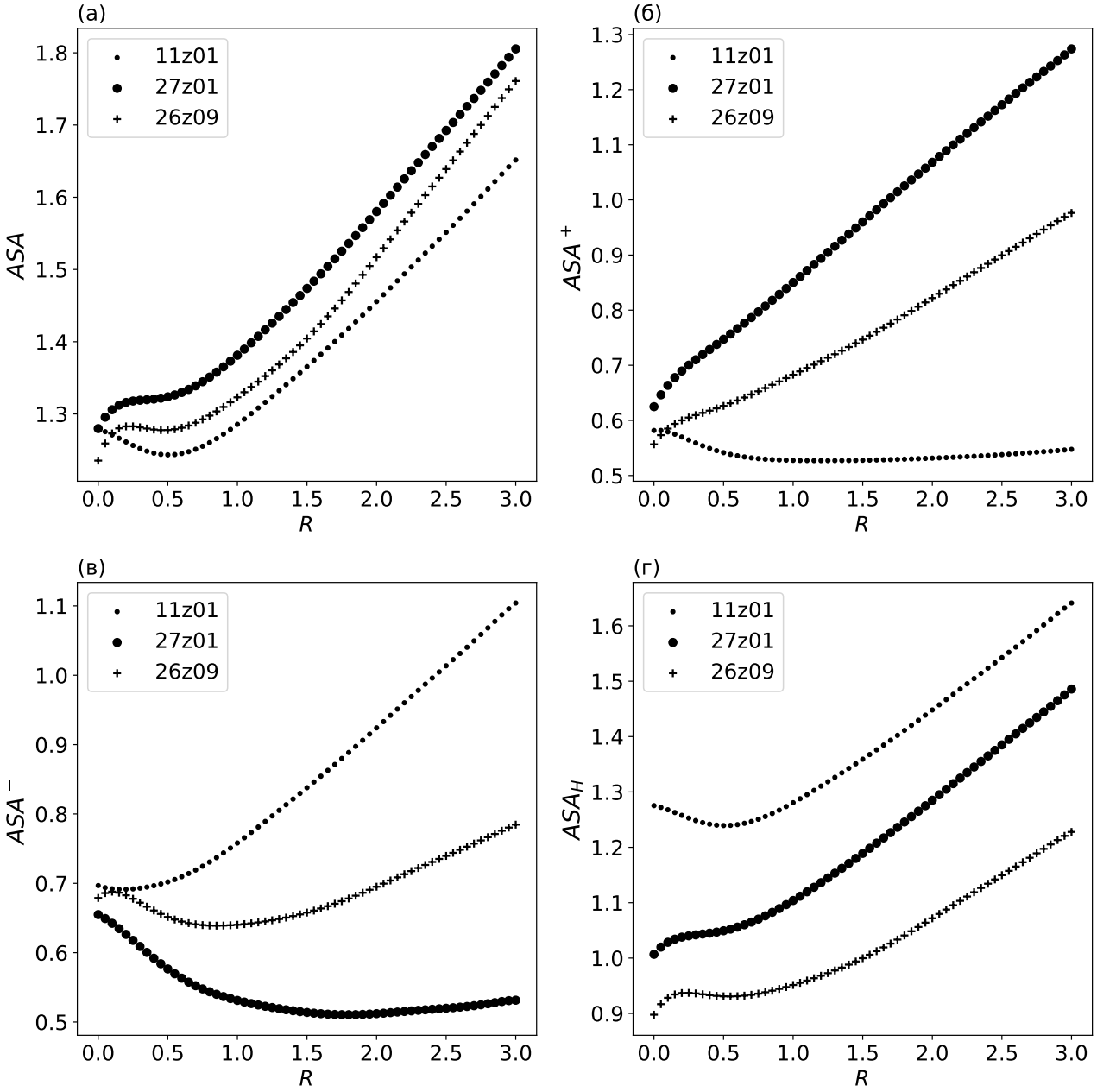


Рисунок 7 – Примеры зависимостей индексов ASA, ASA^+, ASA^-, ASA_P, от радиусов пенетрантов трех полимеров из различных классов.11z01 поливинил хлорид из класса виниловых полимеров, 27z01 поликарбонат из поликарбонатов, 26z09 поли(фенолфталеинфталат) из полиэфиров.

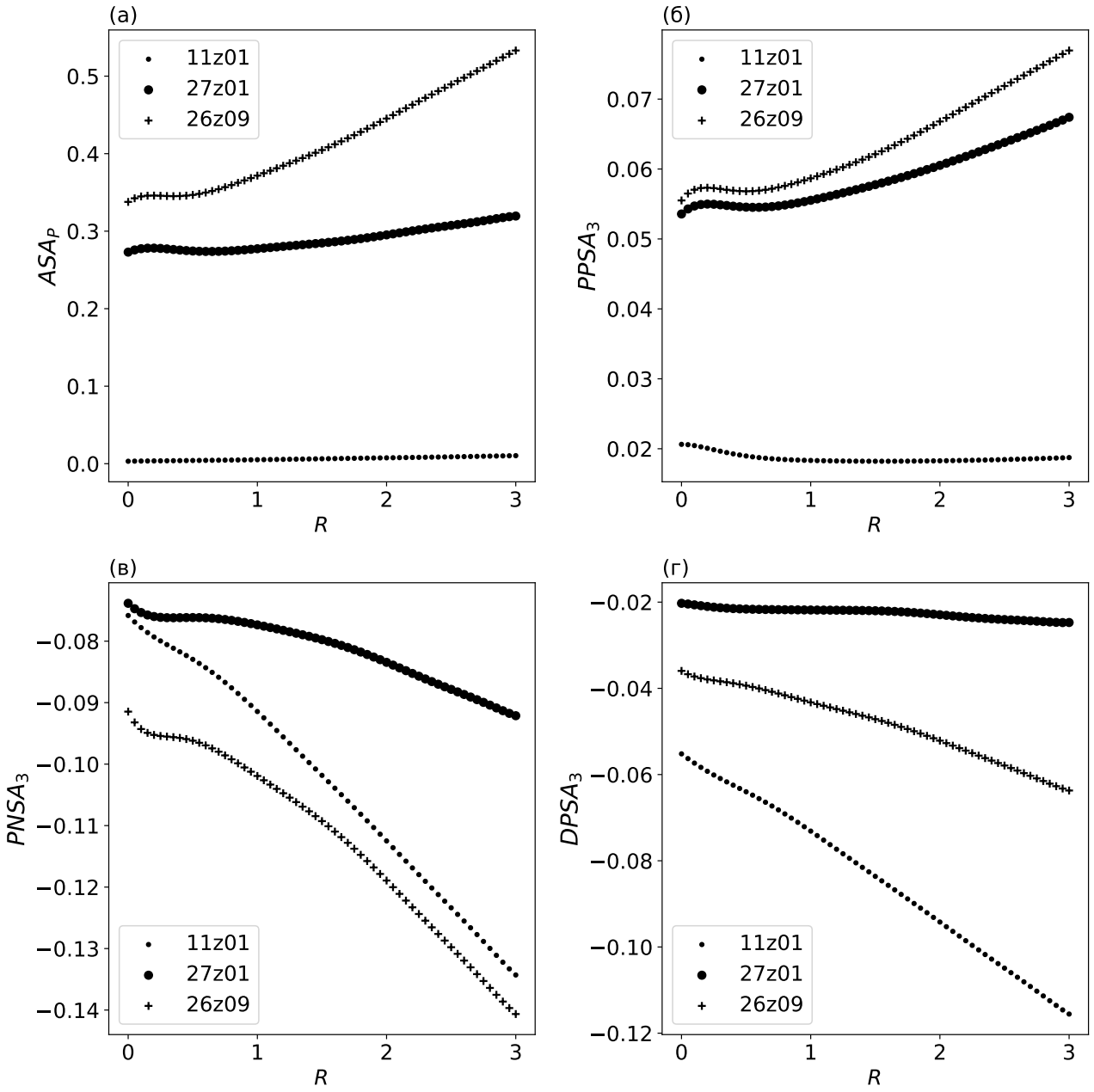


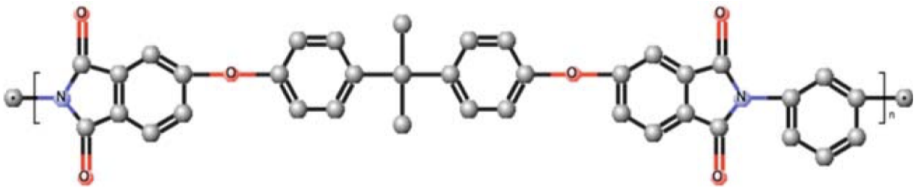
Рисунок 8 – Примеры зависимостей индексов ASA_H, $PPSA_3$,$PNSA_3$ и $DPSA_3$ от радиусов пенетрантов трех полимеров из различных классов. 11z01 поливинил хлорид из класса виниловых полимеров, 27z01 поликарбонат из поликарбонатов, 26z09 поли(фенолфталеинфталат) из полиэфиров.

mPDA (Ultem) (ее повторяющееся звено показано на рисунке 9а, а конформация олигомера вместе с его молекулярной поверхностью показана на рисунке 9б и 9в соответственно.). Области красного цвета на рисунке 9в имеют отрицательный частичный заряд, а области синего цвета – положительный. Поверхности молекул газа и полимера, имеющие заряды противоположного знака, будут притягиваться, а поверхности с зарядом одного знака будут отталкивать друг друга. Предполагается, что амплитуда силы притяжения / отталкивания для пары взаимодействующих элементарных заряженных областей молекул газа и полимера пропорциональна произведению их частичных зарядов.

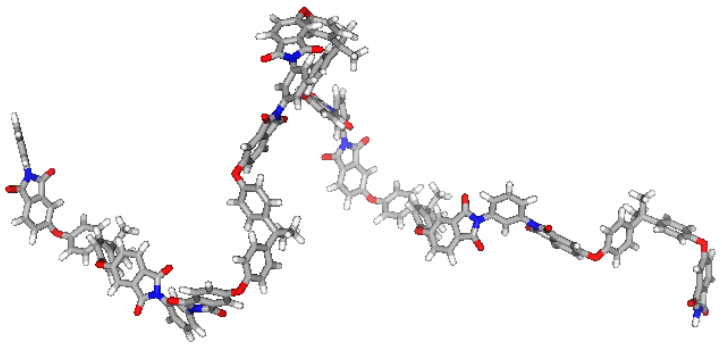
Понятно, что доступная площадь поверхности (ASA, ASA^+ и т. д.), рассчитанная для полимера, характеризует систему «газ-полимер», поскольку каждая площадь рассчитывается с использованием определенного радиуса проникающего газа. Ниже нам потребуются показатели, характеризующие только полимер. Чтобы выявить такие производные индексы, отметим, что ASA (усредненная по конформациям) может быть аппроксимирована линейной зависимостью от радиуса R (газа) газа-пенетранта (см. типичные линии интерполяции для различных классов полимеров на рисунке 10). Зависимость ASA от характеристик пары «газ-полимер» можно записать в виде:

$$ASA(polymer, gas) = c_{ASA}(polymer) + d_{ASA}(polymer)R(gas),$$

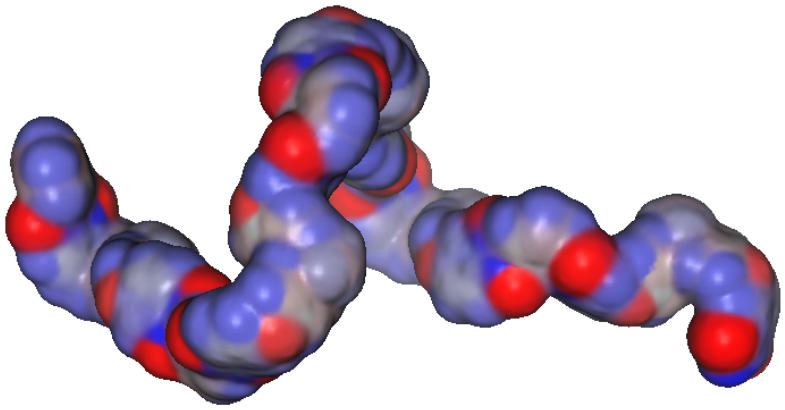
где $c_{ASA}(polymer)$(Å^2 моль/см^3) и $d_{ASA}(polymer)$(Å моль/см^3) зависят от формы конформации олигомера. Следовательно, совокупность площадей поверхности, рассчитанных для одного полимера при различных радиусах газа-пенетранта, порождает два новых дескриптора полимера: «начальная» площадь поверхности c_{ASA} и наклон площади поверхности d_{ASA}. Эти производные индексы оцениваются для каждого полимера методом наименьших квадратов из серии значений ASA, рассчитанных для эффективных радиусов рассматриваемых газов на некотором диапазоне радиусов «обкатки». Границы этого



(а) Звено полимера.



(б) Конформация молекулы полимера.



(в) Площадь доступной поверхности *ASA*.

Рисунок 9 – Пример полиимида BPADA-mPDA (Ultem).

диапазона – минимальный и максимальный радиусы «обкатки» – являются настраиваемыми параметрами предлагаемого метода.

Для других используемых индексов на основе площади поверхности (ASA^+, ASA^-, ASA_P, ASA_H, $DPSA_3$, $PPSA_3$, $PNSA_3$) аналогичным образом строятся линейные зависимости с эффективным радиусом газа-пенетранта, и таким же образом каждый индекс, основанный на площади поверхности, дает пару производных индексов, которые характеризуют геометрию олигомера и заряд поверхности (например, c_{ASA^+}; d_{ASA^+} и т. д.). Все эти индексы используются позже при прогнозировании транспортных характеристик полимеров.

Корреляция коэффициента растворимости со свободным объемом не была в центре внимания данного исследования, поскольку для точной оценки свободного объема требуется расширенная модель упаковки полимерных цепей, которая отсутствует в рассматриваемой упрощенной одноцепочечной модели молекулярной конформации. Тем не менее, были попытки добавить молекулярный объем Ван-дер-Ваальса $VDWV$ в список молекулярных индексов, рассматриваемых как очень приблизительный показатель свободного объема, поскольку, согласно двухрежимной модели сорбции (1.11), сорбционная емкость Ленгмюра CH'_H является частью растворимости, а коэффициент CH'_H, очевидно, коррелирует с долей свободного объема, оцененной как

$$FFV = 1 - 1.3VDWV\rho. \tag{2.1}$$

Кроме того, известно, что CH'_H ведет себя так же, как свободный объем ν_f: сообщалось о многочисленных корреляциях для CH'_H, параллельных корреляциям для ν_f [62]. Аналогичный подход к оценке свободного объема был принят в [56].

Следует подчеркнуть, что неточность, возникающая из-за пренебрежения влиянием свободного объема на значение коэффициента растворимости, не должна быть очень большой. Известно, что коэффициент растворимости при бесконечном разбавлении (или при достаточно низком давлении) может быть

выражен как $S = k_D + CH'_H b$, где k_D и b – параметры двухрежимной сорбции. Для высокопроницаемых полимеров с большим свободным объемом, таких как ПТМСП, выполняется следующее неравенство: $k_D \ll CH'_H b$, и пренебрежение ролью упаковки цепей может быть источником ошибок. Однако, как будет показано далее, большинством объектов анализа были полимеры с относительно низкой проницаемостью и свободным объемом, поэтому роль этого фактора сведена к минимуму.

Таким образом, в настоящем разделе предложено использовать коэффициенты линейной регрессии, аппроксимирующей на разных участках кривую зависимости поверхностных геометрических индексов (ASA, ASA^+ и других) от радиуса «обкатки» для предсказания транспортных характеристик полимеров. В следующем разделе описывается способ построения регрессионной модели.

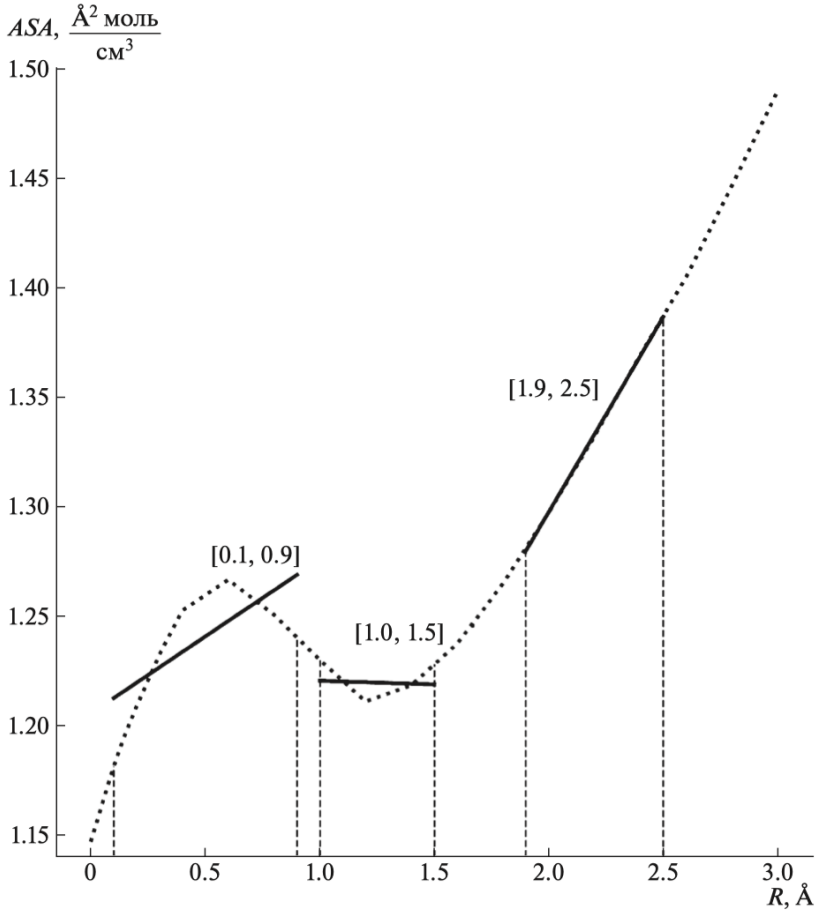


Рисунок 10 – Пример построения линейной аппроксимации зависимости $ASA(R)$ для полимера ПВТМС на разных отрезках $[R^-, R^+]$.

2.1.3 Регрессионная модель

В этой работе статистический подход Ямпольского и др. [114] был расширен за счет рассмотрения площади поверхности полимера в дополнение к площади поверхности молекулы газа. Более того, в работе [44] было показано, что *MaxPA* – максимальная площадь проекции молекулы газа (где атомы представлены своими ван-дер-ваальсовыми сферами) – позволяет еще лучше описать линейную зависимость логарифма коэффициента растворимости газа в полимере для большинства популярных в мембранной технологии легких газов (за исключением окиси углерода, демонстрирующей девиантное поведение):

$$\log S(gas, polymer) = a(polymer) MaxPA(gas) + b(polymer). \qquad (2.2)$$

При этом коэффициенты a и b линейной зависимости являются функциями полимера p, отличаясь для различных полимеров. В приближении первого порядка предполагается линейная связь между коэффициентами регрессии a и b, с одной стороны, и дескрипторами полимерных цепей (c_{ASA}, d_{ASA} и другие, описанные выше), с другой стороны. Следовательно, если обозначить геометрические индексы полимера буквами I_1,. . ., I_n, коэффициенты линейной регрессии (2.2) могут быть записаны как

$$a(polymer) = a_0 + \sum_{j=1}^{n} a_j I_j(polymer) \qquad (2.3)$$

$$b(polymer) = b_0 + \sum_{j=1}^{n} b_j I_j(polymer) \qquad (2.4)$$

где единицы измерения коэффициентов a_j и b_j зависят от единиц измерения соответствующих геометрических индексов.

Подставив (2.3) и (2.4) в (2.2), получим

$$\log S(gas, polymer) = \left(a_0 + \sum_{j=1}^{n} a_j I_j(polymer) \right) \cdot MaxPA(gas) + b_0 + \sum_{j=1}^{n} b_j I_j(polymer). \qquad (2.5)$$

Эту форму можно свести к стандартной линейной регрессии, добавив вспомогательные переменные «квадратичный член», вычисленные как

$$Q_j(\cdot) := MaxPA(gas) \times I_j(\cdot), j = 1, ..., n. \tag{2.6}$$

Тогда

$$\begin{aligned} \log S(gas, polymer) = a_0 \cdot MaxPA(gas) + b_0 + \\ \sum_{j=1}^{n} a_j Q_j(gas, polymer) \ + \sum_{j=1}^{n} b_j I_j(polymer). \end{aligned} \tag{2.7}$$

Линейная регрессия (2.7) может быть построена с использованием стандартных методов ANOVA и других статистических инструментов для линейной регрессии, что подробно будет описано в разделе 2.2.3.

Коэффициент $a(polymer)$ (с размерностью $[\log(cm^3(STP)/cm^3(cmHg))\text{Å}^2]$) представляет особый интерес, поскольку играет ключевую роль в анализе селективности полимеров.

Селективность проницаемости полимера *polymer* по отношению к паре газов gas и $gas^{'}$ определяется как

$$\alpha_P(polymer|gas, gas^{'}) := \frac{P(polymer, gas)}{P(polymer, gas^{'})}. \tag{2.8}$$

Селективность растворимости $\alpha_S(polymer|gas, gas^{'})$ и селективность диффузии $\alpha_D(polymer|gas, gas^{'})$ определяются таким же образом, как отношения соответствующих коэффициентов растворимости и диффузии.

Используя уравнение $P = S \cdot D$ и записав селективность по проницаемости в логарифмической шкале (что типично для диаграммы Робсона, см. раздел 1.1), можно записать $\alpha_P(polymer|gas, gas^{'})$ как сумму селективности растворимости и селективности диффузии, также выраженных в логарифмической шкале:

$$\begin{aligned} \log(\alpha_P(polymer|gas, gas^{'})) := \log(\alpha_S(polymer|gas, gas^{'})) + \\ +\log(\alpha_D(polymer|gas, gas^{'})). \end{aligned} \tag{2.9}$$

Основное внимание уделяется второму члену, то есть селективности растворимости. Записывая коэффициенты растворимости по формуле (2.7), селективность растворимости сводится к следующему выражению:

$$\log(\alpha_P(\text{polymer}|\text{gas}, \text{gas}^{'})) := a(\text{polymer}) \cdot [MaxPA(\text{gas}) - MaxPA(\text{gas}^{'})]. \quad (2.10)$$

Сохраняется только коэффициент $a(polymer)$ из регрессии (2.2) в выражении (2.10) и, следовательно, все остальные коэффициенты и переменные не имеют отношения к анализу селективности растворимости.

С одной стороны, из выражения (2.2) можно сделать вывод, что чем больше разница максимальных площадей проекции молекул растворенных веществ gas и $gas^{'}$, тем больше селективность растворимости для этой пары газов и любого данного полимера $polymer$. С другой стороны, абсолютное значение селективности растворимости увеличивается по величине $a(polymer)$, и полимеры, имеющие максимум $a(polymer)$ демонстрируют максимальную селективность растворимости для любой данной пары газов gas и $gas^{'}$. Следовательно, значение $a(polymer)$ можно рассматривать как «универсальную селективность» полимера.

2.2 Алгоритмы предсказания транспортных характеристик полимерных мембран

В данном разделе подробно разобраны численные методы и алгоритмы, применяемые для моделирования транспортных характеристик в ППКПЦ. В начале раздела описан метод получения конформаций полимерных цепей, которые получаются в результате проведения специфической процедуры молекулярно-механического моделирования. Затем разобран модифицированный метод Ли-Ричардса, позволяющий рассчитать площадь поверхности «обкатки» молекулы полимера газом-пенетрантом, а также поверхностно-зарядные геометрические индексы. В конце главы приведен метод построения регрессионных моделей

предсказания транспортных характеристик на основе производных геометрических индексов.

2.2.1 Метод получения конформаций

При разработке метода ППКПЦ требовалось разработать численный метод оптимизации потенциальной энергии молекулы полимера, на основе методов молекулярно-механического моделирования, с целью генерации представительного набора конформаций полимерных цепей. Разработанный численный метод должен быть применим для специфических полимеров, используемых в мембранном газоразделении, и удовлетворять следующим требованиям:

1. точность – результат моделирования должен соответствовать или быть максимально близким к тому, как в реальности располагается молекула в полимерной мембране, а также .
2. устойчивость – метод должен обеспечивать стабильность получаемых результатов и их воспроизводимость,
3. экономичность – время вычисления конформаций полимерных цепей не должно превышать несколько часов.

В первой версии метода, представленной в работе [44], для каждого полимера создавался олигомер длиной порядка 200-600 атомов, состоящий из нескольких мономерных звеньев полимера. Для этой цепочки генерировались 6 случайных конформаций различной геометрии, получаемой оптимизацией в эмпирическом силовом поле Дрейдинга [70] из различных начальных позиций атомов. Несмотря на то, что с помощью полученных конформаций был довольно успешно предсказан один из транспортных параметров полимерных мембран – растворимость S – данный подход не соответствовал приведенным выше критериям. Самым важным недостатком является нереалистичность многих полученных конформаций.

В работе [43] был предложен модифицированный метод, решающий проблему нереалистичности конформаций. Для задания случайного начального поло-

жения атомов к моделируемой полимерной цепи проводилось молекулярнодинамическое моделирование при температуре от 300 K до 3000 K в течение 1000 итераций, после чего полученная структура оптимизировалась по свободной энергии в эмпирическом поле MM2. Данный подход показывает неплохие результаты в плане реалистичности конформаций, но, как и подход из работы [44], имеет ряд недостатков при реализации в виде программного обеспечения, о чем рассказано в главе 3.

