

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

На правах рукописи

Владимирова Надежда Игоревна

**Подходы QSPR в аналитической химии: *in silico* прогнозирование свойств
ионоселективных электродов и глубоких эвтектических растворителей**

Научная специальность 1.4.2. Аналитическая химия

Диссертация на соискание ученой степени кандидата химических наук

Научный руководитель:
доктор химических наук, профессор
Кирсанов Дмитрий Олегович

Санкт-Петербург

2024

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	5
Глава 1. ЛИТЕРАТУРНЫЙ ОБЗОР	14
1.1. Поиск количественных соотношений структура-свойство: теоретические основы, методология и применение в химии.....	14
1.2. Deskрипторы	19
1.2.1. Структурные deskрипторы	19
1.2.2. Геометрические (3D) deskрипторы.....	23
1.2.3. Квантово-химические deskрипторы	26
1.2.4. Электростатические deskрипторы	27
1.3. Математические методы для обработки данных.....	29
1.3.1. Метрики и валидация.....	30
1.3.2. Множественная линейная регрессия	33
1.3.3. Проекция на латентные структуры	34
1.3.4. Метод опорных векторов.....	38
1.3.5. Нейронные сети.....	39
1.4. Применение QSPR в аналитической химии	42
1.5. Цели и задачи работы	46
Глава 2. ПРИМЕНЕНИЕ QSPR-МОДЕЛИРОВАНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ СЕЛЕКТИВНОСТИ КАРБОНАТНЫХ ИОНОСЕЛЕКТИВНЫХ ЭЛЕКТРОДОВ	51
2.1. Экспериментальная часть	52
2.1.1. Выбор deskрипторов	52
2.1.2. Ионоселективные сенсоры с новыми лигандами в составе	54
2.1.3. QSPR моделирование.....	57
2.2. Обсуждение результатов	59

Заключение к Главе 2	67
Глава 3. ПРИМЕНЕНИЕ QSPR-МОДЕЛИРОВАНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ ЧУВСТВИТЕЛЬНОСТИ ИОНОСЕЛЕКТИВНЫХ ЭЛЕКТРОДОВ К КАТИОНАМ ТЯЖЕЛЫХ МЕТАЛЛОВ (Cu^{2+} , Cd^{2+} И Pb^{2+})	69
3.1. Экспериментальная часть	70
3.1.1. Выбор дескрипторов	70
3.1.2. QSPR моделирование	71
3.2. Обсуждение результатов.....	72
Заключение к Главе 3	81
Глава 4. ПРИМЕНЕНИЕ QSPR-МОДЕЛИРОВАНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ ФИЗИКО-ХИМИЧЕСКИХ СВОЙСТВ ГЛУБОКИХ ЭВТЕКТИЧЕСКИХ РАСТВОРИТЕЛЕЙ НА ОСНОВЕ ХОЛИН ХЛОРИДА	83
4.1. Экспериментальная часть	84
4.1.1. Выбор дескрипторов	84
4.1.2. QSPR моделирование	86
4.2. Обсуждение результатов	89
4.2.1. Плотность	89
4.2.2 Проводимость.....	92
4.2.3. Вязкость.....	94
4.2.4. Коэффициент преломления	96
4.2.5. Статистическая проверка моделей.....	98
4.2.6. Оценка прогностической эффективности моделей.....	100
Заключение к Главе 4	102
ЗАКЛЮЧЕНИЕ	103
СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ.....	105
СПИСОК ЛИТЕРАТУРЫ.....	106

ПРИЛОЖЕНИЕ 115

ВВЕДЕНИЕ

В современной науке наблюдается стремительное развитие вычислительных подходов, которые всё чаще используются для решения сложных химических задач. Одним из таких подходов является поиск количественных соотношений структура-свойство (Quantitative Structure-Property Relationships, QSPR), позволяющий прогнозировать свойства новых соединений и материалов на основе их молекулярной структуры. QSPR основан на использовании математических моделей, связывающих молекулярные дескрипторы – характеристики, описывающие структуру молекулы в численном виде – с интересующими свойствами. Этот подход значительно ускоряет и удешевляет процесс разработки новых веществ и материалов с заданными характеристиками, снижая потребность в дорогостоящих и трудоемких экспериментах. QSPR позволяет исследователям получать ценную информацию о свойствах соединений, минуя этап их синтеза и экспериментального тестирования, что помогает в разработке новых лекарственных средств, растворителей, всевозможных материалов. Этот метод находит широкое применение в различных областях, включая фармацевтику, экологию, материаловедение и химическую инженерию.