В численных методах создания репрезентативных конформаций молекул используются две основные стратегии поиска: систематическая и стохастическая. В первом подходе каждая способная к вращению связь систематически отбирается через дискретные интервалы, что ограничивает его использование молекулами с небольшим количеством вращающихся связей. Стохастические методы случайным образом выбирают конформационное пространство молекулы и, таким образом, могут применяться к более гибким молекулам. Также различные методы используют разную степень экспериментальных данных для генерации конформаций. Существуют методы, использующие предопределенные библиотеки углов вращения и конформаций колец. С другой стороны, в дистанционно-геометрическом подходе используется меньшее количество эмпирической информации (длины связей, валентные и торсионные углы). Дистанционно-геометрический подход – это быстрый в вычислительном отношении метод создания конформаций, но у него есть недостаток, заключающийся в том, что ограничения, основанные исключительно на расстоянии, имеют тенденцию приводить к искажению ароматических колец. Чтобы исправить это, результирующие конфигурации часто минимизируются с помощью силового поля, что увеличивает вычислительную сложность и время вычисления. В предлагаемом методе используются результаты работ [32, 90], представляющие альтернативную стратегию, которая сочетает в себе дистанционно-геометрический подход и торсионные углы, полученные из кристаллографиче-

ских данных малых молекул. Углы вращения описываются ранее разработанным набором иерархически структурированных шаблонов SMARTS.

Оригинальность предложенного метода моделирования определяется использованием процедуры, аналогичной процедуре полимеризации при создании полимерного образца, которая позволила кардинально уменьшить время вычислений, за счет пошаговой оптимизации потенциальной энергии молекулы с дополнительными граничными условиями.

Численный метод получения конформаций реализован алгоритмом 1.

Входные данные: Структура мономера, N – число звеньев

Выходные данные: Конформация

1 $k := 1$

2 **до тех пор, пока** $k \neq N$ **выполнять**

3 Загрузка структуры k-ого мономера и удаление атомов водорода

4 Расположение в пространстве атомов методом ETKDG k-ого мономера со случайными начальными координатами

5 Присоединение k-ого мономера к полимеру

6 Фиксация координат атомов мономеров до $k - 1$

7 Расположение в пространстве всех атомов полимера

8 $k := k + 1$

9 **конец цикла**

10 Добавление атомов водорода

11 Молекулярно-механическое моделирование в силовом поле MMFF94

Алгоритм 1– Алгоритм получения конформаций.

Вначале производится загрузка структуры звена полимера (мономера). На втором шаге для ускорения вычислений на дальнейших этапах алгоритма скрываются атомы водорода H. На третьем шаге производится расположение атомов мономера в трехмерном пространстве. На четвертом шаге предобработанный мономер добавляется к цепочке мономеров, обработанных аналогичным обра-

зом ранее. На пятом шаге координаты всех атомов молекулы, за исключением последних двух мономеров в цепочке, фиксируются. На шестом шаге с помощью метода ETKDG [32, 90] производится поиск допустимого положения двух последних мономерных звеньев в трехмерном пространстве, который является результатом комбинирования стандартного метода метрической геометрии с различными экспериментальными фактами и знаниями из органической химии. Например, с информацией о предпочтительных торсионных углах, полученных из экспериментально определенных кристаллических структур (ETDG), или с использованием знаний о том, что ароматические кольца должны быть плоскими или что связи, связанные с тройными связями, являются коллинеарными. Далее алгоритм повторяется, пока количество мономеров не достигнет заданного числа, или число атомов в молекуле не превысит выставленный лимит. На финальных этапах алгоритма атомы водорода возвращаются в молекулы и методом градиентного спуска производится минимизация потенциальной энергии всей молекулы в эмпирическом силовом поле MMFF94. Таким образом моделируется процесс последовательной полимеризации молекулы, когда все новые присоединяемые мономеры влияют лишь на соседние с ними мономеры цепи. При этом для реалистичности получаемых конформаций метод учитывает экспериментальные значения торсионных углов и использует базу универсальных правил взаимного расположения атомов, например, обязательное нахождение в одной плоскости атомов бензольного кольца.

Для обеспечения устойчивости метода для каждого полимера вычисляются параметры стольких конформаций, сколько необходимо для получения представительной выборки. Примеры полученных конформаций представлены на рисунке 12.

Для оценки времени вычисления конформации случайным образом отобраны 50 полимеров, для каждого из которых построены олигомерные цепи различной длины от 100 до 600 атомов. Для каждой олигомергой цепи вычислено по 6 конформаций. На рисунке 11 представлен график зависимости времени вычис-

ления конформации от числа атомов в полимере. Синим изображена медиана по времени, а серыми областями выделены области 25-75 и 10-90 процентилей. Вычисления производились с помощью языка Python на серверных мощностях (Ubuntu версии 18.04.6, процессор Intel(R) Xeon(R) Silver 4210 CPU 2.20GHz) с использованием распараллеливания по потокам. Увеличение числа атомов в цепи влечет нелинейный рост времени вычисления конформаций. При длине олигомерной цепи более 700 атомов время вычисления одной конформации выходит за сутки, что может доставить неудобства при работе на персональном компьютере.

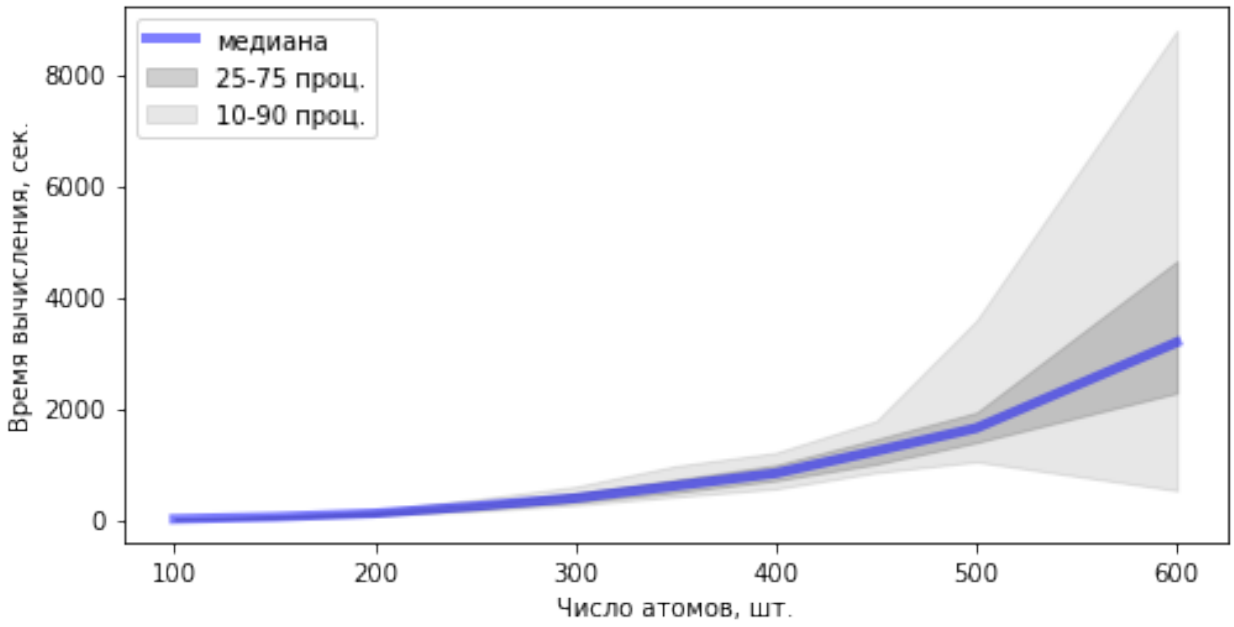


Рисунок 11 – График зависимости времени вычисления конформации от числа атомов в полимере.

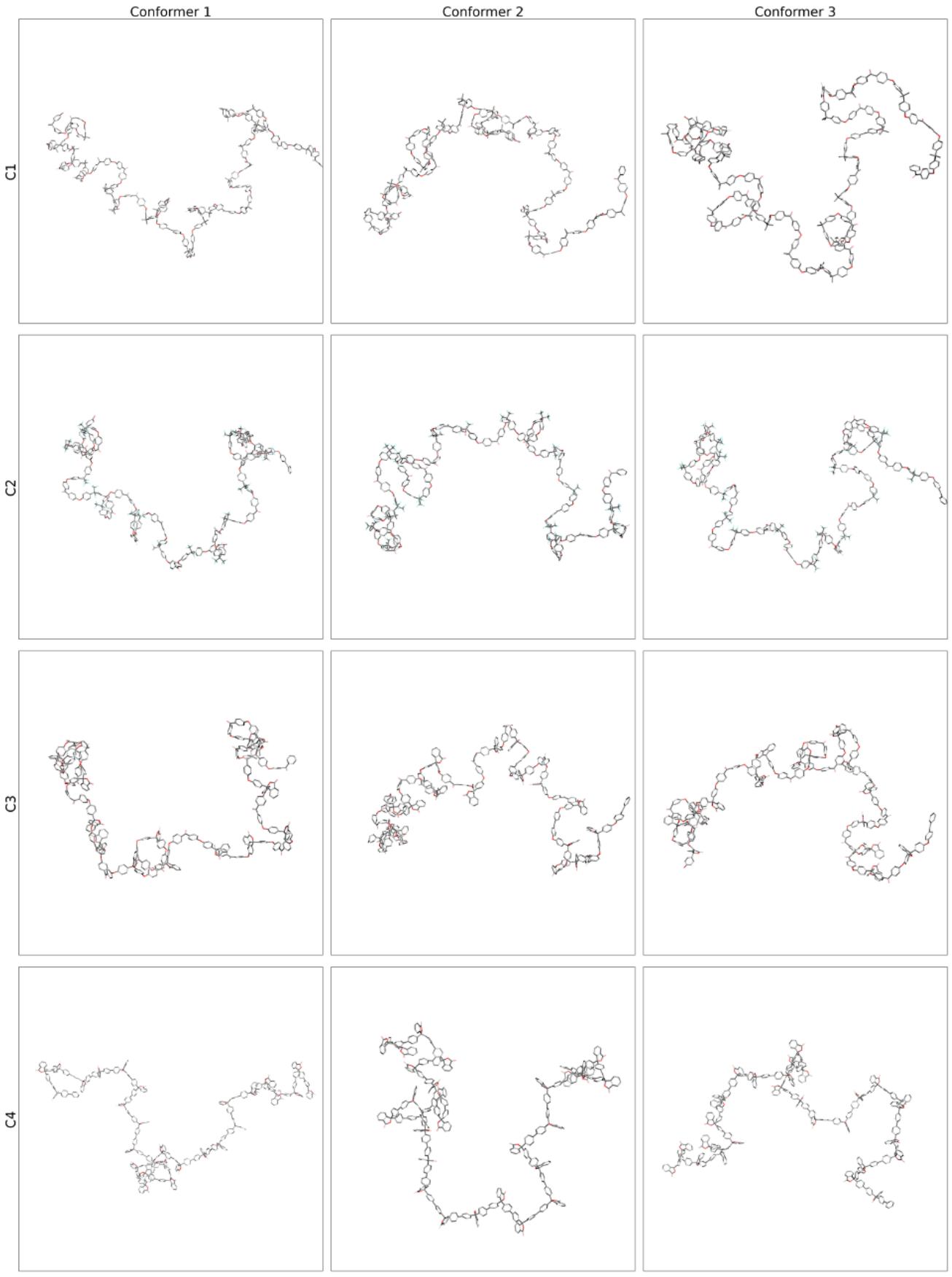


Рисунок 12 – Примеры 3 из 10 случайных конформаций для полимеров C_1, C_2, C_3, C_4 из подраздела 3.2.

2.2.2 Метод вычисления геометрических индексов

Задача вычисления площади поверхности доступной для «обкатки» ASA сводится к задачи вычисления площади криволинейной поверхности в трехмерном пространстве. Для ее решения известны аналитические, так и численные методы. Впервые численный метод расчета площади поверхности доступной для «обкатки» был представлен Ли и Ричардсом в работе [67] в 1971 году, где поверхность аппроксимируется объединением множеств параллельных срезов, а площадь вычисляется интегрированием по этому множеству. Вторым популярным численным методом стал метод Шрайка-Рипли, описанный в работе [99] 1973 года, где поверхность каждой ван-дер-ваальсовой сферы аппроксимируется набором контрольных точек и производится подсчет числа точек не перекрываемых другими. ASA можно рассчитать с произвольной точностью, улучшив разрешение обоих методов аппроксимации. Площадь поверхности также может быть рассчитана аналитически [36], что полезно, когда необходим градиент, или с помощью различных других приближений, приспособленных для различных целей [97, 109].

Опишем принцип работы алгоритма Шрайка-Рипли. Первым шагом на молекуле задается равномерно распределенная сетка из контрольных точек t_i, равноудаленных от каждого гидратированного (расширенного) центрального атома. Затем каждая контрольная точка определяется как скрытая соседним атомом (пробным атомом с центром A и гидратированным радиусом r), если $r > d(A, t_i)$. Для каждого центрального атома ASA рассчитывается путем умножения числа доступных для растворителя контрольных точек на значение площади поверхности, соответствующее каждой контрольной точке.

В отличие от алгортима Шрайка-Рипли, где используется подход на основе сетки, в алгоритме Ли-Ричардса используется расчета длины дуги окружности. Опишем принцип работы алгоритма Ли-Ричардса. Пусть атом i имеет радиус Ван-дер-Ваальса r_i, сфера, которой обкатывается молекула, (или зонд)

имеет радиус r_P, и когда они складываются, мы получаем расширенный радиус $R_i = r_i + r_P$. Сфера радиуса R_i с центром в центре атома i представляет собой объем, недоступный центру зонда для «обкатки». Таким образом, чтобы получить ASA для молекулы необходимо рассчитать площади поверхности расширенных сфер, доступных для «обкатки» зондов, и просуммировать их. Алгоритм Ли-Ричардса вычисляет площадь поверхности «обкатки», производя множество срезов молекулы, вычисляя длину контуров, доступных для зонда, в каждом срезе, а затем суммируя длину, умноженную на толщину среза.

С помощью программы FreeSASA [37], о которой будет рассказано подробнее в разделе 3.1.2, было проведено сравнение работы алгоритмов Ли-Ричардса и Шрайка-Рипли на примере полимера поливинилтриметилсилан (PVTMS), изображение которого и пример конформации, приведены на рисунке 13.

Сравнение проводилось для ряда радиусов «обкатки» в промежутке от 0.01 до 5 Å с шагом 0.01 Å. Расчеты проводились для полярной, неполярной и общей площадей «обкатки», рассчитанной с помощью алгоритмов Ли-Ричардса и Шрайка-Рипли с различным количеством срезов: 20, 100, 200. Результаты приведены на рисунке 14. Алгоритм Шрайка-Рипли показывает менее стабильные результаты, получается, что кривая зависимости геометрических индексов от радиуса «обкатки» – не гладкая, что, очевидно, не соответствует действительности.

В отличие от него алгоритм Ли-Ричардса показывает стабильные результаты и полученные кривые зависимости легче аппроксимировать. По этой причине метод Ли-Ричардса был взят за основу при реализации численного метода вычисления площадей поверхности доступных для «обкатки». В отличие от метода Ли-Ричардса, предложенный метод позволяет вычислять интегралы функции частичных атомных зарядов по площади поверхности молекулы доступной для «обкатки».

Алгоритм 2, реализующий численный метод вычисления геометрических индексов, состоит из нескольких этапов. Вначале на вход подается конформация

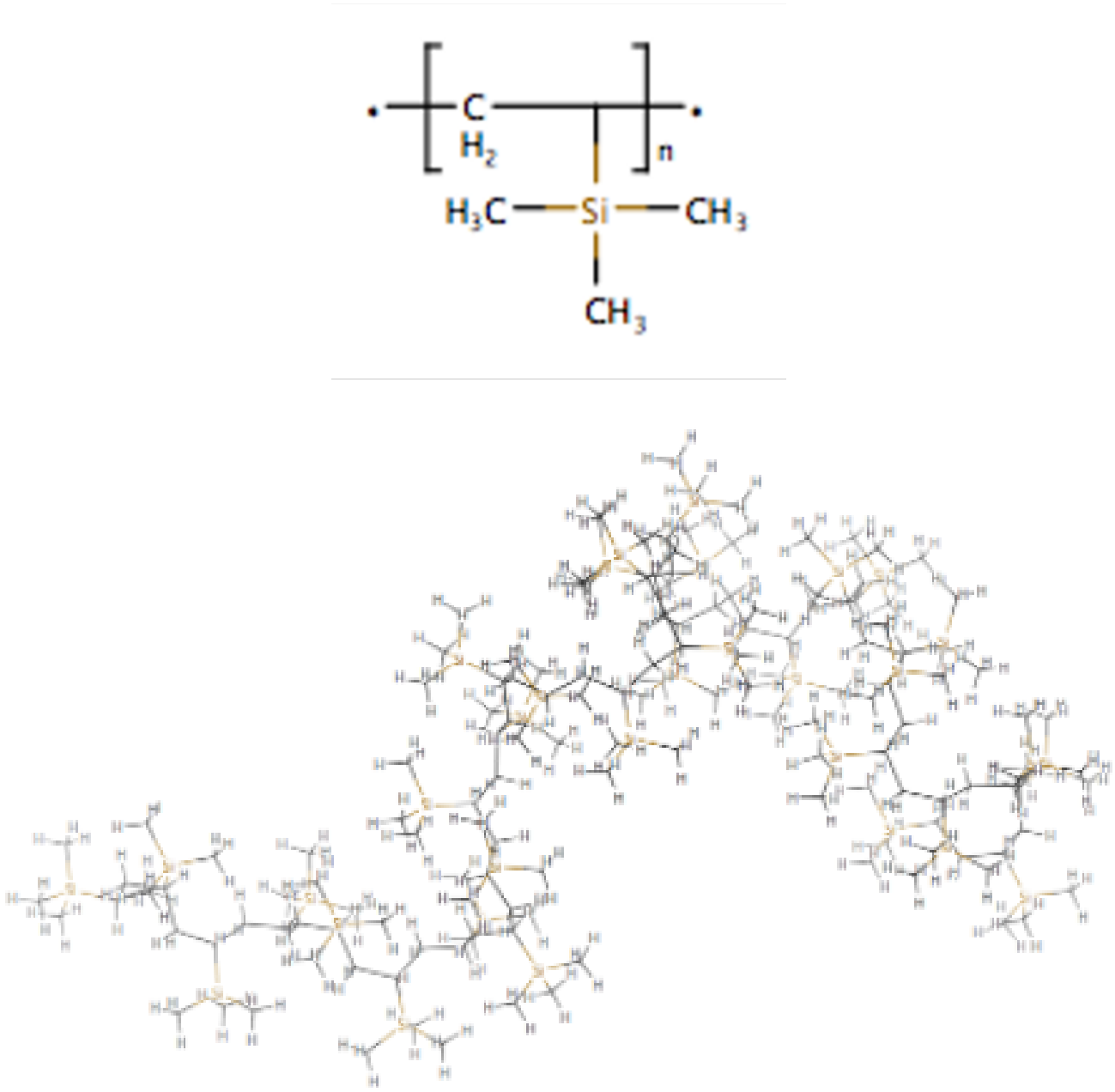


Рисунок 13 – Структура мономерного звена поливинилтриметилсилана (PVTMS) и его конформация.

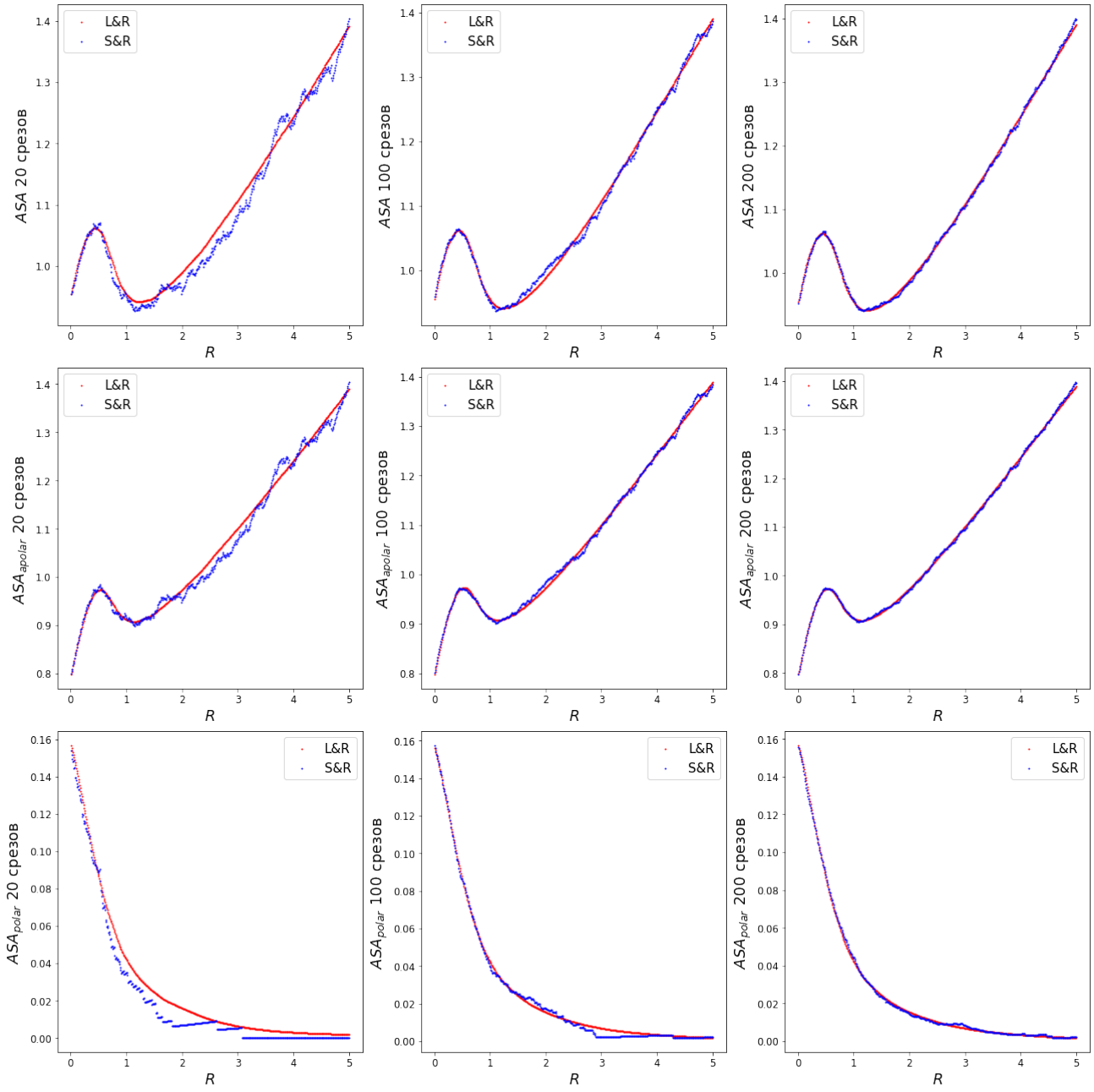


Рисунок 14 – Сравнение рассчитанных площадей «обкатки» с помощью алгоритмов Ли-Ричардса и Шрайка-Рипли на примере поливинилтриметилсилана (PVTMS).

молекулы полимера. На втором шаге производится извлечение координат, присвоение атомам радиусов Ван-дер-Ваальса и атомных номеров.

Согласно определению геометрических индексов из таблицы 1, индексы ASA^+ и ASA^- являются суммой инкрементов площадей атомов заряженных соответственно положительно и отрицательно, а такие геометрические индексы, как ASA_P, ASA_H, $DPSA_3$, $PPSA_3$ и $PNSA_3$ для вычисления требуют не просто знак заряда, но и его значение. Для этого на третьем шаге алгоритма для каждого атома вычисляется его частичный заряд по методу Гастайгеру–Марсели [40,41]. Частичные заряды, рассчитанные по данному методу, дают важный вклад в предсказательную силу моделей, разрабатываемых в данной работе, поэтому ниже метод Гастайгера-Марсили описан подробнее.

Метод Гастайгера-Марсили является конформационно независимым методом, в котором используются данные измерений, а именно, орбитальная электроотрицательность и топология молекулы. Это один из первых численных методов расчета частичного заряда, основанный на электростатической модели, приводящей к частичному выравниванию орбитальной электроотрицательности. В методе функция, соединяющая три значения электроотрицательности атома в его анионном, нейтральном и катионном состояниях, аппроксимируется параболой вида 2.11:

$$\chi_i = a_i q_i^2 + b_i q_i + c_i \tag{2.11}$$

где χ – электроотрицательность атома и функция заряда атома q_i, a_i, b_i, c_i – эмпирические параметры для данного атома, которые оцениваются с использованием потенциалов ионизации и сродства к электрону.

На каждой итерации метода заряд перемещается к более электроотрицательному атому в связи. Перемещенный частичный заряд и его направление определяются разностью электроотрицательностей атомов на каждом конце связи. Затем алгоритм релаксации повторяется несколько раз, уменьшая заряд, перемещаемый с каждой итерацией. Таким образом, с каждым шагом итера-

ции на атом оказывается влияние следующих друг за другом соседних атомов. Итерации повторяются до тех пор, пока значения частичных зарядов не стабилизируются и не будут получены окончательные частичные заряды.

Метод Гастайгера-Марсили является конформационно независимым методом, в котором используются данные измерений, а именно, орбитальная электроотрицательность и топология молекулы. Это один из первых численных методов расчета частичного заряда, основанный на электростатической модели, приводящей к частичному выравниванию орбитальной электроотрицательности. Частичные заряды вычисляются в два этапа. На первом этапе каждому атому в молекуле присваивается некоторый заряд. На втором этапе эти первоначальные заряды затем распределяются между связями, перемещая определенное количество заряда от одного атома к другому, причем на каждой итерации метода заряд перемещается к более электроотрицательному атому в связи. Величина заряда зависит от разности электроотрицательностей связанных атомов. Вычисление частичных зарядов сводится к итерационному решению системы квадратных уравнений.

Перемещенный частичный заряд и его направление определяются разностью электроотрицательностей атомов на каждом конце связи. Затем алгоритм релаксации повторяется несколько раз (по умолчанию восемь проходов), ослабляя заряд, перемещаемый с каждой итерацией.

где χ – электроотрицательность атома и функция заряда атома q_i, a, b, c – эмпирические параметры для данного атома. На каждой итерации метода заряд перемещается к более электроотрицательному атому в связи. Величина заряда зависит от разности электроотрицательностей связанных атомов. После на каждой итерации метода при движении заряда правая часть уравнений пересчитывается до тех пор, пока значения не стабилизируются и не будут получены окончательные частичные заряды.

На четвертом шаге используется алгоритм Ли-Ричардса, описанный выше, для вычисления инкрементов площади – площади поверхности доступной для

«обкатки» для каждого из атомов, составляющих молекулу. Вычисления производятся для радиусов «обкатки» из набора [0, 0.05, 0.1, 0.15, 0.2, 0.3, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8, 2, 2.2, 2.4, 2.6, 2.8, 3], а затем аппроксимируются на диапазон значений радиусов с шагом 0.05 в пределах от 0 до 3 Å. Выбор диапазона и шага изменения в данном ряде радиусов газов-пенетрантов осуществлялся согласно таблице эффективных радиусов по Теплякову и Мересу (таблица 2).

На пятом шаге с использованием полученных инкрементов площади и частичных зарядов для каждого радиуса газа-пенетранта из вышеуказанного набора вычисляются геометрические индексы ASA^+, ASA^-... согласно формулам из таблицы 1. Результатом работы алгоритма являются кривые зависимости геометрических индексов от величины радиуса газа-пенетранта в пределах от 0 до 3 Å.

Точность численного метода вычисления геометрических индексов регулируется параметром – количеством срезов на которые разбивается каждая ван-дер-ваальсова сфера. Численные эксперименты показали, что для практических вычислений достаточно разбивать сферу на 100 срезов. Устойчивость метода определяется сходимостью вычисляемых геометрических индексов при увеличении числа срезов. Алгоритм допускает реализацию с использованием параллельных вычислений, что позволяет снизить время вычисления геометрических индексов для одной конформации до нескольких минут. Таким образом на основе методов Ли-Ричардса и Гастайгера-Марсили предложен новый численный метод вычисления поверхностно-зарядных геометрических индексов, которые впоследствии используются для предсказания транспортных характеристик стеклообразных полимеров.

Входные данные: Конформация

Выходные данные: $ASA(R)$, $ASA^{+}(R)$, $ASA^{-}(R)$, $ASA_H(R)$,$ASA_P(R)$, $PPSA_3(R)$,$PNSA_3(R)$, $DPSA_3(R)$]

цикл *всех атомов в конформации* **выполнять**
    Присвоение R и атомных номеров
    Расчет заряда атома с помощью алгоритма Гастайгера-Марсили
конец
цикл *всех* $R \in [0, 3]$ **выполнять**
    цикл *всех атомов в конформации* **выполнять**
        Расчет с помощью алгоритма Ли-Ричардса инкремента площади атома для R
    конец
конец
цикл *всех* $I \in$ *[*ASA, ASA^{+}, ASA^{-}, ASA_H,ASA_P, $PPSA_3$,$PNSA_3$, $DPSA_3$*]* **выполнять**
    цикл *для всех* $R \in [0, 3]$ **выполнять**
        Вычисление $I(R)$ согласно таблице геометрических индексов
    конец
конец

Алгоритм 2– Алгоритм получения геометрических индексов.

2.2.3 Подход к обучению регрессионной модели

Итак, в предыдущих подразделах 2.1.2 и 2.1.3 описаны методы построения набора объясняющих переменных – параметров линеаризации c_i, d_i кривых зависимости геометрических индексов ASA, ASA^+, ASA^-, ASA_P, ASA_H, $DPSA_3$, $PPSA_3$, $PNSA_3$ (см. таблицу 1) от радиуса «обкатки» R. Таким образом, после линеаризиации кривых для каждого геометрического индекса получаем два производных геометрических индекса – c_i и d_i. Всего получается набор 18 переменных, описывающих влияние полимера на транспортные характеристики пары «газ-полимер». Число переменных еще удваивается за счет рассмотрения мультипликативных индексов $Q_i = I_i \cdot MaxPA$ (см. раздел 1.3.10).

В связи с большим числом объясняющих переменных при относительно небольшом объеме выборки для предотвращения переобучения было необходимо использовать ту или иную процедуру отбора переменных. Пошаговая регрессия – это метод пошагового итеративного построения регрессионной модели, который включает выбор независимых переменных для использования в окончательной модели. Он включает последовательное добавление или удаление потенциальных объясняющих переменных и проверку статистической значимости после каждой итерации. Основная цель пошаговой регрессии состоит в том, чтобы с помощью серии тестов (например, F-тестов, t-тестов) найти набор независимых переменных, которые значительно влияют на зависимую переменную. Пошаговая регрессия может быть получена либо путем тестирования одной независимой переменной за раз и включения ее в модель регрессии, если она статистически значима, либо путем включения в модель всех потенциальных независимых переменных и исключения тех, которые не являются статистически значимыми. Некоторые используют комбинацию обоих методов, поэтому существует три подхода к пошаговой регрессии:

- *Прямой отбор* начинается с отсутствия переменных в модели, проверяется каждая переменная по мере ее добавления в модель, затем сохраняются

те, которые считаются наиболее статистически значимыми, – процесс повторяется до тех пор, пока результаты не станут оптимальными.

- *Обратное исключение* начинается с набора независимых переменных, удаляя по одной, а затем проверяя, является ли удаленная переменная статистически значимой.
- *Двунаправленное исключение* представляет собой комбинацию первых двух методов, которые проверяют, какие переменные следует включить или исключить.

В ППКПЦ используется регрессия с двунаправленным отбором переменных на основе статистики Фишера F: переменная с наименьшей вероятностью $p(F)$ включается в регрессию, если $p(F)$ меньше заданного порога f_{in}, переменная исключается, если для нее $p(F)$ больше заданного порога f_{out}. Процесс регрессии с пошаговым отбором переменных является мощным инструментов в руках статистиков, однако, у него много критиков и даже есть призывы вообще отказаться от использования этого метода. Статистики отмечают несколько недостатков этого подхода, в том числе неверные результаты и необходимость значительных вычислительных мощностей для разработки сложных регрессионных моделей. Поэтому важно следить за стабильностью получаемых результатов, а также разрабатывать программное обеспечение с учетом необходимости распараллеливания вычислений. Несмотря на эти недостатки, метод отбора переменных (stepwise regression) показывает на используемых данных лучшую точность чем другие методы регуляризации (гребневая регрессия и L_1-регуляризация), поэтому он использован в составе ППКПЦ.

Использование всей кривой зависимости площади «обкатки» от радиуса газа-пенетранта в процессе аппроксимации неэффективно ввиду огромного числа порождаемых этими кривыми переменных. Помимо этого, при использовании метода пошагового отбора переменных при построении регрессий, появляются два пороговых значения f_{in} и f_{out}, которые также необходимо варьировать для получения наиболее стабильного и точного решения. Поэтому при реализа-

ции метода необходимо провести выбор оптимального диапазона линеаризации кривых и значений порогов f_{in} и f_{out}.

Подход к обучению регрессии представлен в виде алгоритма 3. На первом шаге производится загрузка кривых зависимости геометрических индексов от радиуса «обкатки» R. На втором этапе производится линеаризация кривых зависимости геометрических индексов от R на ряде диапазонов $[R^-, R^+]$. Таким образом, на каждом диапазоне линеаризации каждому индексу из таблицы 1) будет соответствовать два новых дескриптора полимера, например, для ASA: «начальная» площадь поверхности c_{ASA} (полимер) и наклон площади поверхности d_{ASA} (полимер). Третий шаг посвящен загрузке экспериментальных данных и их разбиению на выборки. Вначале производится разбиение данных на тестовую и обучающую выборки, а затем разбиение обучающей выборки на 5 подвыборок для проведения процедуры кросс-валидации. На четвертом этапе выбор оптимального диапазона $[R^-, R^+]$ линеаризации кривых ASA, ..., $PNSA_3$ и оптимальных параметров шаговой регрессии f_{in}, f_{out} производится 5-блочной кросс-валидацией; а именно, обучающая выборка случайным образом разбивается на пять частей и для каждой комбинации R^- и R^+ в диапазоне от 0 до 3 Å при условии $R^- < R^+$, а также значений f_{in} и f_{out} в диапазоне от 0.01 до 0.1, при условии $f_{in} < f_{out}$, и для каждой такой комбинации строится пять регрессий с прямым пошаговым отбором переменных на основе обучающей выборки с исключенным первым, вторым, ..., пятым блоком.

На пятом этапе по исключенным блокам вычисляется коэффициент корреляции R. Далее вычисляется средний коэффициент корреляции по исключенным блокам. В итоге выбираются несколько комбинаций параметров R^-, R^+ и f_{in}, f_{out} дающих максимальные средние коэффициенты корреляции. Далее на седьмом этапе для каждого такого набора вычисляется регрессия на полной обучающей выборке и выбирается набор объясняющих переменных, показывающий стабильный и наиболее точный результат. Таким образом получается

финальная регрессия с отобранными объясняющими переменными на фиксированном оптимальным диапазоне $[R^-,\ R^+]$.

Входные данные: $(ASA(R),..., DPSA_3(R))$, база данных транспортных характеристик

Выходные данные: $[R^-,\ R^+]$, $(f_{in},\ f_{out})$, коэффициенты регрессионной модели

цикл *$[R_i^-,\ R_i^+]$ в диапазоне от 0 до 3* Å ***при условии*** $R_i^- < R_i^+$ **выполнять**
  Вычисление коэффициентов c_I и d_I из формулы (3).
конец
Разбиение данных на обучающую и валидационную выборки
Разбиение обучающей выборки на 5 подвыборок
цикл *$[R_i^-,\ R_i^+]$ в диапазоне от 0 до 3* Å ***при условии*** $R_i^- < R_i^+$ **выполнять**
  цикл *$(f_{in}^j,\ f_{out}^j)$ в диапазоне от 0.01 до 0.1* ***при условии*** $f_{in}^j < f_{out}^j$ **выполнять**
    цикл *каждой k-ой подвыборки из 5 подвыборок* **выполнять**
      Построение регрессии с двунаправленным отбором переменных с исключенной k-ой подвыборкой
      Вычисление R^2
    конец
    Вычисление среднего $R^2_{mean_k}$ по исключенным блокам
  конец
конец
Выбор $[R^-_{opt},\ R^+_{opt}]$ и $(f_{in}^{opt},\ f_{out}^{opt})$: $R^2 = \max_{k=1,...,L} R^2_{mean_k}$
Построение регрессии с параметрами $[R^-_{opt},\ R^+_{opt}]$ и $(f_{in}^{opt},\ f_{out}^{opt})$ на обучающей выборке

Алгоритм 3– Алгоритм предсказания транспортных характеристик полимерных мембран.

3 Комплекс программ для предсказания транспортных характеристик полимерных газоразделительных мембран

В главе описывается комплекс программ для предсказания характеристик полимерных газоразделительных мембран (см. Свидетельство о государственной регистрации программы для ЭВМ в Приложении 2). В начале главы описывается структура автономных блоков, составляющих комплекс программ: блок интерфейсов данных (1), блок построения конформаций (2), блок вычисления параметров молекул (3.1) и геометрических индексов (3.2) и блок построения регрессионных моделей (4) и моделей кластеризации (5). Затем предлагается два типовых сценария использования комплекса программ, условно называемые исследовательским и пользовательским.

При разработке комплекса программ для реализации метода ППКПЦ на основе описанных в предыдущих разделах математических моделей и алгоритмов преследовалось несколько целей. Комплекс программ должен:

1. иметь блочную архитектуру, позволяющую использовать блоки автономно,
2. обеспечивать стабильность получаемых результатов и их воспроизводимость,
3. иметь возможность распараллеливания на кластере и приемлемое время расчета одного полимера,
4. быть применимым для специфических полимеров, используемых в мембранном газоразделении,
5. иметь возможность автоматизации,
6. использовать свободно распространяемое ПО.

Разработанный комплекс программ является удобным, за счет реализации в

среде Python, универсальным, за счет блочной архитектуры, быстрым, за счет использования технологий параллельных вычислений. На рисунке 15 представлена функциональная структура разработанного комплекса программ. Ниже подробно рассмотрена реализация и детали каждого из перечисленных блоков.

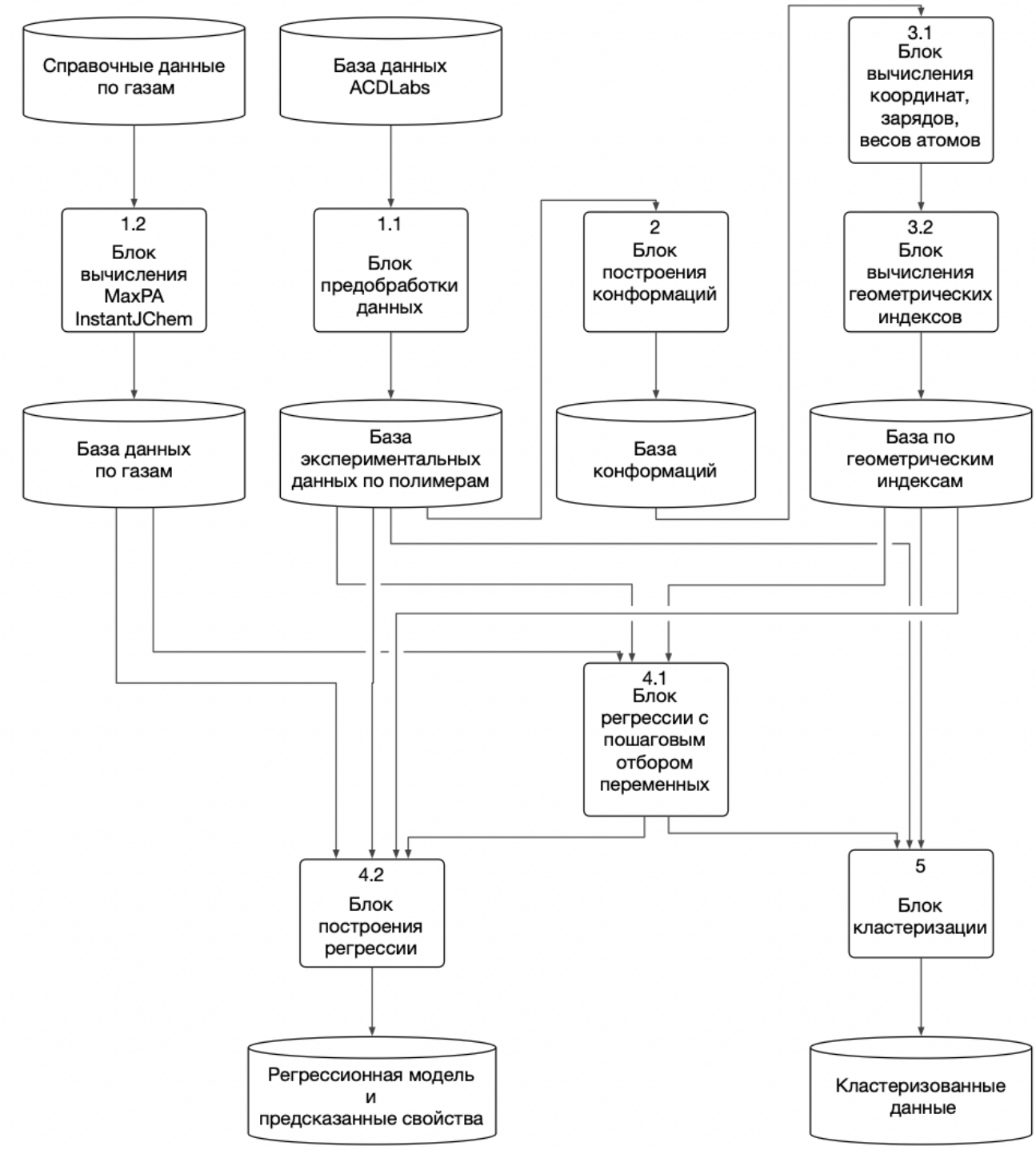


Рисунок 15 – Функциональная структура разработанного комплекса программ.

3.1 Составляющие комплекса программ

3.1.1 Блок интерфейсов данных

Изначально, идеей было создание и настройка единого программного продукта(среды) для хранения, моделирования и обработки данных. Для этих целей отлично подходил продукт Instant JChem компании ChemAxon [60]. Instant JChem поддерживает широкий спектр реляционных баз данных как локально (Derby входит в комплект), так и удаленно (Oracle, MySQL, PostgreSQL и т. д.). Подключения к базе данных можно легко создавать и администрировать, используя данные существующего хранилища данных или импортируя собственные данные для его заполнения. Химические структуры могут быть приведены к стандартному виду, и есть возможность добавлять дополнительные поля с различными физико-химическими свойствами. Дополнительные поля могут быть различных форматов: текстовые, числовые или вычисляемые. За счет использования вычисляемых полей возможно довольно быстро для всей базы вычислить молекулярные массы, различные топологические индексы и многое другое.

Отдельно стоит отметить возможность пользовательской настройки базы данных в соответствии со своими предпочтениями и создание программируемых кнопок с использованием специальных скриптов в среде Groovy Script. В работе [44] Instant JChem использовался в качестве одного из основных инструментов для организации базы данных и проведения моделирования. На рисунке 16 приведен снимок экрана разработанного пользовательского интерфейса в Instant JChem. Опишем принцип работы специально разработанной базы данных для реализации метода ППКПЦ.

1. Новые структуры и экспериментальные данные импортируются в таблицу полимеров (рисунок 17), при необходимости редактируются и вносятся дополнительные данные.

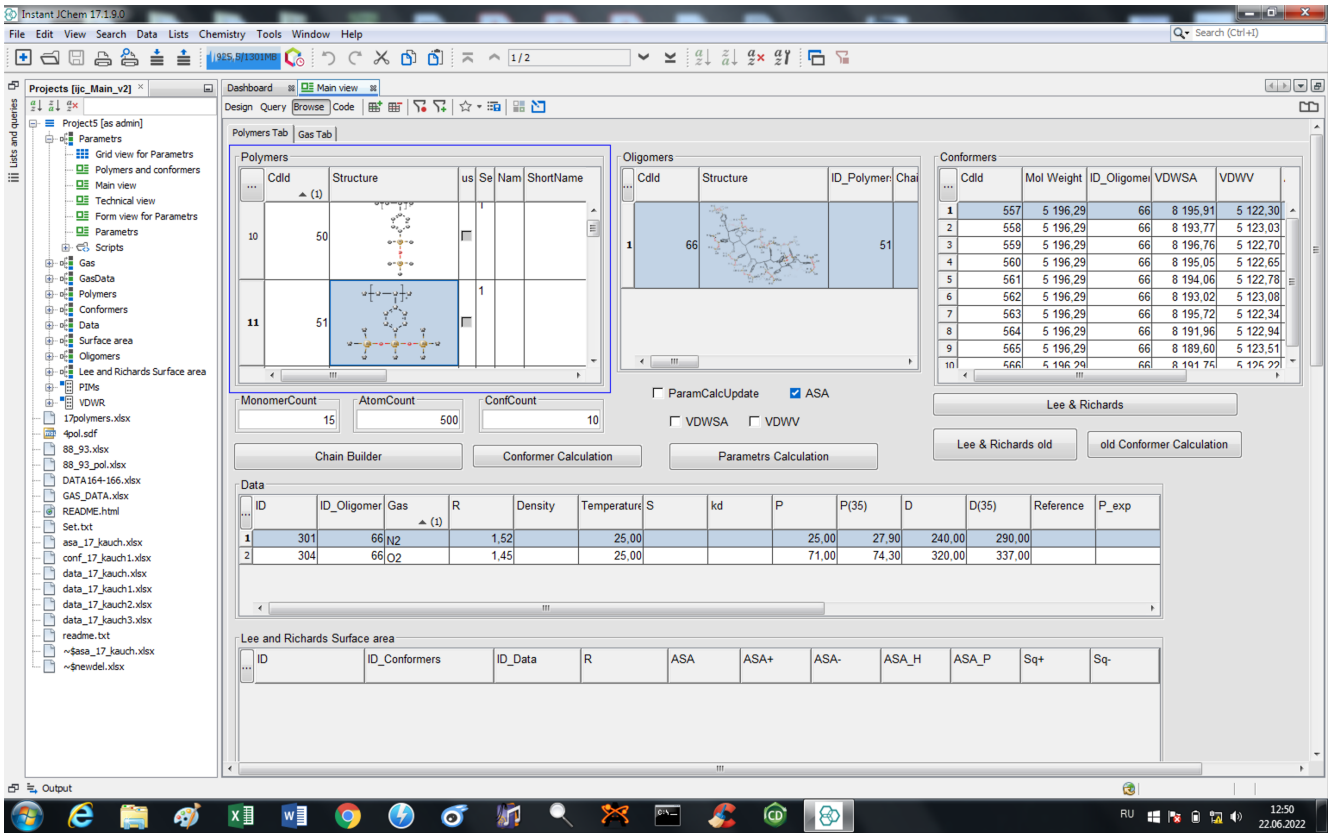


Рисунок 16 – Среда моделирования Instant JChem.

2. Задаются параметры построения олигомерной цепи: число мономерных звеньев или максимальное число атомов.
3. С помощью программируемой кнопки производится построение олигомерной цепи, согласно заданным параметрам.
4. Задается число конформаций, и с помощью программируемой кнопки производится их построение.
5. С помощью программируемой кнопки производится расчет геометрических индексов на основе алгоритма Ли-Ричардса, который был запрограммирован дополнительно, так как встроенный построитель конформаций не подходил для специфических полимеров, используемых в мембранном газоразделении, и выдавал нереалистичные конформации.
6. Полученные геометрические индексы экспортируются. Для построения регрессионных моделей, используется IBM SPSS Statistics.
7. Для конечного пользователя (например, химика, занимающимся синтезом

полимеров) разработанная регрессия переносится в вычисляемое поле в базе.

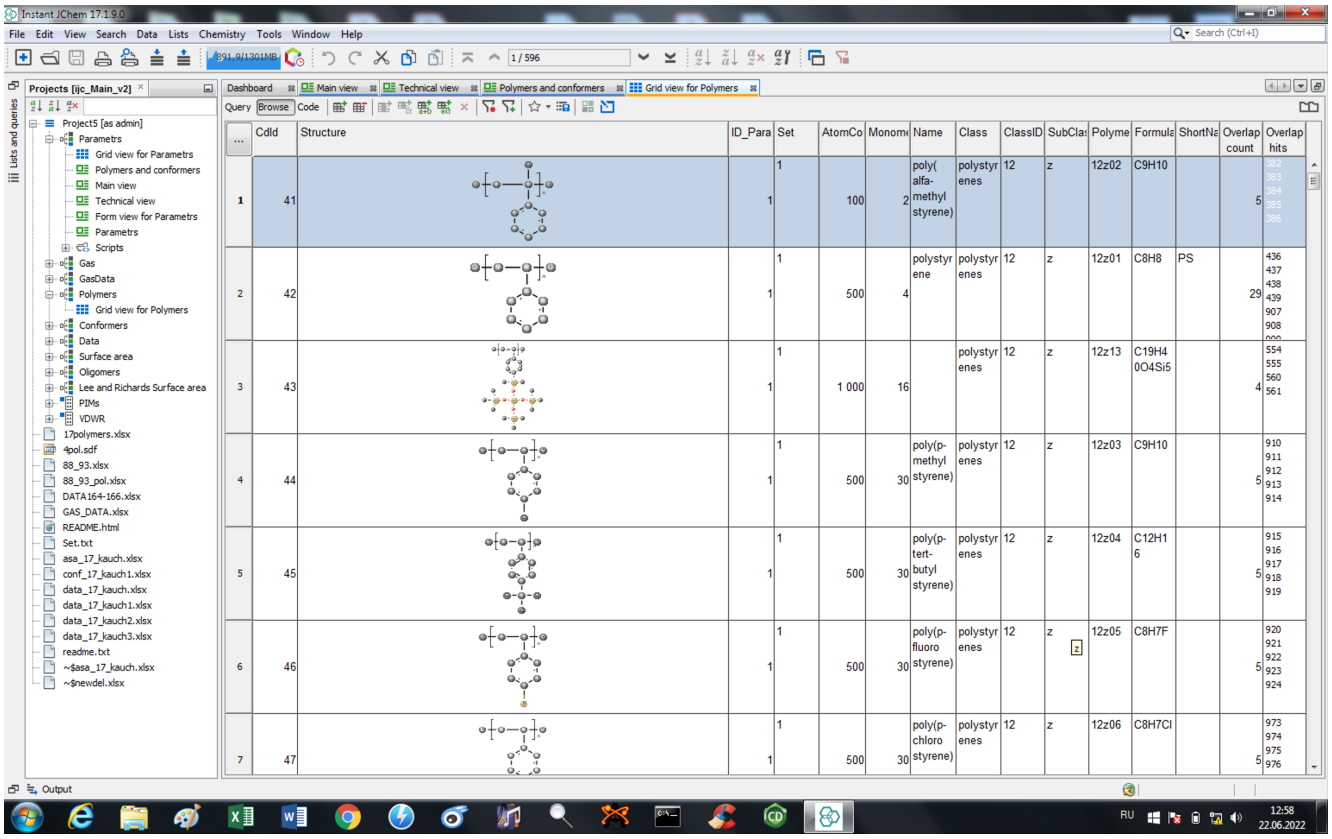


Рисунок 17 – Таблица полимеров в среде моделирования Instant JChem.

Разработанный конвейер обработки данных позволял конечному пользователю получить предсказания транспортных характеристик полимеров, используя всего один программный продукт для хранения и пополнения базы данных и для проведения моделирования. В процессе работы, несмотря на красивый интерфейс и удобство использования для конечного пользователя, были выявлены критические минусы подобного подхода. Используемые скрипты написаны на языке Groovy Script, который является специфическим и в данной реализации не позволяет, например, использовать параллельные вычисления, что является критичным для функционирования комплекса программ. Также в нем отсутствуют многие специализированные пакеты программного обеспечения, например статистические. Поэтому полная автоматизация и реализация исследовательского сценария использования затруднены или скорее невозмож-

ны. Еще одним минусом является зависание интерфейса при работе с большими данными. Помимо этого, Instant JChem является коммерческим продуктом, требующим покупки лицензии, хотя стоит отметить наличие академической лицензии, предлагаемой компанией за цитирование в публикациях в международных журналах.

Для удовлетворения всех требований к итоговому комплексу программ, описанных в начале главы, был осуществлен переход в среду программирования Python с широким использованием возможностей программного пакета RDKit. Пакет RDKit [86] – набор инструментов с открытым исходным кодом – широко используется научным сообществом для решения различных задач в области хемоинформатики и машинного обучения. Основные структуры данных и алгоритмы RDKit написаны на C++, что обеспечивает высокое быстродействие. RDKit также имеет оболочки на Python, Java и C#, что делает его удобным в использовании.

Для осуществления хранения и передачи структур полимеров между Базой данных и разработанным комплексом программ используется формат файлов SDF, описанный в разделе 1.3.1). В RDKit для чтения входного файла в формате SDF используется библиотека PandasTools, переводящая данные в табличный вид и позволяющая взаимодействовать как со структурами полимеров, так и с экспериментальными, техническими и прочими данными.

Согласно рисунку 18, на вход блока интерфейсов данных подаются исходные данные (например, из базы ACDLabs) в формате SDF. Полученные данные преобразуются в подходящий формат для использования в библиотеке PandasTools. Символы «*», означающие границы мономерного звена полимера(мономера), заменяются на специальные символы для того, чтобы можно было различить левую и правую границу мономера. Затем, преобразованный файл в формате SDF разбивается на отдельные файлы SDF со структурой и информацией по каждому полимеру. Таким образом на выходе блока получается множество файлов SDF со структурой и информацией по каждому полимеру.

База данных для популярных газов сформирована в формате XLSX, в ней представлены называния газов, их молекулярный вес, радиус Ван-дер-Ваальса и максимальная площадь проекции газа (MaxPA). Однако, при необходимости добавить новый газ нужно вычислить максимальную площадь проекции газа (MaxPA) путем создания или загрузки его структуры в формате SDF в MarvinSketch. Далее, используется процедура расположения молекулы в пространстве и вычисляются геометрические дескрипторы, в том числе и MaxPA. Вычисленная MaxPA добавляется в базу данных газов в формате XLSX.

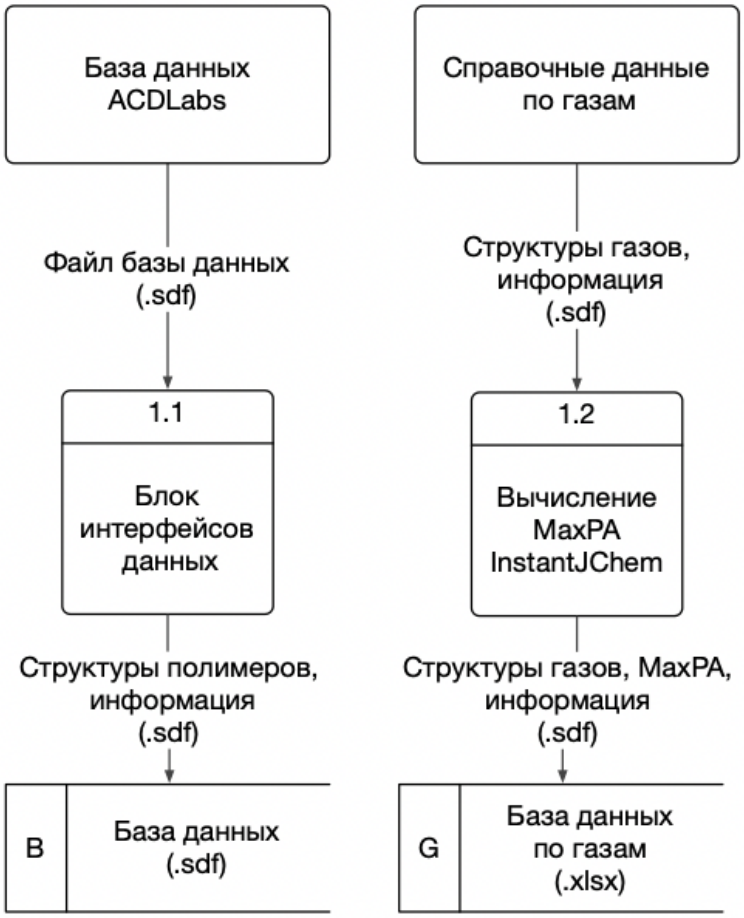


Рисунок 18 – Диаграмма потоков данных блока интерфейсов данных.

3.1.2 Блок построения конформаций молекул

В певых версиях метода ППКПЦ для каждого полимера в среде молекулярного моделирования InstantJChem [60] создавался олигомер длиной по-

рядка 200-600 атомов, состоящий из нескольких мономерных звеньев полимера. Для этой цепочки средствами Conformer Plugin [29] пакета InstantJChem ChemAxon генерировались 6 случайных конформаций различной геометрии, получаемых с помощью оптимизации в эмпирическом силовом поле Дрейдинга [70] из различных начальных позиций атомов. Дальнейший расчет необходимых для построения регрессионной модели индексов также производился средствами InstantJChem. Несмотря на то, что с помощью полученных конформаций был довольно успешно предсказан один из транспортных параметров полимерных мембран – растворимость S, данный подход не соответствовал многим критериям. Самым важным недостатком является нереалистичность многих полученных конформаций, а также сложность в расширении данного решения на серверные платформы.

В работе [43] решена проблема нереалистичности полученных конформаций, за счет проведения их моделирования в программе в пакете PerkinElmer Chem3D версии 15.1.0.144 [26] по аналогичной методике. Для задания случайного начального положения атомов к моделируемой полимерной цепи проводится молекулярно-динамическое моделирование при температуре от 300 K до 3000 K в течение 1000 итераций, после чего полученная структура оптимизируется по свободной энергии в эмпирическом поле MM2. Данный подход показывает неплохие результаты в плане реалистичности конформаций, но лишает возможности автоматизации процесса молекулярно-механического моделирования и, тем более, не позволяет использовать серверные мощности для распараллеливания вычислений.

В процессе поиска было опробовано несколько различных решений молекулярно-механического и молекулярно-динамического моделирования, например, LAMMPS [65], OpenMM [31], Hoomd-Blue [16, 42] и другие. Некоторые решение не позволяли автоматизировать процесс (LAMMPS), другие (OpenMM, Hoomd-Blue) были разработаны скорее для биополимеров и не име-

ли готовых силовых полей для работы со стеклообразными полимерами из мембранного газоразделения.

В работе для расположения молекулы в 3D пространстве используется гибкая процедура EmbedMolecule [32, 89, 90], интегрированная в среду RDKit. Начальные координаты атомов молекулы могут быть определены как через собственные значения матрицы расстояний, так и через случайное расположение атомов, причем значение генератора случайных чисел может быть зафиксировано, что позволит позже получить такую же молекулу. Затем производится молекулярно-механическое моделирование в эмпирическом силовом поле.

На диаграмме на рисунке 20 блок построения конформаций молекул представлен под цифрой (2). На вход блока подается структура полимерного звена в формате SDF, а также параметры, такие как: число мономеров, силовое поле (MMFF94, MMFF94s, UFF) и другие технические параметры. Согласно методу, описанному в подразделе 2.2.1, с помощью процедуры EmbedMolecule производится расположение в пространстве атомов мономерного звена полимера, к нему добавляется аналогично обработанное звено, но с другими случайными координатами. Все координаты атомов, кроме координат атомов последних двух добавленных мономерных звеньев, фиксируются, а два свободных звена проходят процедуру минимизации энергии в силовом поле MMFF94 методом градиентного спуска. На выходе данного блока идет SDF файл с рассчитанными координатами конформации молекулы полимера. Примеры полученных с помощью данной процедуры конформаций представлены на рисунке 12. Таким образом на выходе блока образуется база конформаций молекул.

Вычисления могут быть распараллеленны по конформациям с помощью встроенной в Python библиотеки multiprocessing, что позволяет одновременно вычислять несколько конформаций различных полимеров. Это необходимо, так как при увеличении числа атомов в молекуле время расчета растет нелинейно и проводить расчеты на размерах молекул более 800 атомов становится затруднительно, что является ограничением данного метода.

3.1.3 Блок вычисления геометрических индексов и параметров молекул

Вычисление *ASA* является обычным расчетом в структурной биологии, поэтому существуют готовые библиотеки для этих целей. FreeSASA [37, 77] – это библиотека C с открытым исходным кодом для вычислений *ASA*, которая предоставляет интерфейсы командной строки и Python в дополнение к API C. Библиотека реализует приближения как Ли-Ричардса, так и Шрейка-Рипли, и имеет широкие возможности настройки, позволяя пользователю контролировать молекулярные параметры, точность и степень детализации.

FreeSASA эффективна для расчета *ASA* структур белков, РНК и ДНК, однако исследуемые нами полимеры, используемые в мембранном газоразделении, являются в основном стеклообразными, поэтому в данной работе используются универсальные значения торсионных углов, зарядов и прочих характеристик. Также FreeSASA использует файлы формата PDB, а в разделе 1.3.1 были описаны причины выбора файлов формата SDF. Поэтому для вычисления *ASA* модифицированный алгоритм Ли-Ричардса был реализован на языке Python.

На диаграмме на рисунке 20 блок вычисления геометрических индексов и параметров молекул представлен под цифрой (3) и разбит на две части. На вход блока вычисления параметров молекул (3.1) подаются координаты атомов(формат SDF). Функциональное назначение блока (3.1) – вычисление атомных номера, радиусов Ван-дер-Ваальса и частичных зарядов атомов по Гастайгеру–Марсели с использованием встроенных в RDKit функций (например, частичные заряды вычислялись процедурой rdPartialCharges библиотеки rdkit.Chem). Вычисленные параметры подаются на вход блока вычисления геометрических индексов (3.2), где с использованием алгоритма Ли-Ричардса вычисляются для каждого атома площади доступные для «обкатки» газом-пенетрантом, и, следом, производится расчет индексов из таблицы 1 (ASA, ASA^+ и т. д.). На выходе блока (3) получаем три файла формата HDF5:

1. с координатами атомов молекулы в трехмерном пространстве, значениями радиусов, частичных зарядов и весов атомов;
2. с рассчитанными значениями геометрических индексов для каждого атома в зависимости от радиуса «обкатки»;
3. с рассчитанными значениями суммарных геометрических индексов в зависимости от радиуса «обкатки».

Все этапы вычисления геометрических индексов эффективно распараллеленны по конформациям молекул, что позволяет массово вычислять геометрические индексы для множества полимеров и их конформаций.

3.1.4 Блок регрессионного анализа

В подразделах 2.1.3 и 2.2.3 описывается принцип построения регрессионной модели и алгоритм отбора объясняющих переменных. На диаграмме на рисунке 20 блок регрессионного анализа представлен под цифрой (4) и разбит на две части. На вход блока регрессионного анализа подаются значения рассчитанных геометрических индексов в формате HDF5, а также экспериментальные значения транспортных характеристик и метки разбиения на тестовую и обучающую выборки в формате XLSX.

В связи с большим числом объясняющих переменных при относительно небольшом объеме выборки для предотвращения переобучения используется процедура `stepwise_selection` (регрессии с шаговым отбором переменных) из [101], реализующая алгоритм регрессии с двунаправленным отбором переменных [33] на основе статистики Фишера F. Первая часть блока (4.1), за счет использования процедуры 5-блочной кросс-валидации, описанной в разделе 2.11, позволяет получить список значимых переменных и оптимальных значений диапазона эффективных радиусов и порогов вхождения и исключения переменных в финальную регрессию, что используется для построения регрессии во второй части блока (4.2).

На выходе данного блока получаем список отобранных объясняющих переменных с коэффициентами, а также метрики на обучающей и тестовой выборках: коэффициент детерминации R^2 и средняя относительная ошибка MRE. Также, с помощью алгоритма `AgglomerativeClustering` пакета `scikit-learn` [11] на основе полученных значимых переменных проводится кластеризация полимеров в блоке (5) на диаграмме на рисунке 20.

При реализации данных блоков (4) и (5) на языке Python были использованы такие программные библиотеки, как: numpy, pandas, scipy, sklearn и statsmodels.

3.2 Сценарии использования комплекса программ

Разработанный комплекс программ рассчитан на два сценария использования: исследовательский и пользовательский. Исследовательский сценарий позволяет расширять обучающее множество за счет новых полимеров, вычислять свои индексы и вносить изменения в уже разработанные, изменять параметры построения регрессионных моделей и строить свои модели на основе произведенных расчетов.

Пользовательский сценарий предлагает использовать разработанные и обученные модели для предсказания транспортных характеристик новых, еще не синтезированных полимеров, а также использовать предложенную в разделе 4.5 кластеризацию для оценки и сравнения значений транспортных характеристик полимеров. В качестве пользователя могут выступать химики, которым необходимо, например, проверить ряд перспективных полимеров и предсказать для них транспортные характеристики без необходимости синтезировать данные полимеры. Согласно диаграмме потоков данных для пользовательского интерфейса (рисунок 19), пользователь должен на вход блока интерфейсов данных (1) подать SDF файл с одной или несколькими структурами мономерных звеньев полимеров. Затем, предобработанные данные подаются на вход блока построения конформаций (2), в котором для каждого из уникальных полимеров стро-

ится 6 конформаций (подробнее в подразделе 3.2). После, в блоке вычисления параметров и геометрических индексов, происходит вычисление необходимых индексов. Затем в блоке регрессионного анализа с использованием разработанных регрессий происходит предсказание транспортных характеристик.

Исследовательский сценарий используется для разработки новых регрессионных моделей на основе пользовательских экспериментальных данных. Данный сценарий использования может быть представлен диаграммой на рисунке 20. Функциональность исследовательского сценария совпадает с полной функциональностью вышеописанных блоков комплекса программ. Основными преимуществами данного режима является:

- возможность добавления новых газов;
- изменение числа конформаций, силового поля и числа фиксируемых мономеров в процессе построения конформаций;
- добавление новых геометрических индексов;
- использование собственных обучающих экспериментальных данных для построения новых регрессионных моделей и моделей кластеризации.

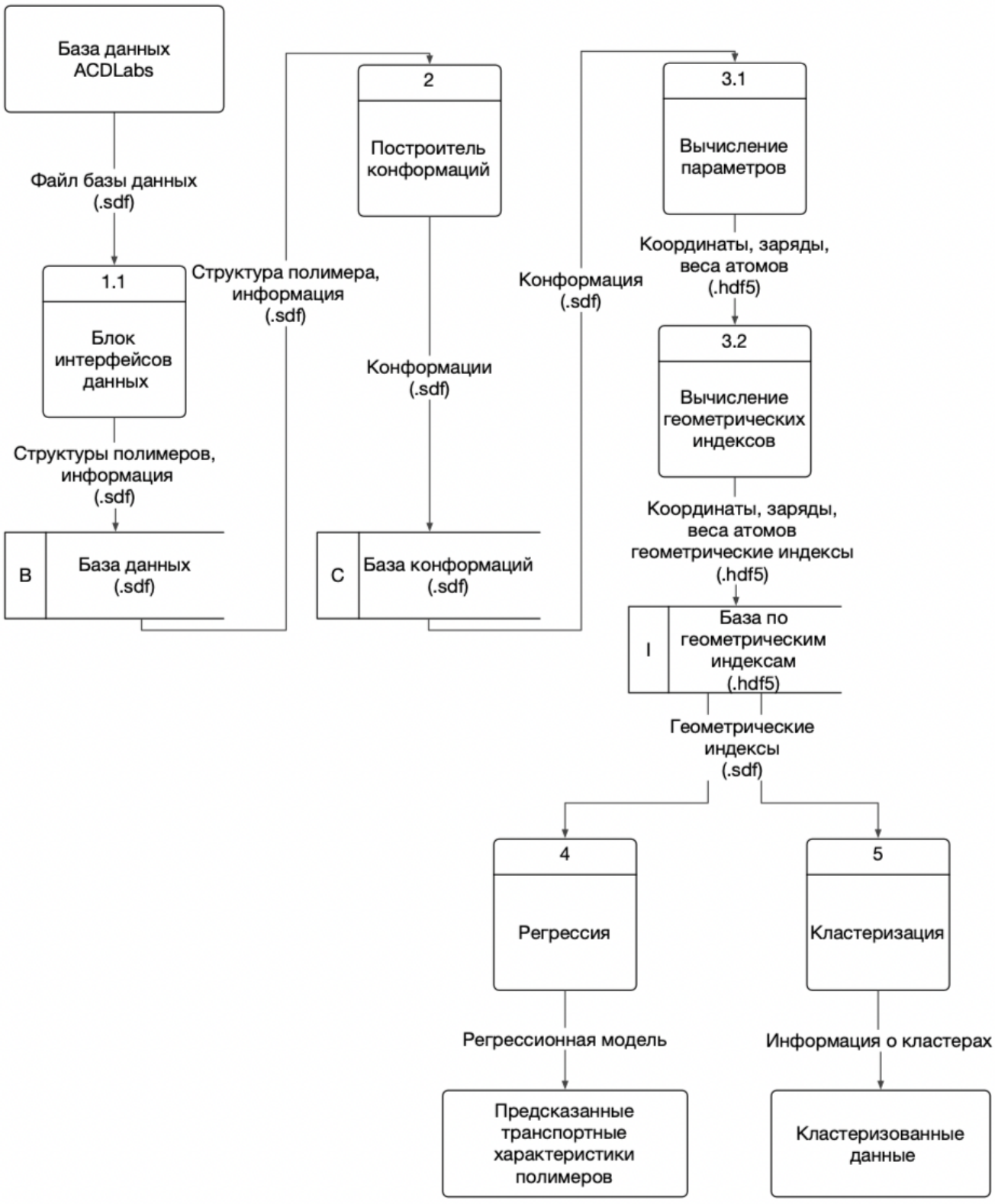


Рисунок 19 – Диаграмма потоков данных пользовательского сценария.

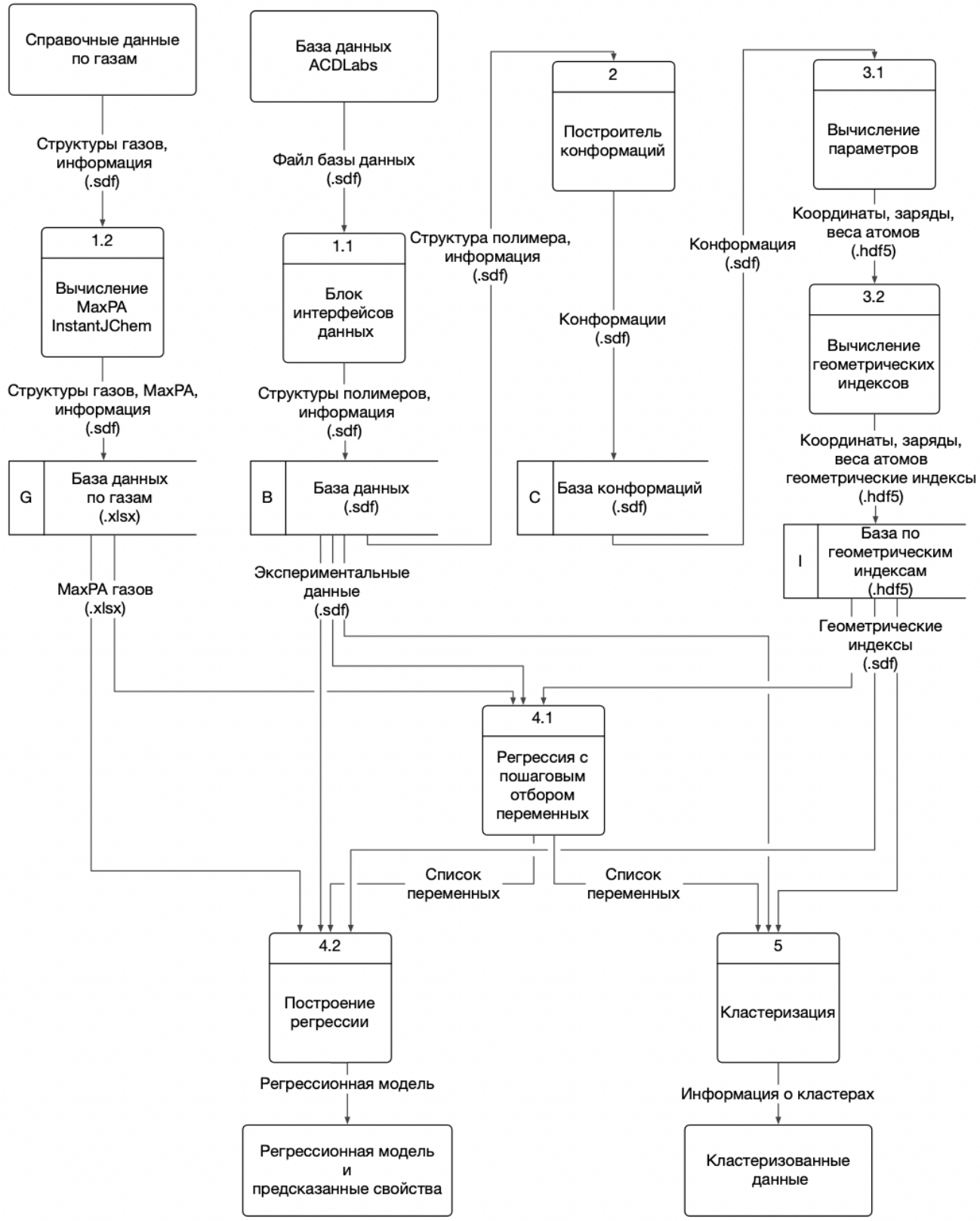


Рисунок 20 – Диаграмма потоков данных исследовательского сценария.

4 Прикладные задачи анализа и синтеза полимерных материалов в интересах мембранного газоразделения

Прикладные исследования в рамках диссертационной работы велись в интересах Лаборатории Мембранного газоразделения Института нефтехимического синтеза им. А.В.Топчиева РАН (ИНХС РАН). Основными направлениями исследований Лаборатории Мембранного газоразделения являются:

- изучение транспортных характеристик полимеров различной химической структуры и физического состояния с целью поиска оптимальных материалов для мембранного газо- и пароразделения, первапорации;
- изучение связи транспортных характеристик полимеров и мембран с их физико-химическим свойствами;
- изучение термодинамических свойств мембранных материалов;
- поиск решения различных практических задач с использованием мембранного газоразделения.

Разработанная и поддерживаемая коллегами из лаборатории Мембранного газоразделения уникальная база данных «Газоразделительные параметры стеклообразных полимеров» (далее – База данных), подробное описание и анализ которой будут представлены ниже, позволяет проводить исследования, с использованием статистики и методов машинного обучения. Поэтому процесс разработки, обучения и тестирования метода ППКПЦ велся с использованием экспериментальных данных из Базы данных.

В главе описаны прикладные проекты, осуществленные с использованием разработанного метода ППКПЦ и реализованного комплекса программ. В начале главы приводится подробное описание Базы данных, предоставленной коллегами из ИНХС РАН. После предлагается процедура обоснования достаточной длины цепочки и количества генерируемых конформаций. Затем описаны

примеры использования разработанных методов для предсказания таких транспортных параметров газоразделительных мембран, как коэффициент растворимости S и константа Генри k_D.

Работы проводились в рамках проектов, перечисленных ниже (см. также Акт о внедрении в Приложении 1).

- Гранты РФФИ:
  - А 17-08-00164 Компьютерное моделирование наноструктуры мембранных материалов: традиционные и новые подходы.
  - мол_а 18-37-00265 Классификация конформационных структур аморфных полимеров в интересах мембранной технологии.
  - Аспиранты 19-37-90004 Математические модели и алгоритмы поиска веществ с заданными физико-химическими свойствами.
- Госзадания фундаментальных исследований ИНХС РАН:
  - 2017 «Мембранное разделение и мембранный катализ в химии, энергетике, экологии: новые мембранные материалы, высокопроизводительные мембраны и процессы на их основе»; шифр 45, 47; Госуд. рег. № АААА-А18-118011990199-9; чл.-к. РАН А.Б. Ярославцев, проф. В.В. Волков.
  - 2018-2019 «Новые материалы и высокопроизводительные мембраны для разделения жидких и газовых смесей; мембранный катализ в химических процессах получения продуктов высокой чистоты»; шифр 45, 47; Госуд. рег. № АААА-А19-119020490055-4.

4.1 База данных «Газоразделительные параметры стеклообразных полимеров»

Как и в любой задаче машинного обучения, правильно собранные и предобработанные данные играют ключевую роль в получении качественного резуль-

тата. В описываемых ниже исследованиях в качестве исходной базы использовалась База данных «Газоразделительные параметры стеклообразных полимеров», которая была создана в лаборатории мембранного газоразделения ИНХС им. Топчиева РАН в 1998 г. База данных является уникальным в своем роде справочником и инструментом прогнозирования газоразделительных параметров полимеров. В мировой практике уже давно обрели популярность базы данных химических структур низкомолекулярных веществ. В последнее время с развитием полимерной промышленности также стали появляться базы данных по основным физико-химическим свойствам полимеров [83–85] и др., некоторые из которых также используют для прогнозирования свойств полимеров. Однако они направлены на применение полимеров в качестве конструкционных материалов, тогда как применение полимеров в качестве материалов мембранного газоразделения остается за кадром. Общее число полимеров в Базе данных превышает 1600. На сегодняшний день представленная База данных не имеет аналогов в мире по объему информации по газоразделительным характеристикам полимерных материалов.

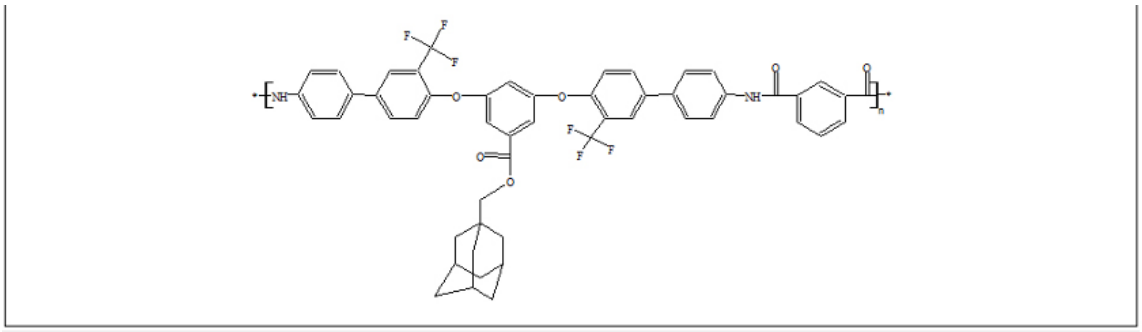
База данных хранится и пополняется в программе ACD Labs. Каждая запись в Базе данных представляется химической структурой в 2D виде и несколькими текстовыми полями от технических до тех, что описывают все известные транспортные характеристики для пары «газ-полимер». На рисунке 21 представлен пример одной записи. Информационные поля в базе можно разбить на 4 группы: технические, описательные, экспериментальные и вычисляемые.

Описательные поля – поля, описывающие полимер и газ:

- Class – класс полимера определяется исходя из его структуры.
- Name – название полимера.
- ShortName – сокращенное название полимера.
- Gas – химическая формула газа.

Технические поля – это данные, необходимые для удобного использования базы, кодирования полимеров и их классов.

- ClassID – идентификационный номер класса.
- SubClassID – идентификатор подкласса, обозначаемый буквой. Классы, внутри которых существуют определенные различия химической структуры.
- PolymerID – идентификационный номер полимера. Пример записи: 23d104 (ClassID, SubClassID, порядковый номер).



Formula: ?·$[C_{52}H_{40}F_6N_2O_6]_n$
FW: $[902.8743]_n$+?
Class: polyamides
ClassID: 21
SubClassID: b
PolymerID: 21b30
Short Name: PA 9C
Reference: Debaditya Bera,Parthasarathi Bandyopadhyay,Sipra Ghosh,Susanta Banerjee, Journal of Membrane Science, 453 (2014), 175-191
Tg, C: 219
Density, g/cm3: 1.34
Texp, C: 35
Pexp, atm: 3.5
Gas: CO2
P, Ba: 13.0
P(35C), Ba: 13.0
Epexp, kJ/mol: 13.6
D*10E8, cm2/s: 2.2
D(35C)*10E8, cm2/s: 2.2
Edexp, kJ/mol: 34.5
comment: Density is measured at 30 oC
Note: Additional information provided on Number average molecular weight, Polydispersity index, FFV, Tensile strength, Young's modulus, dielectric constant
<End of List>

Рисунок 21 – Скриншот одной записи из Базы данных.

Ниже рассмотрим экспериментальные и вычисляемые поля. Экспериментальные данные представляют наибольший интерес в данной базе, так как многие годы они скрупулезно собирались вручную.

- Общая информация об экспериментальных данных:
  - Texp – температура эксперимента, °C,
  - Tg – температура стеклования,
  - Density – плотность полимера,
  - Reference – источник (статья). Заполняется в таком порядке: перечисляются авторы, название журнала, номер выпуска, страницы, год (указывается в скобках).
- Информация по газопроницаемости:
  - *P, Ba* – Коэффициент проницаемости при температуре эксперимента,
  - *P(35C), Ba* – Коэффициент проницаемости при температуре 35°C, рассчитанный,
  - *Epexp, kJ/mol* – Энергия активации проницаемости, измеренная экспериментально,
  - *Epcalc, kJ/mol* – Энергия активации проницаемости, рассчитанная.
- Информация по диффузии:
  - *D10E8, cm2/s* – Коэффициент диффузии при температуре эксперимента,
  - *D(35C)10E8, cm2/s* – Коэффициент диффузии при температуре 35°C, рассчитанный,
  - *Edexp, kJ/mo* – Энергия активации диффузии, измеренная экспериментально,
  - *Edcalc, kJ/mo* – Энергия активации диффузии, рассчитанная.
- Информация по растворимости:
  - *S10^2, cm3(STP)/(cm3cmHg)* – Коэффициент растворимости при температуре эксперимента,

- *S10^2 (35C), cm3(STP)/(cm3cmHg)* – Коэффициент растворимости при температуре 35°C, рассчитанный,
- *Solution Heat (Hs) exp, kJ/mol* – Теплота сорбции, измеренная экспериментально,
- *Solution Heat (Hs) calc, kJ/mol* – Теплота сорбции, рассчитанная,
- *kd, cm3(STP)/(cm3cmHg)* – константа растворимости по закону Генри,
- *CH', cm3(STP)/cm3* – описывает адсорбционную способность полимера по отношению к конкретному газу при определенной температуре $T < Tg$,
- *b, 1/cmHg* – адсорбционный коэффициент, зависящий от энергии адсорбции и температуры.

В связи с тем, что качество исходных данных определяет точность и работоспособность итоговых моделей, был подробно изучен процесс создания Базы данных. Пополнение Базы данных производится в два этапа:

1. поиск литературных источников, которые ранее не были включены в Базу данных
2. анализ данных, представленных в литературных источниках, и внесение информации в Базу данных

Первый этап – поиск научных статей и других источников, в которых были бы представлены новые структуры полимерных материалов и/или данные по газоразделительным характеристикам полимерных материалов. В том числе, рассматриваются обзоры, в которых описаны тенденции последних лет, которые представляли собой весомый источник информации для внесения наиболее актуальных структур в Базу данных. Затем проводится дополнительный поиск по авторам статей, касающихся синтеза и характеризации новых полимерных структур. Таким образом, набирается ряд новых литературных данных для

обновления Базы данных и, соответственно, для повышения точности прогнозирования параметров полимерных материалов при использовании данных из базы в качестве обучающей выборки.

Стоит отметить, что тенденция к созданию исследователями полимерных материалов с улучшенными свойствами приводит к проблемам в отношении распределения данных – смещению данных в сторону более высокопроницаемых материалов. Кроме того, так как исследования материалов мембранного газоразделения возможно проводить только в условиях, когда из полимера формируется пленка достаточной механической прочности, исследователи наиболее активно занимаются синтезом и характеризацией определенных классов полимерных материалов, в частности, полиимидов. Поэтому, в Базе данных можно увидеть существенный сдвиг в сторону полиимидов, что будет показано ниже во время анализа Базы данных.

На втором этапе проводят анализ данных, представленных в каждой статье, и производится внесение этих данных в Базу. Вначале вносится источник данных в поле *Reference*. После этого, вносят описательную инфорамцию о паре «газ-полимер», а также температуру, при которой был проведен эксперимент. В случае если параметры определяли при разных температурах, создают несколько записей для одной пары «газ-полимер» и литературного источника. Таким образом, разделяют свойства, измеренные при одних и при других условиях. После вносятся физическо-химические свойства полимера и информация о газоразделительных параметрах: коэффициенте проницаемости, коэффициенте диффузии, коэффициенте растворимости, энергии активации диффузии, энергии активации проницаемости, теплоте сорбции. В статье могут быть одновременно приведены коэффициент диффузии и коэффициент растворимости. В случае, если данные параметры измерены экспериментально с использованием разных методов, эти параметры приводятся в разных записях Базы данных. Однако чаще имеет место случай, когда один из данных коэффициентов измерен экспериментально, тогда как другой рассчитан по формуле (1.8). В таком

случае в Базе данных приводят только экспериментальные значения, что позволяет избежать ошибок в интерпретации.

Для того чтобы использовать данные для прогнозирования, все газоразделительные параметры приводят к одной температуре. Так как в литературных источниках преобладает температура эксперимента 35°C, то она была выбрана в качестве стандартной. В качестве температурных коэффициентов газоразделительных параметров используют энергию активации проницаемости, энергию активации диффузии и теплоту сорбции. Если эти параметры измерены экспериментально и приведены в работах, то их обязательно вносят в Базу данных и используют при необходимости пересчета газоразделительных параметров к другой температуре. Если же этих данных нет в работе, то проводят расчет этих параметров по схеме, приведенной в работе [1].

Почти все источники данных могут содержать ошибки ручного ввода. В Базу данных вручную вводятся данные из различных литературных источников, которые также могут содержать ошибки. Поэтому, был проведен разведочный анализ Базы данных, по результату которого:

- перепроверены и исправлены выбросы,
- удалены дубликаты,
- устранены некоторые пропуски как технических так и экспериментальных данных.

По распределению по классам полимеров на май 2022, представленному на рисунке 22, видно, что более трети Базы данных составляют полиимиды, порядка 15% – сополимеры (которые не используются в данной работе ввиду особенностей их строения), 11% – полиацетилены, и от 5 до 6 % – полинорборнены, полиамиды, полимеры на основе простых и сложных эфиров. Любой из остальных классов полимеров занимает менее 3% Базы данных. Подобное неравномерное распределение данных может губительно сказаться на разраба-

тываемых моделях, что, по возможности, будет учитываться при построении моделей.

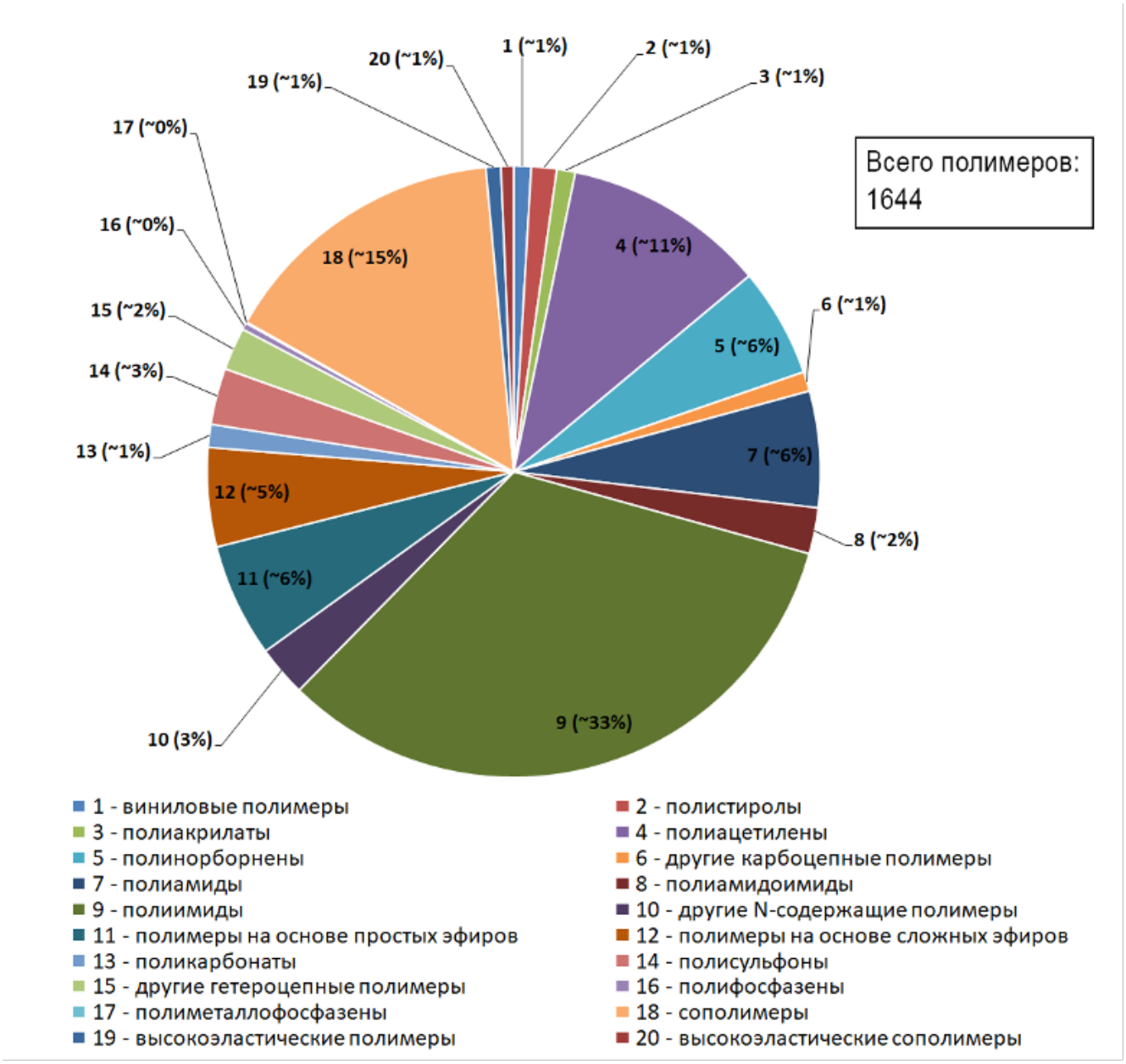


Рисунок 22 – Соотношение в % классов полимеров в Базе данных ИНХС РАН.

Другой сложностью при работе с Базой данных является большое число пропусков экспериментальных данных, например, в случае, когда авторы работ приводят данные только по проницаемости и не измеряют значения коэффициентов диффузии. На момент мая 2022 года в Базе данных сделано 11108 записей (что почти в 1,5 раза больше, чем на момент начала разработки метода ППКПЦ). Из таблицы 3, где приведено число пропусков в данных для полей,

наиболее важных для разработки моделей, видно, что при необходимости предсказать коэффициент растворимости S, нужно воспользоваться формулой (1.8) для его вычисления, и только после этого использовать его в качестве «экспериментального» значения, что негативно повлияет на качество итоговой модели. Также видно, что крайне мало необходимых экспериментальных данных по константе Генри k_D.

Таблица 3 – Демонстрация числа пропусков данных среди наиболее важных полей.

Название поля	Число не пустых полей, шт	Доля непустых полей, %
P, Ba	10732	96.6
P(35C), Ba	7725	69.5
D10E8, cm2/s	5101	45.9
D(35C)10E8, cm2/s	3155	28.4
Epexp, kJ/mol	1157	10.4
Edexp, kJ/mol	787	7.1
S10^2,cm3(STP)/(cm3cmHg)	407	3.7
S10^2 (35C), cm3(STP)/(cm3cmHg)	182	1.6
kd, cm3(STP)/(cm3cmHg)	267	2.4
Density	6579	59.2

Базу данных возможно экспортировать в формате .sdf как одним компактным файлом, так и каждую полимерную структуру отдельным файлом, что позволяет легко импортировать ее в разработанный комплекс программ и не требует дополнительных преобразований форматов данных.

4.2 Обоснование достаточной длины полимерной цепи и количества генерируемых конформаций

Для обоснования достаточной длины полимерной цепи и количества генерируемых конформаций для устойчивости получаемых далее результатов были сгенерированы по 6 конформаций различной длины (200, 300, 400, 600, 800 атомов без учета водородов) для 40 различных полимеров. Для них были рассчитаны индексы из всего набора для ряда значений R из диапазона от 1Å до 3Å. На рисунке 23 в качестве примера представлены полученные зависимости трех индексов (при различных R) от размера молекулы для 16 случайных полимеров из 40 обсчитанных. Кривые построены по усредненным значениям индексов по 6 конформациям для каждого полимера. Из рисунка видно, что значения индексов начинают сходиться на длине молекулы в 600 атомов, которая и была взята в качестве базовой длины.

По аналогии с достаточной длиной молекулы доказывается достаточность усреднения по 6 конформациям для получения устойчивых значений индексов (см. рисунок 24). Для 40 различных полимеров было сгенерировано по 10 конформаций. Для всего ряда значений R из диапазона от 1Å до 3Å были рассчитаны индексы и получены кривые зависимости индексов от радиуса «обкатки». Из рисунка 24 видно, что значения индексов начинают сходиться при усреднении значений по 5-6 конформациям. Поэтому достаточным является генерация 6 конформаций на один полимер.

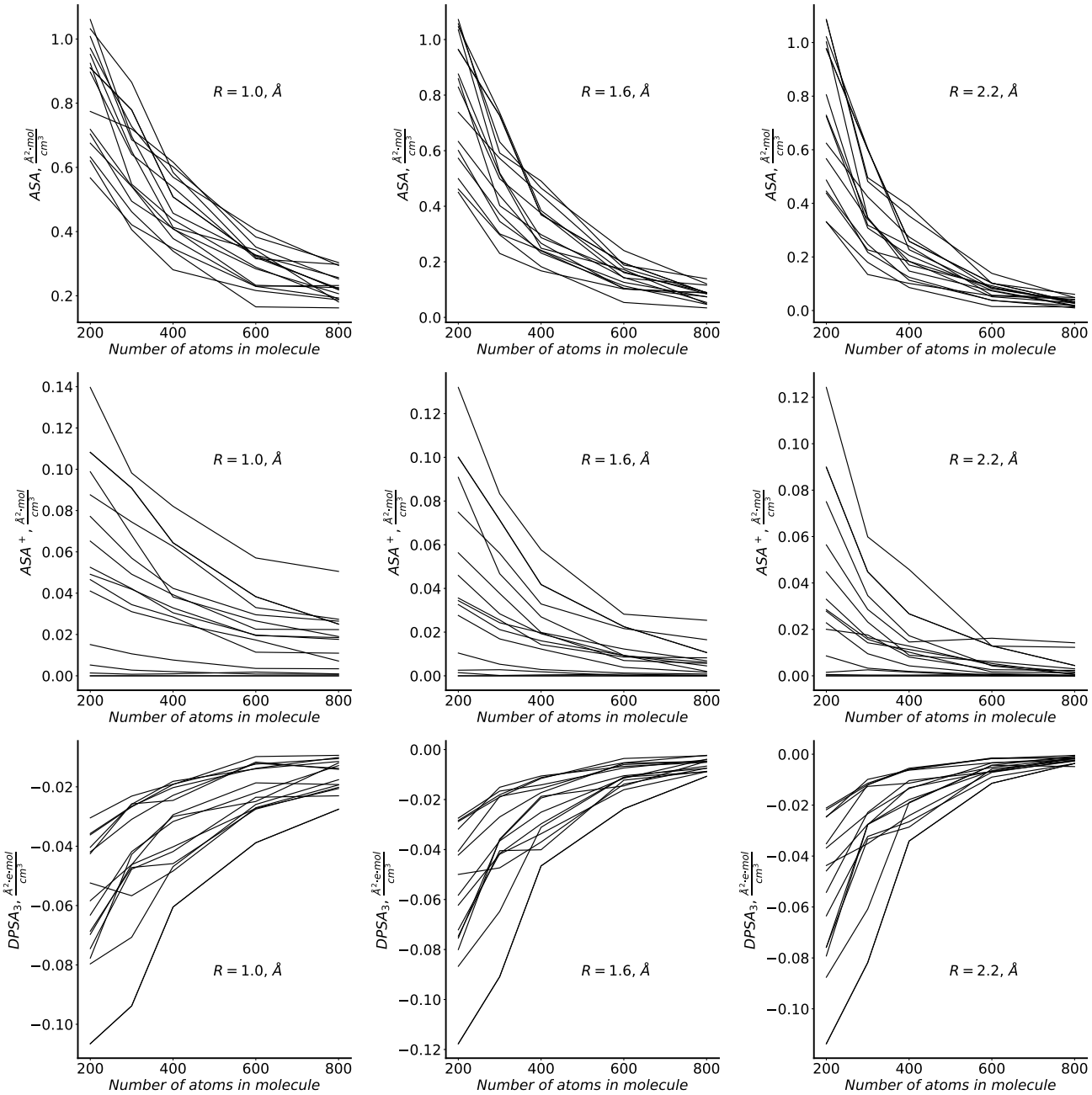


Рисунок 23 – Зависимости трех индексов от размера молекулы для 16 случайных полимеров.

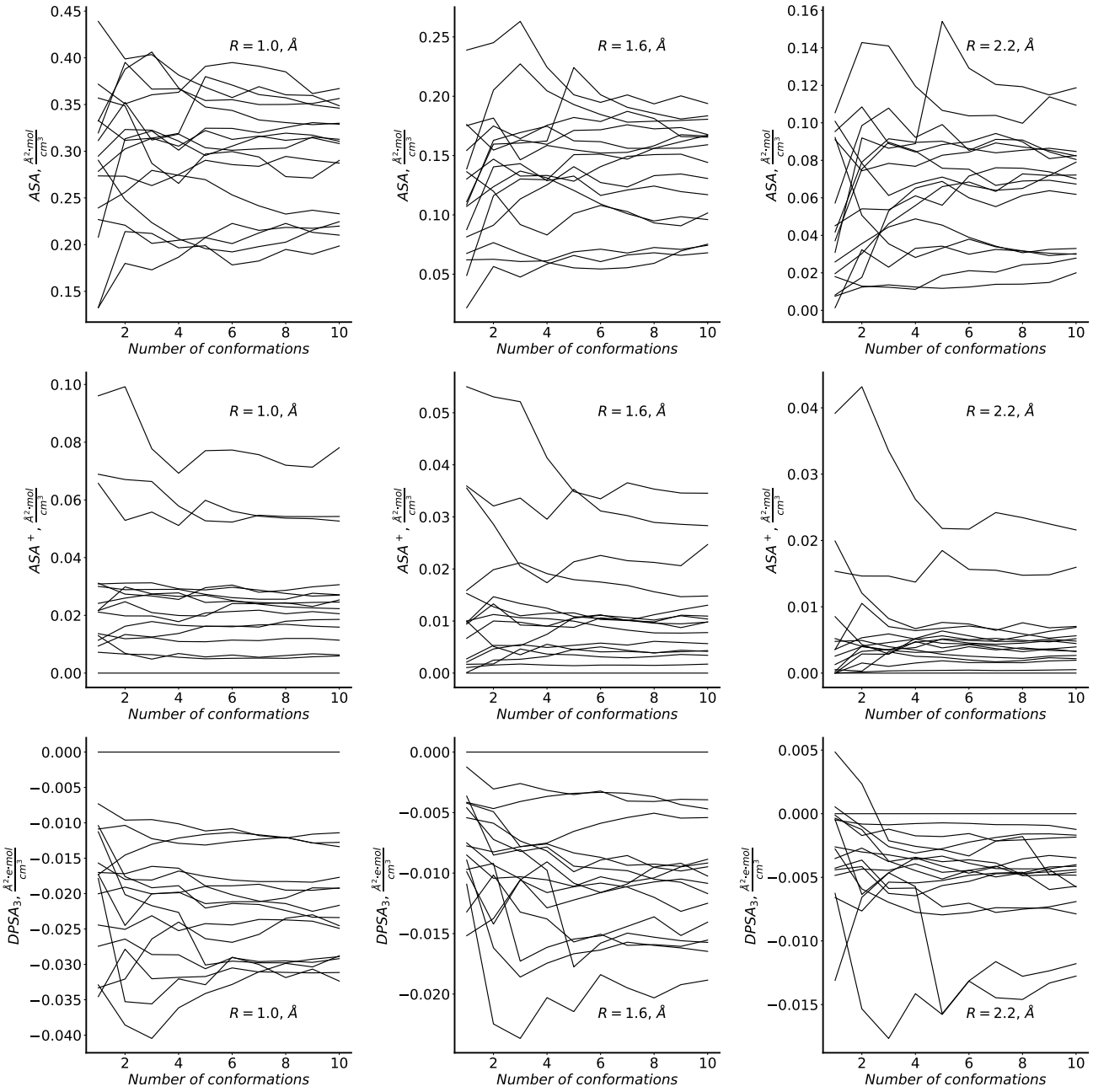


Рисунок 24 – Зависимости значений трех усредненных индексов от количества конформаций, по котором проводилось усреднение, для 16 случайных полимеров.

4.3 Универсальная формула для предсказания коэффициента растворимости S

Стеклообразные аморфные полимеры являются неравновесными объектами, что обосновывает широкий разброс экспериментальных значений газотранспортных параметров. Поэтому такие методы прогнозирования, как аддитивные методы, являются привлекательными для прогнозирования коэффициентов диффузии и коэффициентов проницаемости газов в полимерах, даже несмотря на то, что отклонение предсказанного параметра от экспериментального значения может быть значительным (до 1 порядка): диапазон значений этих параметров для всех полимеров составляет до 8 порядков, что, в целом, делает погрешности прогнозирования приемлемыми. Тем не менее, для коэффициента растворимости ситуация иная. Общий разброс экспериментальных значений данного коэффициента для всех полимеров составляет 2 порядка, поэтому аддитивные методы позволяют достичь точности не большей, чем просто выбор любого возможного значения наугад.

Тем не менее, прогнозирование коэффициента растворимости является важной задачей, так как позволяет оценить без привлечения эксперимента, какой механизм селективности пар газов реализуется в конкретном полимере. Существует два типа селективности в полимерах: диффузионная селективность (обусловленная разностью диффузии) и термодинамическая селективность (обусловленная разностью растворимостей). В то время как диффузионная селективность определяется различиями в размерах пенетрантов и структурой свободного объема в полимере, термодинамическая селективность определяется комплексом факторов, описывающих взаимодействие молекул газа с материалом мембраны. Уточнение механизма селективности позволяет глубже изучить фундаментальные основы транспорта газов в полимерных материалах.

В разделах 2.1.3 и 2.2.3 описана общая методика построения и алгоритм обучения регрессии для предсказания любой транспортной характеристики, вопрос

лишь в использовании корректного набора данных, достаточного для обучения и тестирования объема. Частными случаями подобных регрессий являются «универсальная» регрессия и «частные» газовые регрессии. В этом разделе описывается применение методики для построения «универсальной» и «частных» газовых регрессий. Задача – получить максимально точный прогноз транспортных характеристик полимерных мембран, в частности, коэффициент растворимости S, для новых полимеров.

В набор данных для обучения и тестирования «универсальной» регрессии вошло 1586 экспериментальных данных по паре «газ-полимер» (383 уникальных полимера, 13 различных классов). Подробнее с составом выборки данных можно ознакомиться в таблице 4. Разбиение на обучающую и тестовую выборку проводилось по уникальным полимерам случайным образом, с условием, что данные по полимерам различных химических классов входят в выборки с учетом их общего числа. Таким образом, во-первых, исключается возможность протечки данных, и, во-вторых, все классы полимеров присутствуют и в тестовой и в обучающей выборках. Размер обучающей выборки 1254 строки по парам «газ-полимер» по 303 уникальным полимерам, а тестовой – 332 строки по парам «газ-полимер» по 80 уникальным полимерам.

В результате процедуры из 2.2.3 был получен оптимальный диапазон $[R^-, R^+] = [0.0, 1.4]$Å линеаризации кривых ASA, ..., $PNSA_3$ и значения оптимальных параметров шаговой регрессии $f_{in} = 0.01$, $f_{out} = 0.07$ давшие в результате кросс-валидации максимальную среднюю корреляцию.

Данным значениям параметров соответствовала регрессия (настроенная уже на полной обучающей выборке), которая состоит всего из 6 переменных:

$$\log S = -2.5497 - 1.4019 c_{ASA} - 13.8327 d_{PNSA_3} + 3.7395 c_{PNSA_3} + MaxPA(0.1820 + 0.2345 c_{ASA^+} - 2.3076 d_{PPSA_3}) \quad (4.1)$$

Коэффициент детерминации на тестовой выборке составил $R^2 = 0.72$, а средняя

Таблица 4 – Сводная таблица по выборке данных.

Класс полимера	Ar	C_2H_4	C_2H_6	CH_4	CO_2	H_2	H_2S	He	Kr	N_2	Ne	O_2	Xe	Всего
сополимеры	0	3	3	9	12	6	0	0	0	12	0	13	0	58
другие N-содержащие полимеры	0	0	0	10	10	12	0	4	0	16	0	14	0	66
другие полимеры с углеродной цепью	0	0	0	1	1	1	0	1	0	1	0	1	0	6
другие гетероцепные полимеры	0	0	0	1	1	0	0	0	0	1	0	1	0	4
полиацетилены	0	0	0	6	6	6	0	6	0	6	0	6	0	36
полиакрилаты	2	0	0	2	3	1	1	1	1	3	1	2	0	17
полиамиды	1	0	0	16	18	1	1	1	0	16	1	16	0	71
полиамидоимиды	0	0	0	9	19	0	0	0	0	9	0	9	0	46
поликарбонаты	2	0	0	11	11	2	0	9	0	11	0	11	0	57
полиэфиры	3	0	2	34	34	0	0	3	0	35	0	38	0	149
полиэфиры	0	0	0	20	26	3	0	2	0	20	0	20	0	91
полиимиды	2	2	2	142	153	73	1	28	0	157	0	137	0	697
полинорборнены	0	9	6	17	13	10	0	0	0	14	0	13	0	82
полистиролы	0	0	0	8	8	0	0	0	0	7	0	7	0	30
полисульфоны	0	0	0	31	31	0	0	0	0	28	0	28	0	118
виниловые полимеры	7	0	1	4	7	5	0	8	5	7	4	8	2	58
Всего	17	14	14	321	353	120	3	63	6	343	6	324	2	1586

относительная ошибка $MPE = 104\%$. Диаграмма рассеяния представлена на рисунке 25.

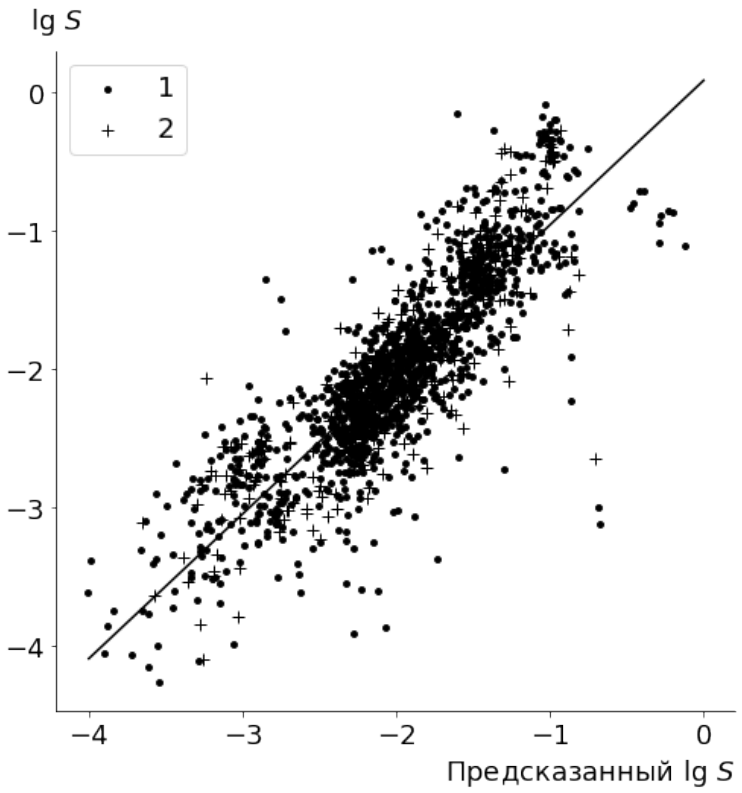


Рисунок 25 – Диаграмма рассеяния для предсказания $\log S$.lg[см^3 (н.у.)/ см^3 (см.рт.ст.)]. 1 – обучающая выборка, 2 – тестовая выборка.

Серьезный вклад в погрешность дали различные экспериментальные данные, полученные при одинаковых условиях различными научными коллективами. В подобных случаях для решения регрессионной задачи нам пришлось брать усредненные экспериментальные данные по S.

Поскольку полимер новый, обучить полимер-специфичную регрессию не получится, так как нет эмпирической информации о его свойствах. Поэтому необходимо либо обучить «универсальную» регрессию, на всем наборе данных, либо газ-специфичную, но на более узком наборе наборе данных по отдельным, так как мы ограничены экспериментальными данными лишь по одному газу. В случае обучения газ-специфичной регрессии, для исключения возможности переобучения, необходимо использовать меньшее количество значимых

предикторов. Исследование подобных регрессий очень важно, поскольку в парах «газ-полимер» последний является гораздо более сложным объектом для количественного описания.

Мотивация рассмотреть отдельные регрессии для каждого газа заключается в том, что «универсальная» регрессия, скрывает индивидуальное поведение каждого газа, что может быть важно для анализа селективности. Следовательно, «частные» регрессии для прогнозирования коэффициентов растворимости могут быть построены для подмножества начального набора данных, ограниченного наблюдениями за отдельными газами. Таблица 5 показывает, что только шесть газов (CH_4, CO_2, CO, N_2, O_2 и H_2) имеют достаточное количество экспериментальных данных для регрессионного анализа.

На основе значимых переменных, полученных для универсальной регрессии, нами были построены частные регрессии для газов He, H_2, O_2, N_2, CO_2, CH_4. Частные регрессии показали несколько лучшее качество. Коэффициенты регрессии для этих газов, а также оценка качества Mean Relative Error (MRE) представлены в Таблице 5 .

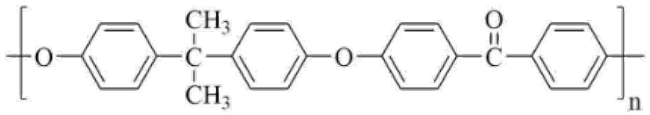
4.4 Предсказание коэффициента растворимости S для задачи поиска высокопроницаемых полимеров

В рамках проекта, результаты которого были позднее опубликованы в работе [14], проводился поиск высокопроницаемых полимеров. Были отобраны потенциальные структуры, и необходимо было определить наиболее перспективный с точки зрения растворимости, чтобы проводить подробные исследования уже с ним. Структуры полимеров представлены на рисунке 26.

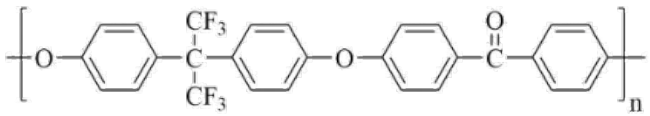
В рамках этой работы для полимеров из работы [14] с использованием частных газовых регрессий мною был предсказан коэффициент растворимости S. Примеры конформаций, полученных для этих полимеров, изображены на ри-

Таблица 5 – Газы-пенетранты в наборе данных и средняя относительная ошибка предсказаний коэффициента растворимости S полученных с «частными» регрессиями.

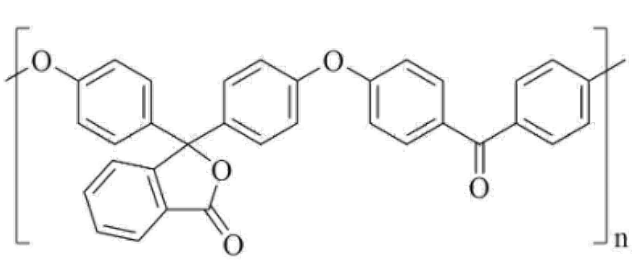
Газ-пенетрант	Обучающая выборка		Тестовая выборка	
	Количество измерений	Средняя относительная ошибка предсказания S, %	Количество измерений	Средняя относительная ошибка предсказания S, %
CH_4	256	93	65	77
CO_2	280	87	73	65
H_2	91	56	29	70
He	49	68	14	100
N_2	270	85	73	59
O_2	256	59	68	53



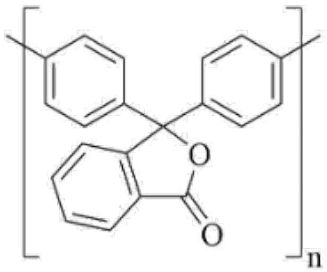
(а) C_1.



(б) C_2.



(в) C_3.



(г) C_4.

Рисунок 26 – Стурктуры полимеров C_1 - C_4.

сунке 12. Коэффициенты растворимости, предсказанные с помощью метода ППКПЦ, представлены в таблице 6.

Предсказанные значения сравнивались с коэффициентами растворимости, рассчитанными методами МАС и ВС (таблица 6), где предсказанный коэффициент растворимости S_{pred} вычислялся как отношение P_{pred}/D_{pred}, где P_{pred} и D_{pred} равны прогнозируемым коэффициентам проницаемости и диффузии, соответственно, полученным с помощью методов МАС или ВС. Экспериментальные коэффициенты растворимости, полученные как P/D (оба параметра экстраполированы до 35°C [112]), также представлены в таблице 6. Следует отметить, что экстраполяция P и D на 35°C была проведена методом из [112] с использованием слабых корреляций ($R^2 = 0,25 - 0,66$), что может привести к большой ошибке в расчетах P (35°C) и D (35°C). Следовательно, ошибка вычисления S (35°C) также может быть высокой. Однако, насколько нам известно, метод из [112] является единственным методом экстраполяции экспериментальных значений P и D до 35°C, когда температурные зависимости недоступны.

Стоит отметить, что средняя абсолютная ошибка (MAPE) и средняя процентная ошибка (MPE), представленные в таблице 6, были рассчитаны для каждого из четырех исследованных полимеров. MAPE были также рассчитаны для отдельных регрессий для различных газов с использованием всего валидационного набора полимеров. Они были следующими: He – 100%, H_2 – 56%, O_2 – 53%, N_2 – 59%, CO_2 – 67%, CH_4 – 77%. Можно заметить, что, хотя общие MAPE являются удовлетворительными, некоторые вызывающие беспокойство большие MAPE были получены для некоторых отдельных полимеров (MAPE для C_1 составляет 175%, а для C_2 достигает 362%).

Диаграмма сравнения $S_{pred} = f(S_{exp})$ представлена на рисунке 27.

Как показано на рисунке, прогнозы методов МАС/ВС и ППКПЦ находятся в пределах одного порядка величины (за исключением двух точек, полученных методом ВС) по сравнению с экспериментальными значениями. Такой уровень точности прогноза не вполне удовлетворителен для коэффициента растворимо-

Таблица 6 – Экспериментальные коэффициенты растворимости газа $S \cdot 10^3$ (см^3(STP) $\cdot$ см$^{-3}$ $\cdot$ см рт.ст.$^{-1}$) в синтезированных полимерах при 35°C и прогнозные данные для тех же полимеров.

Полимер	Тип данных	He	H_2	O_2	N_2	CO_2	CH_4	$MAPE$, %	MPE, %
C_1	Эксперимент	0.25	1.0	2.9	1.3	52	7.3	-	
	MAC	-	-	2.3	0.93	12	2.2	50	33
	BC	-	-	2.2	1.0	19	2.6	44	30
	SPCSBP	0.76	2.4	7.8	5.9	77	18	175	-175
C_2	Эксперимент	0.49	1.0	2.9	1.8	44	7.7	-	
	MAC	-	-	4.0	1.7	18	3.6	40	13
	BC	-	-	4.8	1.8	36	5.0	30	-3
	SPCSBP	2.1	4.5	14	11	140	39	362	-362
C_3	Эксперимент	0.40	1.4	5.9	3.7	110	15	-	
	MAC	-	-	2.5	1.5	26	6.4	63	42
	BC	-	-	7.2	1.0	29	2.9	63	34
	SPCSBP	1.2	2.9	8.3	7.6	100	25	86	-84
C_4	Эксперимент	0.89	2.3	9.1	6.6	190	33	-	
	MAC	-	-	5.2	3.2	66	23	47	65
	BC	-	-	18	0.47	11	2.1	94	64
	SPCSBP	1.2	4.0	12	12	140	32	42	-32

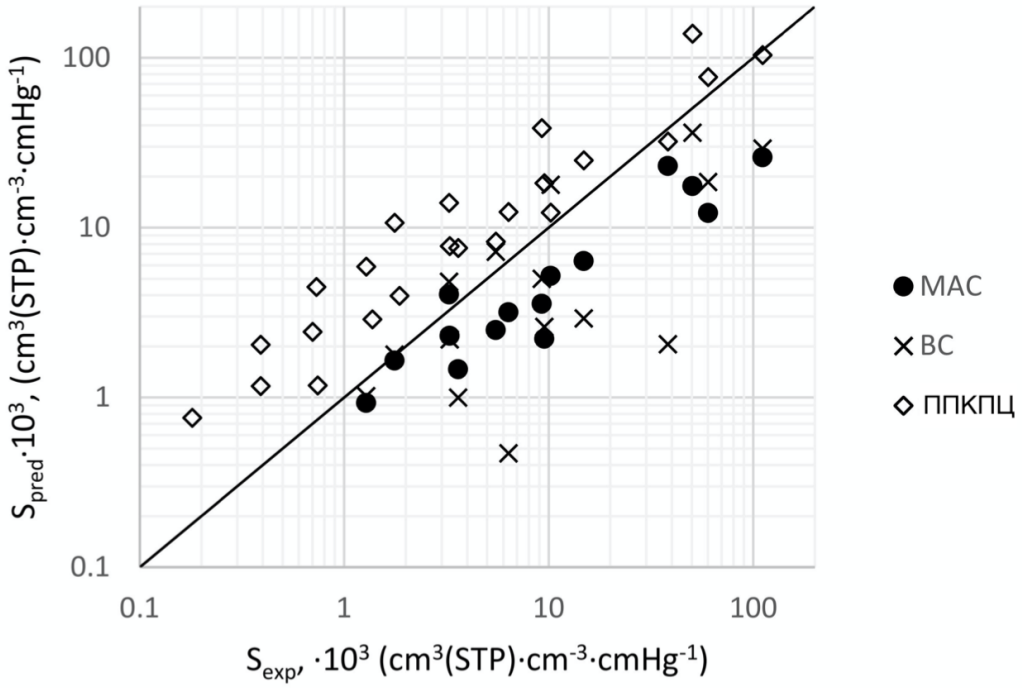


Рисунок 27 – Диаграмма сравнения расчетных и экспериментальных значений коэффициента растворимости.

сти; однако можно сделать некоторые интересные наблюдения для дальнейшего развития методов. Стоит отметить, что методы МАС/ВСи ППКПЦ показали противоположные отклонения; в основном отрицательно в случае МАС/ВС и в основном положительно в случае ППКПЦ. Используя результаты расчета MPE, можно ясно видеть, что наиболее заниженные (положительные числа MPE) значения P_{pred} в сочетании с завышенными (отрицательные числа MPE) значениями D_{pred} приводят к ожидаемому заметному занижению S.

Метод ППКПЦ базируется на физической интуиции, поскольку он включает моделирование длинноцепочечных фрагментов и основан на теории, связывающей площади поверхности полимера и газа с растворимостью газа в полимерах [114]. Однако, если сравнить ошибки прогноза для обсуждаемых полимеров, становится очевидным, что ППКПЦ дает хорошие прогнозы коэффициента растворимости для C_3 и C_4, в то время как для C_1 и C_2 наблюдаются значительные ошибки. Основное различие между полимерами C_3 – C_4 и C_1 – C_2 заключается в наличии фталидной группы в первой паре и наличии группы $–C(CH_3)_2$

или –$C(CF_3)_2$ во второй. Мы рассчитали MAPE для поли (ариленэфиркетона) с фталидной группой и без ее присутствия в используемом наборе обучающих данных. Они составили 93% и 14% для –$C(CH_3)_2$ или –$C(CF_3)_2$ и фталидсодержащих поли (ариленэфиркетонов) s, соответственно. Похоже, что ошибка метода сохраняется и для –$C(CH_3)_2$ или –$C(CF_3)_2$ содержащих поли (ариленэфиркетонов) s, поскольку MAPE для таких полимеров в шесть раз выше, чем для других. Возможное объяснение явления может быть связано с тем обстоятельством, что (i) индексы, используемые в методе ППКПЦ, не описывают особенности групп–$C(CH_3)_2$ или –$C(CF_3)_2$ в поли (ариленовых) структурах эфиркетона и их влияние на растворимость газа; (ii) экспериментальных данных для поли (ариленэфиркетона) с такими фрагментами недостаточно для обучения регрессии.

4.5 Предсказание константы Генри k_D

Как показано в предыдущем разделе, ППКПЦ позволяет с удовлетворительной точностью предсказать коэффициенты растворимости около 20 газов в примерно 60 стеклообразных полимерах разной структуры. Межу тем, в этом подходе почти полностью игнорируется влияние свободного объема на коэффициенты растворимости, хотя известно, что многие полимеры, в том числе представляющие интерес в качестве мембранных материалов, отличаются повышенным свободным объемом и бóльшими коэффициентами растворимости, чем обычные стеклообразные полимеры [25, 56, 79].

Мерой свободного объема может служить ленгмюровская сорбционная емкость в модели двойной сорбции C'_H [114]. В рамках данной модели, по формуле (1.11) коэффициент растворимости при бесконечном разбавлении (низких давлениях сорбата) может быть представлен как $S = k_D + C'_H \cdot b$, где k_D – коэффициент растворимости «популяции Генри» молекул сорбата, то есть параметр, характеризующий растворимость в более плотных областях полимерной матрицы, а параметр b является константой равновесия для молекул сорбатов

в элементах свободного объема. Для высокопроницаемых полимеров выполняется неравенство $C'_H \gg k_D$. Таким образом, параметр k_D не должен зависеть от свободного объема в полимере. В связи с этим представляет интерес применение ППКПЦ для предсказания константы закона Генри k_D .

Набор экспериментальных значений k_D из 94 измерений был взят из Базы данных ИНХС. Набор построен на основе анализа литературы и включает в себя данные по растворимости 13 газов в 40 полимерах. Сводная таблица по классам полимеров и присутствующим газам представлена в таблице 7. Проблемой является то, что эксперименты проводились при разной температуре от 278 до 308 K. В большинстве работ отсутствовала информация о температурной зависимости k_D, и использованные в работе значения k_D являются усредненными в указанном температурном интервале.

В связи с отсутствием экспериментальных величин плотности образцов значения индексов не нормализовались на плотность (что равнозначно предположению о средней плотности исследуемых стеклообразных полимеров 1 г/см^3). Отметим отсутствие среди исследуемых газов CO (как и для растворимости в разделе 3.2), а также CO_2, который, предположительно, также демонстрирует индивидуальный характер.

Набор данных был разбит на обучающую (76 измерений) и тестовую (17 измерений) выборки. Предполагалось, что искомая регрессия имеет вид, аналогичный выражению (2.2):

$$\log k_D(gas, polymer) = a(polymer) \cdot MaxPA(gas) + b(polymer) \qquad (4.2)$$

где коэффициенты a и b зависят от полимера. Как и в разделе 3.2), они подбирались согласно процедуры из 2.2.3 в виде взвешенной суммы коэффициентов линейной аппроксимации кривых $ASA(R), \ldots, PNSA_3(R)$ на отрезке [2.2 Å; 2.6 Å]. Также в качестве объясняющей переменной использовался коррелянт свободного объема (согласно 2.1), ван-дер-ваальсов объем $VDWV$ молекулы (для

Таблица 7 – Сводная таблица по классам полимеров и присутствующим газам в выборке для предсказания k_D.

Химический класс	Ar	C_2H_4	C_2H_6	C_3H_6	C_3H_8	C_4H_{10}	CH_4	H_2	He	N_2	N_2O	O_2	Xe	Всего
виниловые полимеры	2		1				3	1	1	2	1		1	12
другие гетероцепные полимеры				1	1								1	3
полиакрилаты	1	1			1		1							4
полиамидоимиды										2		2		4
полиацетилены	1		2		2	1	2	1		2	1	1	2	15
полиимиды		1		7	2		2			3		2		17
поликарбонаты						1	4			1			1	7
полинорборнены	5		5	1			5		3	5				24
полисульфоны							1							1
полиэфиры							1							1
полиэфиры							2			2		1		5
Всего	9	2	8	9	6	2	21	2	4	17	2	6	5	93

проверки высказанного выше предположения о том, что для предсказания k_D показатели свободного объема менее значимы).

В результате применения шаговой регрессии (со значениями оптимальных параметров шаговой регрессии $f_{in} = 0.01$, $f_{out} = 0.07$) была получена формула

$$\log k_D = MaxPA(0.12 - 2.96 d_{PPSA_3}) - 4.11, \tag{4.3}$$

задействующая единственную переменную, характеризующую полимер – это чувствительность индекса $PNSA_3$ (площади отрицательно заряженной доступной поверхности с учетом величины частичного заряда) к радиусу «обкатки» (Обучающая выборка: $N = 76$, $R^2 = 0,79$, $RMSE = 0.12$ (среднеквадратичная ошибка). Тестовая выборка: $N = 17$, $R^2 = 0.81$, $RMSE = 0.09$). Значения $\log k_D$ в обучающей выборке, а также остатки, распределены нормально.

Меньшее число значимых переменных обусловлено небольшим объемом выборки, не позволяющим выделить тонкие эффекты. Отметим, что ван-дер-

ваальсов объем оказался незначимой переменной; это подтверждает предположение о том, что константа закона Генри k_D в первом приближении не зависит от свободного объема полимерной матрицы.

Коэффициент детерминации регрессии составил 0.79 (скорректированный $R^2 = 0.78$) на обучающей выборке (при корреляции 0.89) и 0.81 на тестовой выборке (при корреляции 0.9), что соответствует средней относительной ошибке предсказания k_D 69% на обучающей выборке и 47% на тестовой выборке. Таким образом, несмотря на меньший объем выборки, отсутствие нормировки объясняющих переменных на экспериментальную плотность образца, а также на отсутствие нормализации значений k_D по температуре, удалось получить даже лучшую точность предсказания, чем для коэффициента растворимости S (см. раздел 3.2).

На рисунке 28 приведена диаграмма рассеяния для регрессии (4.3).

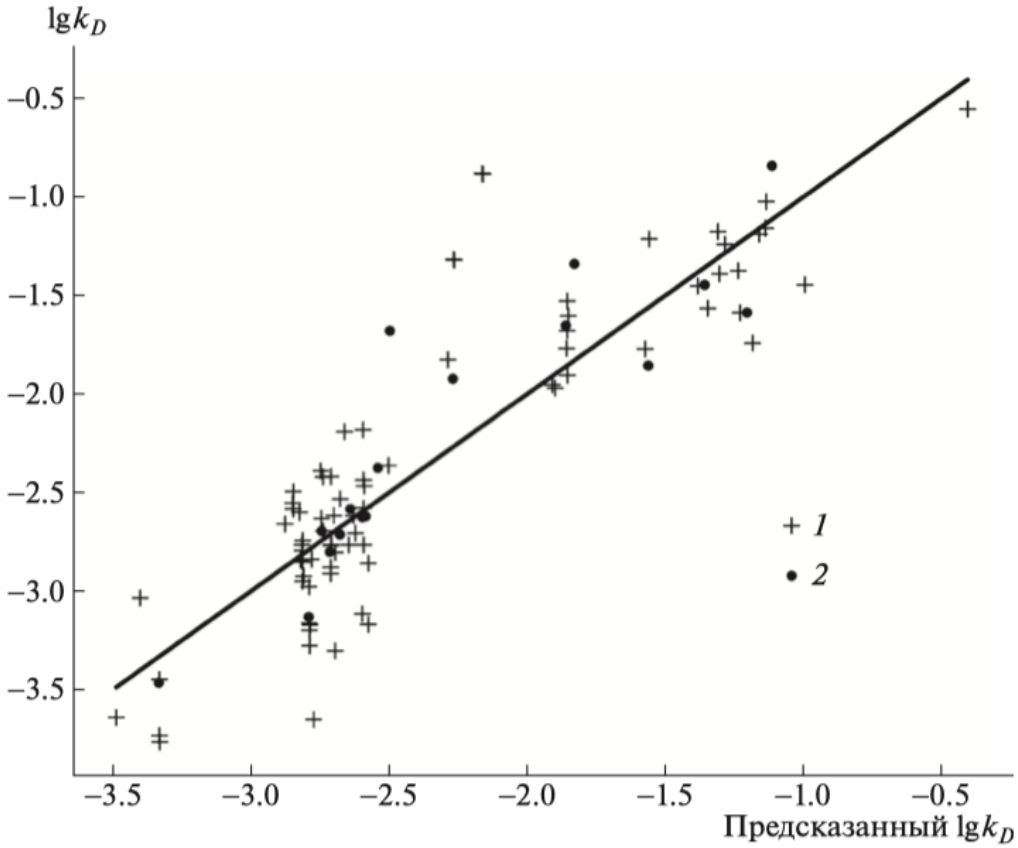


Рисунок 28 – Диаграмма рассеяния для предсказания logk_D, lg[см^3 (н.у.)/ см^3 (см.рт.ст.)]. 1 – обучающая выборка, 2 – тестовая выборка.

4.6 Методы кластеризации для анализа и предсказания транспортных характеристик

Исследования, описанные в предыдущих разделах, показали, что анализ геометрии короткого отрезка полимерной цепи (в том числе, геометрии доступной поверхности молекулы) позволяет сделать содержательные выводы о транспортных характеристиках полимерного материала. Регрессионные модели из разделов 3.2-4.5 эффективно предсказывают коэффициент растворимости и константу Генри стеклообразных полимеров. Анализируя значимые переменные и их вклад в числовую характеристику той или иной транспортной характеристики, химик-исследователь получает некоторые указания о том направлении, в котором необходимо изменять молекулу для получения экстремальных транспортных характеристик.При этом нет прямой связи между химическим классом полимера и его транспортными характеристиками – полимеры разных классов могут демонстрировать экстремальные характеристики.

Однако зачастую исследователю непросто связать молекулярную структуру потенциального полимера с ее поверхностными характеристиками, и хотелось бы иметь более простые и визуально представимые правила. Так, было высказано предположение, что молекулы близкой геометрической формы независимо от их химического класса обладают схожими транспортными характеристиками, а для классификации геометрических форм можно использовать поверхностные геометрические индексы. В настоящем разделе описывается вариант использования ППКЦ для построения не регрессий, но классификаторов на основе тех же геометрических индексов, описанных в разделе 2.1.2.

Для решения задачи кластеризации из Базы данных был отобран 401 уникальный полимер, для каждого из которых существует несколько записей, содержащих экспериментальные данные по взаимодействию пары «газ-полимер» и условия, при которых оно осуществлялось (всего 2663 записи). База данных содержит больше уникальных структур, однако для задачи кластеризации необ-

ходимо выдержать баланс в данных по их химическим классам. Поэтому структуры отбирались по правилу:

- не более 42 структур одного химического класса
- если более, то выбираем те, у которых есть экспериментальные данные по *P, D, S*.
- если меньше, то выбирались все структуры.

Таким образом было выбрано 397 уникальных полимеров различных классов (см. таблицу 8).

После создания выборки для каждого полимера, согласно обоснованию ранее в разделе 3.2, было сгенерировано по 6 конформаций размером более 600 атомов (не считая атомы водорода). Примеры полученных конформаций можно увидеть на рисунке 12. Далее для каждой конформации для каждого R в диапазоне 0Å до 3Å рассчитаны геометрические индексы.В отличие базового варианта метода ППКПЦ, в котором выбирается оптимальный диапазон $[R^-, R^+]$ линеаризации полученных зависимостей и полученные коэффициенты наклона и смещения используются в качестве объясняющих переменных регрессии, здесь используется вся кривая. Таким образом, одна конформация представлена 8 векторами (по числу геометрических индексов из таблицы 1 или 160 переменными (20 вычисленных значений геометрического индекса в зависимости от радиуса «обкатки» на каждый из 8 индексов), а полимер – усредненными переменными по шести конформациям.

Для кластеризации полимеров был использован агломеративный метод (разновидность иерархической классификации) [11]. Агломеративный метод кластеризации был запущен для числа кластеров в диапазоне $k = 2, \ldots, 30$. Согласно значениям силуэтного коэффициента, индекса Калински-Харабаша (CH) и индекса Дэвиса-Болдуина (DBI) для построенных кластеризации было выбрано значение $k = 15$. Для иллюстрации качества полученной кластеризации с помощью алгоритма t-SNE [103] приведено размещение построенных кластеров

Таблица 8 – Сводная таблица по классам полимеров для задачи кластеризации.

Хим. класс	Номер кластера															Σ
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
полиакрилаты	1	3	1	9		1										15
полиэфиры		2	18	1	1						4		4	4	8	42
полиэфиры		1	8		12	4	2		2	5			3			37
полифосфазены								3								3
полиацетилены					1	4		3				11				19
полинорборнены			5	7	2	2	2	9	1	6			3			37
полисульфоны					19	1				18					2	40
другие N-содержащие		4	1	5	3						4		9			26
полиамиды			11	1	16	3			2				4			37
полистиролы	2		3	1	2	1		13								22
виниловые полимеры	1			3		2	1	5	1				1			14
поликарбонаты			14			1							1		3	19
полиимиды		3	3	5	6						9		16			42
полиамидоимиды		2		2	1		1				11		23			40
другие углеродные цепи						1				2						3
другие гетероцепные								1								1
Σ	4	15	64	34	63	20	6	34	6	31	28	11	64	4	13	397

полимеров в двумерном пространстве (см. рисунок 29). Алгоритм t-SNE является техникой нелинейного снижения размерности, хорошо подходящей для вложения данных высокой размерности при визуализации в пространстве низкой размерности (двух- или трехмерное).

Для сопоставления результатов кластеризации с экспериментальными значениями транспортных характеристик, полученные кластеры были перенесены на диаграмму Робсона, популярную среди специалистов по мембранному газоразделению (см. раздел 1.1). Диаграмма Робсона строится для пары газов (например кислород-азот) в двойных логарифмических координатах $\alpha = P_{O_2}/P_{N_2}$ – селективность при разделении пары газов (кислород-азот) и P_{O_2} -коэффициента проницаемости для более проницаемоего газа (кислорода).

Экспериментальные данные по коэффициенту проницаемости и диффузии брались из Базы данных. В связи с тем, что экспериментальные данные были получены при различных температурах, первоочередной задачей стало приведение их к единому значению 308K. Для этого использовался алгоритм из [1]. После приведения к общей температуре 308K были построены диаграммы по имеющимся в Базе данных экспериментальным данным по коэффициентам проницаемости P и диффузии D. Затем внесены данные о полученных кластерах (см. рисунок 30 и 31 соответственно). Для удобства анализа полученных диаграмм точки отдельных кластеров помещались на фоне основного пятна диаграммы Робсона, а также отмечались посчитанные за вычетом явных выбросов центры масс кластеров и всей выборки.

Выделенные кластеры оказываются тесно связанными с транспортными характеристиками полимеров. Расположение центра масс кластера, относительно центра масс всей выборки, форма кластера и его состав демонстрируют эти связи. Первое, что стоит отметить – точки большинства кластеров расположены на диаграммах Робсона довольно кучно. Следовательно, в кластере оказались материалы со схожими транспортными характеристиками. Но кластеризация построена только на основе формы и геометрии конформаций молекул полиме-

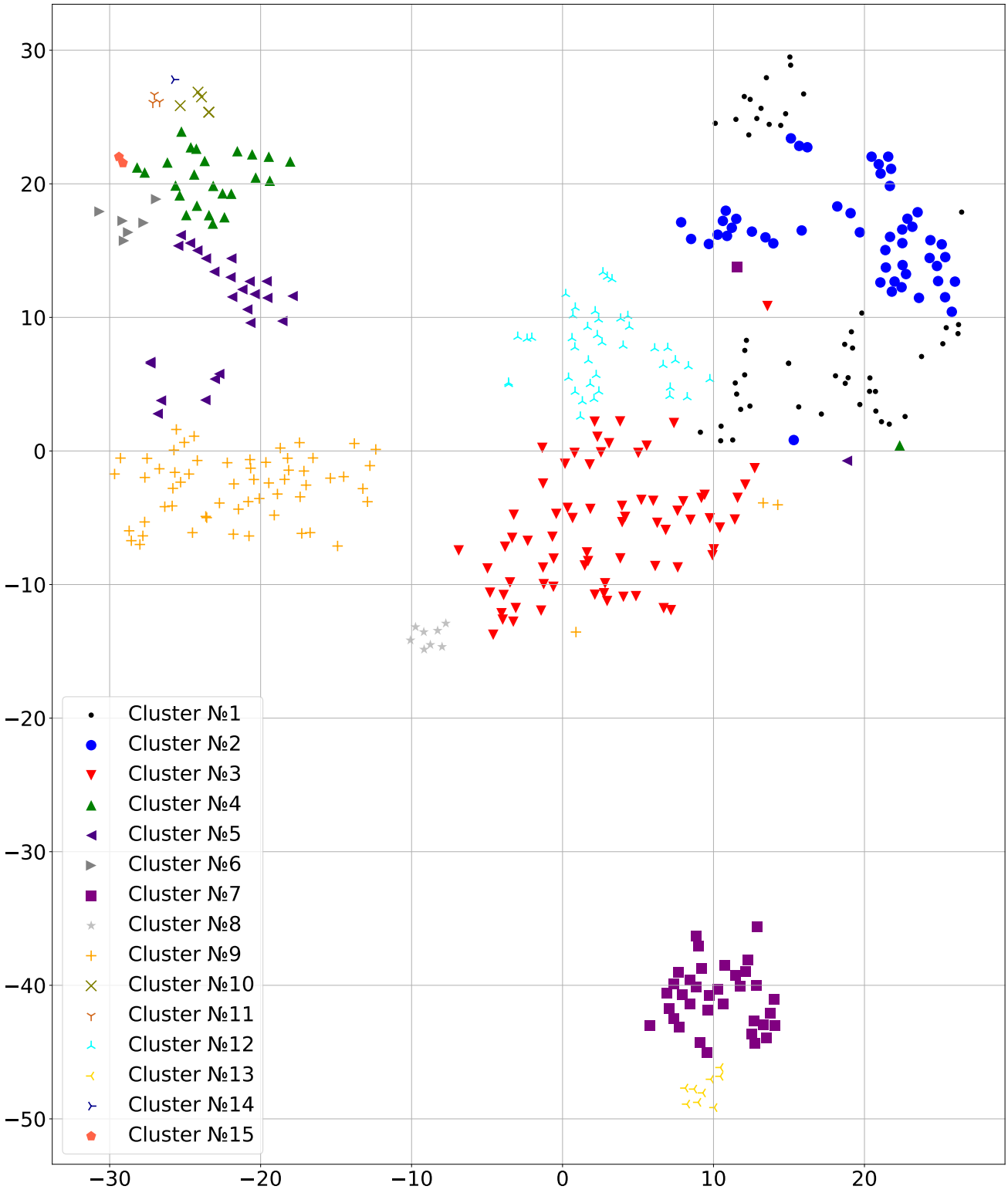


Рисунок 29 – Размещение кластеров полимеров в двумерном пространстве с помощью алгоритма t-SNE.

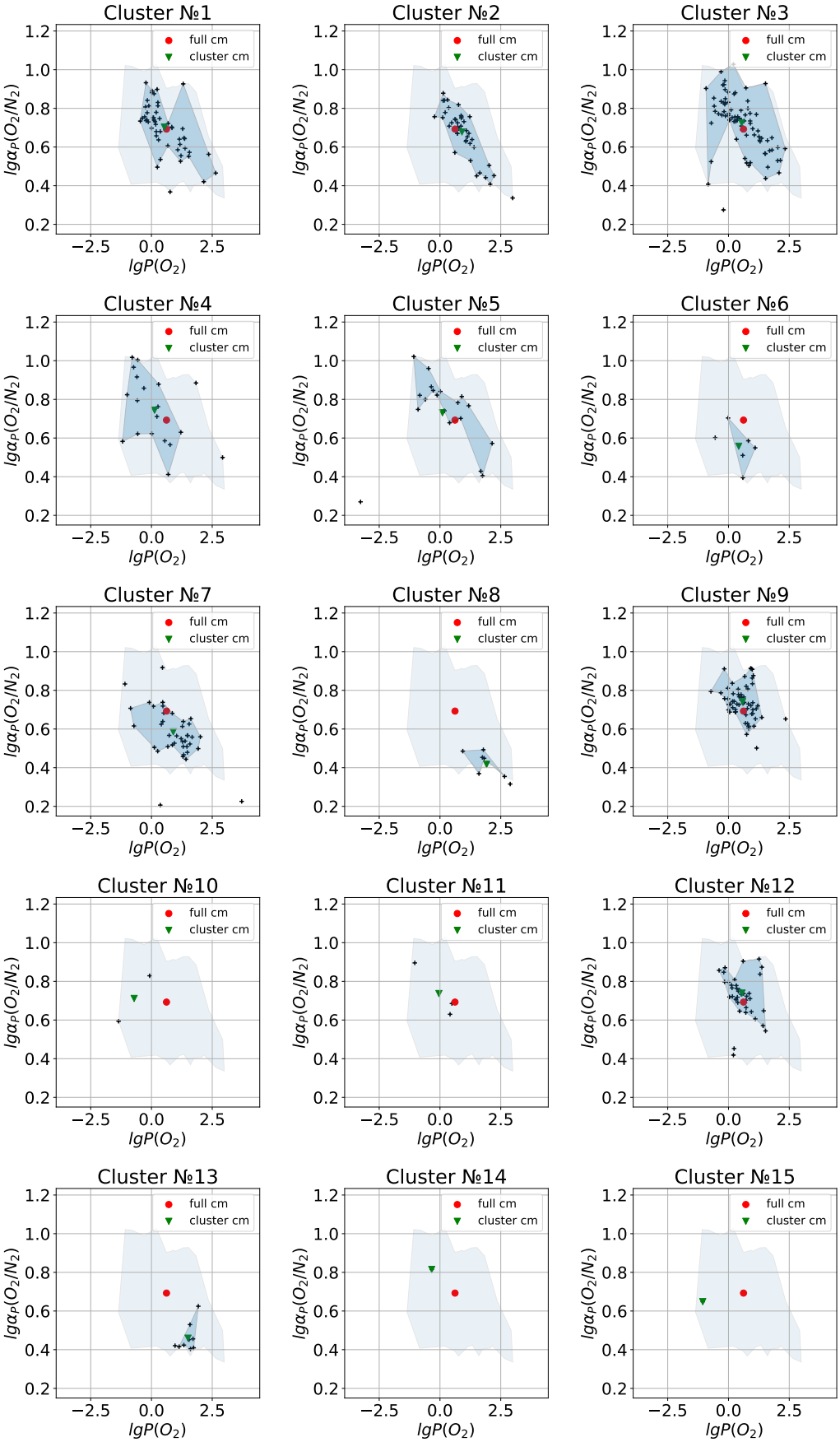


Рисунок 30 – Полимерные кластеры на диаграмме проницаемости. Центр всей выборки данных показан красным кружком, а центр кластера – зеленым треугольником.

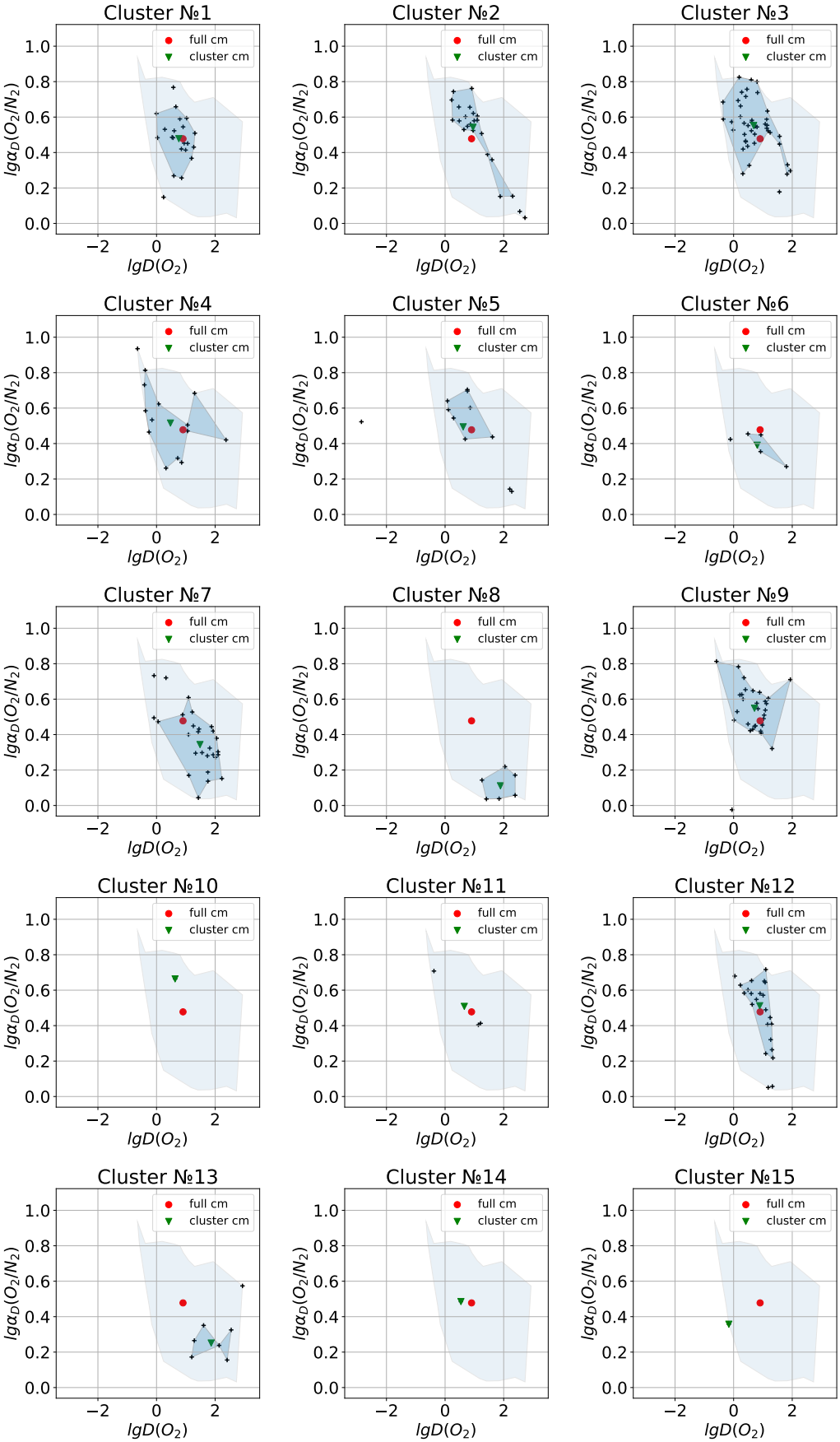


Рисунок 31 – Полимерные кластеры на диаграмме диффузии. Центр всей выборки данных показан красным кружком, а центр кластера – зеленым треугольником.

ра. Таким образом, полученные результаты подтверждают высказанное предположение о том, что транспортные характеристики полимерных материалов в большей степени определяются геометрическими свойствами молекул чем химическим классом веществ, из которых эти молекулы составлены.

Рассмотрим кластеры в отдельности. Наиболее показательные зависимости можно найти на диаграмме Робсона для селективности диффузии на рисунке 31. Так, центры масс больших кластеров 3, 5, 13 смещены в левый верхний квадрант общей диаграммы. Это говорит о высокой селективности по D и, при этом, с невысоким коэффициентом диффузии. Из рисунка 30 видно, что полимеры из этих кластеров демонстрируют хорошее сочетание селективности по проницаемости и самой проницаемости. Вероятно, к ним можно отнести кластер 11, однако сделать однозначный вывод не представляется возможным по причине недостатка экспериментальных данных по коэффициенту диффузии. Полимеры из кластеров 6, 8, 12 в своем большинстве оказались низкоселективными по проницаемости и диффузии, однако при этом высоко проницаемыми и высокодиффузными.

Можно отметить, что в целом кластеризация не зависит от химического класса полимеров. Так рассмотренные выше кластеры 3, 5, 6, 8, 13 состоят из полимеров различных химических классов. Однако, есть и исключения. Например, все высокодиффузные и высокопроницаемые полимеры с низким значением селективности из 12 кластера относятся к *полиацетиленам*. Все *полифосфазены* вошли в кластер 8, большая часть *поликарбонатов* вошла в кластер 3, а *полисульфоны* разделились на два кластера – 5 и 10. При этом, часто схожие по структуре и свойствам *полиамидоимиды* и *полиимиды* в основном разделились по кластерам 11 и 13. Теперь обратимся к строению самих полимеров. Так, кластер 6 состоит только из полимеров с объемными фторсодержащими заместителями, которые при этом относятся к различными химическим классам. В сочетании с информацией о том, что полимеры этого кластера высокопроницаемые и высокодиффузные и при этом низкоселективные, можно судить о

последствиях добавления объемных фторсодержащих заместителей. При этом, кластер 7, содержащий в основном полимеры с *пентафторфенильной* группой, является скорее наоборот низкоселективным и низкодиффузным, что говорит о том, что не всегда наличие большого количества молекул фтора приводит к свойствам, схожим с со свойствами полимеров из кластера 6. Также интересно сравнить кластеры 6 и 10. В кластере 10 также есть объемные фторсодержащие заместители, однако мономерные звенья чаще в разы длиннее и бóльшая часть полимеов относится к классам *полиамидоимиды* и *полиимиды*, которые отсутствуют в кластере 6.

Проведенный анализ позволяет говорить о глубокой зависимости транспортных характеристик полимеров от формы и геометрии конформаций молекул полимера. Полученная кластеризация позволяет выделить признаки и характеристики молекул полимеров с различными экстремальными характеристиками, что будет несомненно полезно в задачах поиска и синтеза новых перспективных полимеров.

Заключение

В процессе работы над диссертацией были получены следующие результаты:

1. Проведен анализ существующих математических моделей стеклообразных полимеров и методов предсказания их транспортных характеристик.
2. Разработан метод моделирования конформаций полимерных цепей, позволяющий получить реалистичные молекулы полимерных цепей, необходимые для вычисления поверхностных и поверхностно – зарядных геометрических индексов. В частности, предложено семейство геометрических молекулярных индексов, основанных на анализе кривых зависимости площади доступной поверхности молекул от радиуса «обкатки».
3. На основе регрессионных моделей разработаны численные методы предсказания транспортных характеристик полимерных мембран.
4. Разработан комплекс программ, позволяющий полностью автоматизировать процесс моделирования полимерных цепей стеклообразных мембранных материалов и предсказания их транспортных характеристик. Все вычисления имеют возможность распараллеливания на кластере и приемлемое (для работы на ПК) время расчета одного полимера.
5. Продемонстрирована эффективность разработанных методов и алгоритмов в задачах предсказания транспортных характеристик полимерных мембран, таких как коэффициент растворимости S и константа закона Генри k_D. Предложена кластеризация конформационных структур аморфных полимеров, основанная лишь на значениях геометрических индексов.

Список литературы

[1] Алентьев А.Ю. Прогнозирование транспортных свойств стеклообразных полимеров: Роль химической структуры и свободного объема / А.Ю. Алентьев: дис. ... докт. хим. наук: 05.17.18. - М., 2003. - 368 с.

[2] Аскадский А.А. Методы расчета физических свойств полимеров // Обзорный журнал по химии. 2015. Том 1. № 2. С. 101-164.

[3] Губко М.В., Милосердов О.А., Ямпольский Ю.П., Алентьев А.Ю., Рыжих В.Е. Универсальная модель для предсказания растворимости газов в стеклообразных полимерах // Тезисы докладов XIII всероссийской научной конференции (с международным участием) Мембраны-2016, 10-14 октября 2016 года, Нижний Новгород. 159-161.

[4] Губко М.В., Милосердов О.А., Ямпольский Ю.П., Рыжих В.Е. Новый метод предсказания растворимости и константы неспецифической сорбции k_D для легких газов в стеклообразных полимерах // Тезисы докладов XIV Всероссийской научной конференции МЕМБРАНЫ - 2019. — Сочи, 2019. — С. 344–346.

[5] Милосердов О.А., Губко М.В. Использование алгоритма Ли-Ричардса для расчета поверхностно-зарядных характеристик макромолекул в задаче предсказания транспортных свойств стеклообразных полимеров / Труды 61-й Всероссийской научной конференции МФТИ "Радиотехника и компьютерные технологии – Москва, 2018. – С. 46-48.

[6] Милосердов О.А., Губко М.В. Классификация конформационных структур аморфных полимеров в интересах мембранной технологии // Труды 60-й научной конференции МФТИ: Радиотехника и кибернетика, – Москва, 2017. – С. 75-76.

[7] Милосердов О.А., Губко М.В. Классификация стеклообразных полимеров по транспортным свойствам на основе локальной геометрии полимерных цепей / Труды 16-й Всероссийской школы-конференция молодых ученых «Управление большими системами» (УБС'2019, Тамбов). Тамбов: Тамбовский государственный технический университет, 2019. С. 107-110.

[8] Милосердов О.А., Губко М.В., Ямпольский Ю.П., Белов Н.А. Классификация конформационных структур аморфных полимеров мембранного назначения / Тезисы докладов XIV Всероссийской научной конференции МЕМБРАНЫ - 2019. — Сочи, 2019. — С. 177–179.

[9] Милосердов О.А., Губко М.В., Ямпольский Ю.П., Рыжих В.Е. Предсказание транспортных характеристик стеклообразных полимеров по зависимости площади макромолекулы от радиуса обкатки / Материалы XV Всероссийской школы-конференции молодых ученых УБС 2018: Управление техническими системами и технологическими процессами, 10-13 сентября 2018 года. – ВГТУ, 2018. – Т. 2, – С. 76-80.

[10] Ямпольский Ю. П. Методы изучения свободного объема в полимерах // Успехи химии. – 2007. – Т. 76. – №. 1. – С. 66-87.

[11] Agglomerative Clustering sklearn [Электронный ресурс]. Доступ URL: scikit-learn.org/stable/modules/generated/sklearn.cluster.AgglomerativeClustering.html (дата обращения: 01.05.2022).

[12] Alder B. J., Wainwright T. E. Studies in molecular dynamics. I. General method // The Journal of Chemical Physics. – 1959. – Vol. 31. – No. 2. – P. 459-466.

[13] Alentev A.Yu, Ronova I.A., Schcukin B.V., Yampolskii Yu P. Correlation between the transport behavior of polyimides and the conformational rigidity of their chains // Polymer Science Series A. – 2007. – Vol. 49. – No. 2. – P. 217-226.

[14] Alentiev A, Chirkov S, Nikiforov R. et al. Structure-Property Relationship on the Example of Gas Separation Characteristics of Poly (Arylene Ether Ketone) s and Poly (Diphenylene Phtalide) // Membranes. – 2021. – Vol. 11. – No. 9. – P. 677.

[15] Alentiev A. Yampolskii Yu P., Ryzhikh V. The database "gas separation properties of glassy polymers"(Topchiev Institute): capabilities and prospects // Petroleum Chemistry. – 2013. – Vol. 53. – No. 8. – P. 554-558.

[16] Anderson J. A., Lorenz C. D., Travesset A. General purpose molecular dynamics simulations fully implemented on graphics processing units // Journal of Computational Physics. – 2008. – Vol. 227. – No. 10. – P. 5342-5359.

[17] Askadskiĭ A. A. Computational materials science of polymers. – Cambridge Int Science Publishing, 2003.

[18] Aspuru-Guzik A., Lindh R., Reiher M. The matter simulation (r) evolution // ACS Central Science. – 2018. – Vol. 4. – No. 2. – P. 144-152.

[19] Auhl R., Everaers R., Grest G. et al. Equilibration of long chain polymer melts in computer simulations // The Journal of Chemical Physics. – 2003. – Vol. 119. – No. 24. – P. 12718-12728

[20] Baker R. W. Future directions of membrane gas separation technology // Industrial & Engineering Chemistry Research. – 2002. – Vol. 41. – No. 6. – P. 1393-1411.

[21] Barrer R. M., Barrie J. A., Slater J. Sorption and diffusion in ethyl cellulose. Part III. Comparison between ethyl cellulose and rubber // Journal of Polymer Science. – 1958. – Vol. 27. – No. 115. – P. 177-197.

[22] Berman H., Henrick K., Nakamura H. Announcing the worldwide protein data

bank // Nature Structural & Molecular Biology. – 2003. – Vol. 10. – No. 12. – P. 980-980.

[23] Bernardo P., Drioli E., Golemme G. Membrane gas separation: a review/state of the art // Industrial & Engineering Chemistry Research. – 2009. – Vol. 48. – No. 10. – P. 4638-4663.

[24] Bicerano J. Prediction of polymer properties third edition, revised and expanded // Plastic Engeneeging-New York. – 2002. – P. 65.

[25] Budd P. M., McKeown N.B., Ghanem B.S. et al. Gas permeation parameters and other physicochemical properties of a polymer of intrinsic microporosity: Polybenzodioxane PIM-1 // Journal of Membrane Science. – 2008. – Vol. 325. – No. 2. – P. 851-860.

[26] ChemDraw Perkinelmerinformatics [Электронный ресурс]. Доступ URL: https://perkinelmerinformatics.com/products/research/chemdraw/ (дата обращения: 03.04.2022).

[27] ChemDraw [Электронный ресурс]. Доступ URL: https://www.perkinelmer.com/es/category/chemdraw. (дата обращения: 04.05.2022).

[28] Chemical Computing Group (CCG) Computer-Aided Molecular Design [Электронный ресурс]. Доступ URL: https://www.chemcomp.com/. (дата обращения: 04.05.2022).

[29] Conformer Plugin Chemaxon [Электронный ресурс]. Доступ URL: https://docs.chemaxon.com/display/docs/Conformer+Plugin (Дата обращения: 30.05.2022).

[30] Dalby A., Nourse J. G., Hounshell W. D. et al. Description of several chemical structure file formats used by computer programs developed at Molecular

Design Limited // Journal of Chemical Information and Computer Sciences. – 1992. – Vol. 32. – No. 3. – P. 244-255.

[31] Eastman P., Swails J., Chodera J. et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics // PLoS Computational Biology. – 2017. – Vol. 13. – No. 7. – P. e1005659.

[32] Ebejer J. P., Morris G. M., Deane C. M. Freely available conformer generation methods: how good are they? // Journal of Chemical Information and Modeling. – 2012. – Vol. 52. – No. 5. – P. 1146-1158

[33] Efroymson M. A., In Mathematical Methods for Digital Computers / Ed. by A. Ralston, H.S. Wilf, Eds. Wiley: New York, 1960. P. 191-203.

[34] Ferrara P., Apostolakis J., Caflisch A. Evaluation of a fast implicit solvent model for molecular dynamics simulations // Proteins: Structure, Function, and Bioinformatics. – 2002. – Vol. 46. – No. 1. – P. 24-33.

[35] Fichthorn K. A., Weinberg W. H. Theoretical foundations of dynamical Monte Carlo simulations // The Journal of Chemical Physics. – 1991. – Vol. 95. – No. 2. – P. 1090-1096.

[36] Fraczkiewicz R., Braun W. Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules // Journal of Computational Chemistry. – 1998. – Vol. 19. – No. 3. – P. 319-333.

[37] FreeSASA [Электронный ресурс]. Доступ URL: https://freesasa.github.io/. (дата обращения: 04.05.2022).

[38] Principles of Molecular Simulation of Gas Transport in Polymers / Theodorou D. // Materials Science of Membranes for Gas and Vapor Separation / Ed. by Yu. Yampolskii, I. Pinnau, B.D. Freema. Wiley: Chichester, 2006. – P. 49-89.

[39] Molecular Simulation of Gas and Vapor Transport in Highly Permeable Polymers / Fried J. // Materials Science of Membranes for Gas and Vapor Separation / Ed. By Yu. Yampolskii, I. Pinnau, B.D. Freeman. Wiley: Chichester, 2006. – P. 95-127.

[40] Gasteiger J., Marsili M. Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges // Tetrahedron. – 1980. – Vol. 36. – No. 22. – P. 3219-3228.

[41] Gasteiger J., Marsili M. A new model for calculating atomic charges in molecules // Tetrahedron Letters. – 1978. – Vol. 19. – No. 34. – P. 3181-3184.

[42] Glaser T. D., Nguyen J. A., Anderson Strong scaling of general-purpose molecular dynamics simulations on GPUs //Computer Physics Communications. – 2015. – Vol. 192. – P. 97-107.

[43] Goubko M. V., MiloserdovO. A, Yampolskii Yu. P., Ryzhikh V. Ye. Prediction of Solubility Parameters of Light Gases in Glassy Polymers on the Basis of Simulation of a Short Segment of a Polymer Chain // Polymer Science, Series A. – 2019. – Vol. 61. – No. 5. – P. 718-732.

[44] Goubko M. V., MiloserdovO. A, Yampolskii Yu. P. et al. A novel model to predict infinite dilution solubility coefficients in glassy polymers // Journal of Polymer Science Part B: Polymer Physics. – 2017. – Vol. 55. – No. 3. – P. 228-244.

[45] Goubko M., Miloserdov O. Simple Alcohols with the Lowest Normal Boiling Point Using Topological Indices // MATCH Commun. Math. Comput. Chem. –2016, – Vol. 75, – No. 1. – P. 29-56.

[46] Graham T. XVIII. On the absorption and dialytic separation of gases by colloid septa //Philosophical transactions of the Royal Society of London. – 1866. – No. 156. – P. 399-439.

[47] Greenfield M. L., Theodorou D. N. Coarse-grained molecular simulation of penetrant diffusion in a glassy polymer using reverse and kinetic Monte Carlo // Macromolecules. – 2001. – Vol. 34. – No. 24. – P. 8541-8553.

[48] Gusev A. A., Suter U. W. Dynamics of small molecules in dense polymers subject to thermal motion // The Journal of Chemical Physics. – 1993. – Vol. 99. – No. 3. – P. 2228-2234.

[49] Halgren T. A. MMFF VI. MMFF94s option for energy minimization studies // Journal of Computational Chemistry. – 1999. – Vol. 20. – No. 7. – P. 720-729.

[50] Halgren T. A. MMFF VII. Characterization of MMFF94, MMFF94s, and other widely available force fields for conformational energies and for intermolecular-interaction energies and geometries // Journal of Computational Chemistry. – 1999. – Vol. 20. – No. 7. – P. 730-748.

[51] Halgren T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94 // Journal of Computational Chemistry. – 1996. – Vol. 17. – No. 5-6. – P. 490-519.

[52] Halgren T. A. Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions // Journal of Computational Chemistry. – 1996. – Vol. 17. – No. 5-6. – P. 520-552.

[53] Halgren T. A. Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94 // Journal of Computational Chemistry. – 1996. – Vol. 17. – No. 5-6. – P. 553-586.

[54] Halgren T. A. Merck molecular force field. V. Extension of MMFF94 using experimental data, additional computational data, and empirical rules // Journal of Computational Chemistry. – 1996. – Vol. 17. – No. 5-6. – P. 616-641.

[55] Halgren T. A., Nachbar R. B. Merck molecular force field. IV. Conformational

energies and geometries for MMFF94 // Journal of Computational Chemistry. – 1996. – Vol. 17. – No. 5-6. – P. 587-615.

[56] Hart K. E., Colina C. M. Estimating gas permeability and permselectivity of microporous polymers // Journal of Membrane Science. – 2014. – Vol. 468. – P. 259-268.

[57] Hasnaoui H., Krea M., Roizard D. Neural networks for the prediction of polymer permeability to gases // Journal of Membrane Science. – 2017. – Vol. 541. – P. 541-549.

[58] Heller S. R. et al. InChI, the IUPAC international chemical identifier // Journal of Cheminformatics. – 2015. – Vol. 7. – No. 1. – P. 1-34.

[59] Introduction / Encyclopedia of membrane science and technology // Hoek E. M. V., Tarabara V. V. (ed.). – Hoboken, NJ : Wiley, 2013. – Vol. 3. – P. 2219-2228.

[60] Instant JChem Chemaxon [Электронный ресурс]. Доступ URL: https://chemaxon.com/products/instant-jchem (дата обращения: 12.04.2022).

[61] Jørgensen P. B., Schmidt M. N., Winther O. Deep generative models for molecular science // Molecular Informatics. – 2018. – Vol. 37. – No. 1-2. – P. 1700133.

[62] Kanehashi S., Nagai K. Analysis of dual-mode model parameters for gas sorption in glassy polymers // Journal of Membrane Science. – 2005. – Vol. 253. – No. 1-2. – P. 117-138.

[63] Kanehashi S., Nagai K. Analysis of dual-mode model parameters for gas sorption in glassy polymers // Journal of Membrane Science. – 2005. – Vol. 253. – No. 1-2. – P. 117-138.

[64] Kirkpatrick P., Ellis C. Chemical space // Nature. – 2004. – Vol. 432. – No. 7019. – P. 823-824.

[65] LAMMPS. Molecular Dynamics Simulator [Электронный ресурс]. Доступ URL: https://lammps.sandia.gov/ (дата обращения: 01.05.2022)

[66] Leay L., Siperstein F. R. Single Polymer Chain Surface Area as a Descriptor for Rapid Screening of Microporous Polymers for Gas Adsorption // Adsorption Science & Technology. – 2013. – Vol. 31. – No. 1. – P. 99-112.

[67] Lee B., Richards F. M. The interpretation of protein structures: estimation of static accessibility // Journal of Molecular Biology. – 1971. – Vol. 55. – No. 3. – P. 379-400.

[68] Maine E., Garnsey E. Commercializing generic technology: The case of advanced materials ventures // Research Policy. – 2006. – Vol. 35. – No. 3. – P. 375-393.

[69] Marvin ChemAxon [Электронный ресурс]. Доступ URL: https://chemaxon.com/products/marvin. (дата обращения: 04.05.2022).

[70] Mayo S. L., Olafson B. D., Goddard W. A. DREIDING: a generic force field for molecular simulations // Journal of Physical Chemistry. – 1990. – Vol. 94. – No. 26. – P. 8897-8909.

[71] Mazo M., Balabaev N., Alentiev A., Yampolskii Yu. Molecular dynamics simulation of nanostructure of high free volume polymers with SiMe3 side groups // Macromolecules. – 2018. – Vol. 51. – No. 4. – P. 1398-1408.

[72] Mazo M., Balabaev N., Alentiev A. et al. Structure and properties of high and low free volume polymers studied by molecular dynamics simulation // Computation. – 2019. – Vol. 7. – No. 2. – P. 27-41.

[73] Merkel T. C., Pinnau I., Prabhakar R., Freeman B. Gas and vapor transport properties of perfluoropolymers // Materials Science of Membranes for Gas and Vapor Separation. – 2006. – Vol. 1. – P. 251-271.

[74] Miloserdov O.A. Classifying Amorphous Polymers for Membrane Technology Basing on Accessible Surface Area of Their Conformations // Advances in Systems Science and Applications. 2020. – Vol. 20, – No. 3. – P. 91-104

[75] Miloserdov O.A., Goubko M.V. QSPR method for prediction of sorption parameters of light gases in glassy polymers / Book of abstracts of the 32nd International Course and Conference on the Interfaces among Mathematics, Chemistry and Computer Sciences: Mathematics, Chemistry, Computing (Dubrovnik, 2021). Zagreb: Croatian Chemical Society, 2021. P. 18.

[76] Mitchell J. K. On the penetrativeness of fluids //Journal of Membrane Science. – 1995. – Vol. 100. – No. 1. – P. 11-16.

[77] Mitternacht S. FreeSASA: An open source C library for solvent accessible surface area calculations // F1000Research. – 2016. – Vol. 5. – P. 189-209.

[78] Mullard A. The drug-maker's guide to the galaxy // Nature. – 2017. – Vol. 549. – No. 7673. – P. 445-447.

[79] Nagai K., Masuda T., Nakagawa T. Poly [1-(trimethylsilyl)-1-propyne] and related polymers: synthesis, properties and functions // Progress in Polymer Science. – 2001. – Vol. 26. – No. 5. – P. 721-798.

[80] Neyertz S., Brown D. Molecular dynamics simulations of oxygen transport through a fully atomistic polyimide membrane // Macromolecules. – 2008. – Vol. 41. – No. 7. – P. 2711-2721.

[81] Norman G. E., Filinov V. S. Investigations of phase transitions by a Monte-Carlo method // High Temperature. – 1969. – Vol. 7. – No. 2. – P. 216-222.

[82] Park M., Salem D., Parviz D. et al. Measuring the accessible surface area within the nanoparticle corona using molecular probe adsorption // Nano Letters. – 2019. – Vol. 19. – No. 11. – P. 7712-7724.

[83] PoLyInfo [Электронный ресурс]. Доступ URL: https://polymer.nims.go.jp/ (дата обращения: 04.05.2022).

[84] Polymer Properties Database [Электронный ресурс]. Доступ URL: http://www.polymerdatabase.com (дата обращения: 04.05.2022).

[85] Polymers: A Property Database [Электронный ресурс]. Доступ URL: http://poly.chemnetbase.com/faces/polymers/PolymerSearch.xhtml (дата обращения: 04.05.2022).

[86] RDKit: Open-source cheminformatics [Электронный ресурс]. Доступ URL: http://www.rdkit.org (дата обращения: 04.05.2022).

[87] Rappé A. K., Casewit C. J., Colwell K. S. et al. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations // Journal of the American Chemical Society. – 1992. – Vol. 114. – No. 25. – P. 10024-10035.

[88] Reymond J. L. The chemical space project // Accounts of Chemical Research. – 2015. – Vol. 48. – No. 3. – P. 722-730.

[89] rdDistGeom module. Module containing functions to compute atomic coordinates in 3D using distance geometry [Электронный ресурс]. Доступ URL: https://www.rdkit.org/docs/source/rdkit.Chem.rdDistGeom.html (дата обращения: 04.05.2022).

[90] Riniker S., Landrum G. A. Better informed distance geometry: using what we know to improve conformation generation // Journal of Chemical Information and Modeling. – 2015. – Vol. 55. – No. 12. – P. 2562-2574.

[91] Robeson L. M. Correlation of separation factor versus permeability for polymeric membranes // Journal of Membrane Science. – 1991. – Vol. 62. – No. 2. – P. 165-185.

[92] Robeson L. M. The upper bound revisited // Journal of Membrane Science. – 2008. – Vol. 320. – No. 1-2. – P. 390-400.

[93] Ronova I. A. Rozhkov E. M., Alentiev A. Yu., Yampolskii Yu. P. Occupied and accessible volumes in glassy polymers and their relationship with gas permeation parameters // Macromolecular Theory and Simulations. – 2003. – Vol. 12. – No. 6. – P. 425-439.

[94] Ronova I. A., Sokolova E. A., Bruma M. Influence of chemical structure of the repeating unit on physical properties of aromatic polymers containing phenylquinoxaline rings // Journal of Polymer Science Part B: Polymer Physics. – 2008. – Vol. 46. – No. 17. – P. 1868-1877.

[95] Royal Geographical Society, 21st Century Challenges (2015) [Электронный ресурс]. Доступ URL: https://21stcenturychallenges.org/challenges/. (дата обращения: 04.05.2022).

[96] Ryzhikh V., TsarevbA D., Alentiev A., Yampolskii Yu. A novel method for predictions of the gas permeation parameters of polymers on the basis of their chemical structure // Journal of Membrane Science. – 2015. – Vol. 487. – P. 189-198.

[97] Sanner M. F., Olson A. J., Spehner J. C. Reduced surface: an efficient way to compute molecular surfaces // Biopolymers. – 1996. – Vol. 38. – No. 3. – P. 305-320.

[98] Schrödinger [Электронный ресурс]. Доступ URL:https://www.schrodinger.com/. (дата обращения: 04.05.2022).

[99] Shrake A., Rupley J. A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin // Journal of Molecular Biology. – 1973. – Vol. 79. – No. 2. – P. 351-371.

[100] Stanton D. T., Jurs P. C. Development and use of charged partial surface area structural descriptors in computer-assisted quantitative structure-property relationship studies // Analytical Chemistry. – 1990. – Vol. 62. – No. 21. – P. 2323-2329.

[101] Stepwise regression algorithm [Online]. Доступ URL: https://datascience.stackexchange.com/questions/937/does-scikit-learn-have-a-forward-selection-stepwise-regression-algorithm (дата обращения: 04.05.2022).

[102] Structure drawing software for academic and personal use. https://www.acdlabs.com/resources/freeware/chemsketch/. (дата обращения: 04.05.2022).

[103] TSNE sklearn [Электронный ресурс]. Доступ URL: https://scikit-learn.org/stable/modules/generated/sklearn.manifold.TSNE.html (дата обращения: 04.05.2022).

[104] Theodorou D. N. Understanding and predicting structure–property relations in polymeric materials through molecular simulations // Molecular Physics. – 2004. – Vol. 102. – No. 2. – P. 147-166.

[105] Thompson A. P., Aktulga H. M., Berger R. et al. LAMMPS-a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales // Computer Physics Communications. – 2022. – Vol. 271. – P. 108171.

[106] Todeschini R., Consonni V. Handbook of molecular descriptors. – John Wiley & Sons, 2008.

[107] Virshup A. M., Contreras-García J., Wipf P. et al. Stochastic voyages into uncharted chemical space produce a representative library of all possible drug-like compounds // Journal of the American Chemical Society. – 2013. – Vol. 135. – No. 19. – P. 7296-7303.

[108] Weininger D., Weininger A., Weininger J. L. SMILES. 2. Algorithm for generation of unique SMILES notation // Journal of Chemical Information and Computer Sciences. – 1989. – Vol. 29. – No. 2. – P. 97-101.

[109] Weiser J., Shenkin P. S., Still W. C. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO) // Journal of Computational Chemistry. – 1999. – Vol. 20. – No. 2. – P. 217-230.

[110] v. Wroblewski S. Ueber die natur der absorption der gase //Annalen der Physik. – 1879. – Vol. 244. – No. 9. – P. 29-52.

[111] Yampolskii Y. Polymeric gas separation membranes // Macromolecules. – 2012. – Vol. 45. – No. 8. – P. 3298-3311.

[112] Yampolskii Y., Shishatskiy S., Alentiev A., Loza K. Correlations with and prediction of activation energies of gas permeation and diffusion in glassy polymers // Journal of Membrane Science. – 1998. – Vol. 148. – No. 1. – P. 59-69.

[113] Yampolskii Y., Shishatskiy S., Alentiev A., Loza K. Group contribution method for transport property predictions of glassy polymers: focus on polyimides and polynorbornenes // Journal of Membrane Science. – 1998. – Vol. 149. – No. 2. – P. 203-220.

[114] Yampolskii Y., Wiley D., Maher C. Novel correlation for solubility of gases in polymers: effect of molecular surface area of gases // Journal of Applied Polymer Science. – 2000. – Vol. 76. – No. 4. – P. 552-560.

Приложение 1. Акт о внедрении результатов диссертационной работы

«УТВЕРЖДАЮ»
Директор
Федерального государственного
бюджетного учреждения науки
Ордена Трудового Красного Знамени
Институт нефтехимического синтеза
имени А. В. Топчиева
Российской академии наук
чл.-корр. РАН, д.х.н. Максимов А. Л.

«__» _____ 2022 г.

АКТ
о внедрении результатов диссертационной работы
Милосердова Олега Александровича на тему «Математическое моделирование полимерных цепей в задачах предсказания транспортных характеристик стеклообразных полимеров» в работы, проводимые в Институте нефтехимического синтеза им. А. В. Топчиева Российской академии наук (ИНХС РАН)

Мы, нижеподписавшиеся сотрудники ИНХС РАН: и.о. заведующего лабораторией Мембранного газоразделения к.х.н. Белов Н. А., в.н.с., д.х.н., проф., Алентьев А. Ю., м.н.с., к.х.н. Рыжих В. Е. составили настоящий акт о том, что разработанный метод прогнозирования транспортных свойств стеклообразных полимеров «Предсказания на основе Поверхности Коротких Полимерных Цепей» и реализованный на его основе программный комплекс для прогнозирования транспортных свойств стеклообразных полимеров, являющиеся результатами диссертационной работы Милосердова Олега Александровича, были использованы в исследованиях лаборатории Мембранного газоразделения при реализации:

- Программы фундаментальных исследований ИНХС за 2017г. по теме «Мембранное разделение и мембранный катализ в химии, энергетике, экологии: новые мембранные материалы, высокопроизводительные мембраны и процессы на их основе»; шифр 45, 47; Госуд. рег. № АААА-А18-118011990199-9;
- Программы фундаментальных исследований ИНХС за 2018-2019г. по теме «Новые материалы и высокопроизводительные мембраны для разделения жидких и газовых смесей; мембранный катализ в химических процессах получения продуктов высокой чистоты»; шифр 45, 47; Госуд. рег. № АААА-А19-119020490055-4;
- Гранта РФФИ № 17-08-00164 «Компьютерное моделирование наноструктуры мембранных материалов: традиционные и новые подходы».

И.о. заведующего лабораторией
Мембранного газоразделения
старший научный сотрудник,
кандидат химических наук — Белов Н. А.

Ведущий научный сотрудник,
доктор химических наук, профессор — Алентьев А. Ю.

Младший научный сотрудник,
кандидат химических наук — Рыжих В. Е.

Приложение 2. Свидетельство о государственной регистрации программы для ЭВМ

3379

РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2022666110

Программный комплекс для прогнозирования транспортных свойств стеклообразных полимеров на основе метода «Предсказания на основе Поверхности Коротких Полимерных Цепей»

Правообладатель: ***Федеральное государственное бюджетное учреждение науки Институт проблем управления им. В.А. Трапезникова Российской академии наук (RU)***

Автор(ы): ***Милосердов Олег Александрович (RU)***

Заявка № **2022662979**
Дата поступления **11 июля 2022 г.**
Дата государственной регистрации
в Реестре программ для ЭВМ ***25 августа 2022 г.***

Руководитель Федеральной службы
по интеллектуальной собственности

Ю.С. Зубов