Классические подходы QSPR требуют больших объёмов данных для построения надежных моделей, обычно речь идет о тысячах соединений, структура и свойства которых используются для создания математической модели. Это существенно затрудняет применение QSPR в аналитической химии, где возможность подбора больших представительных однородных наборов данных весьма ограничена. Например, для того чтобы прогнозировать аналитические характеристики потенциометрических сенсоров с полимерными пластифицированными мембранами на основе структурной формулы ионофора, необходим набор экспериментальных данных по сенсорам, изготовленным на основе структурно схожих ионофоров. При этом составы мембран (полимер,

растворитель-пластификатор, ионообменная добавка) должны быть идентичными в этой обучающей выборке. Также должны совпадать и условия проведения потенциометрических измерений (концентрационные диапазоны определения чувствительностей, набор мешающих ионов для определения селективности), и экспериментальные протоколы определения характеристик сенсоров. Очевидно, что различные научные группы, работающие в этой области, проводят эксперименты несколько по-разному, и прямое сопоставление результатов, полученных разными исследователями, не всегда возможно. В таких условиях создание обучающей выборки для QSPR, содержащей хотя бы несколько десятков соединений, охарактеризованных в схожих условиях, представляется серьезной проблемой. Схожие ограничения существуют и во многих других областях аналитической химии, например, при разработке экстрагентов с требуемыми свойствами. Преодоление таких ограничений и исследование потенциала QSPR для разработок новых материалов в аналитической химии является важной и актуальной задачей.

В настоящей работе для изучения возможностей QSPR в решении задач аналитической химии выбраны два популярных современных направления исследований: разработка потенциометрических сенсоров и разработка глубоких эвтектических растворителей (ГЭР) с требуемыми свойствами. Для исследований перспектив использования QSPR в вышеупомянутых направлениях были определены конкретные задачи, среди которых исследование возможности *in silico* прогнозирования чувствительности и селективности потенциометрических сенсоров с полимерными пластифицированными мембранами на основе ионофоров и на возможности прогнозирования плотности, проводимости и вязкости ГЭР. В качестве модельных систем для исследований были выбраны сенсоры на основе фосфорил ацетамидов с выраженной чувствительностью к ионам тяжелых металлов, карбонат-селективные сенсоры и ГЭР на основе холин хлорида.

Актуальность исследования

Разработка новых подходов к QSPR-моделированию, дающих возможность построения прогностических моделей на основе ограниченного набора обучающих данных, является крайне востребованным инструментом для решения актуальных задач аналитической химии, однако не исследованным в должной степени до настоящего времени. Предложенный подход может быть применен к широкому кругу аналитических задач, в данной работе было выделено 2 направления, актуальность которых продиктована следующими соображениями. Разработка новых ионоселективных электродов (ИСЭ) для определения тяжёлых металлов и гидрофильных анионов является важной задачей в области создания инструментов для экологического мониторинга, контроля качества пищевых продуктов, медицинской диагностики. QSPR-моделирование может значительно ускорить и оптимизировать этот процесс, позволяя прогнозировать свойства сенсоров на основе структуры ионофоров и других компонентов мембраны, что сократит время и затраты на разработку новых ИСЭ.

Прогнозирование физико-химических свойств глубоких эвтектических растворителей также является крайне актуальной проблемой. ГЭР представляют собой новый класс «зелёных» растворителей с уникальными свойствами, которые делают их перспективными для применения в различных областях, таких как «зелёная» химия, фармацевтика и энергетика. QSPR-моделирование позволит прогнозировать свойства ГЭР на основе их состава, что существенно упростит поиск оптимальных составов для конкретных применений, способствуя развитию устойчивых технологий и снижению негативного воздействия на окружающую среду.

Цели и задачи исследования

Цель данного исследования заключается в разработке и проверке QSPR-моделей, адаптированных для работы с ограниченными наборами данных, для

прогнозирования свойств ионоселективных электродов и глубоких эвтектических растворителей.

Для достижения этой цели поставлены следующие **задачи**:

1. Разработка QSPR-моделей для прогнозирования селективности потенциометрических сенсоров к гидрофильным анионам карбоната на основе ионофоров с акцепторными заместителями вблизи карбонильной группы.

2. Разработка QSPR-моделей для прогнозирования чувствительности потенциометрических сенсоров к катионам тяжелых металлов (Ca^{2+} , Cd^{2+} , Pb^{2+}) на основе дифенилфосфорил ацетамидных ионофоров.

3. Разработка QSPR-моделей для прогнозирования плотности, проводимости и вязкости ГЭР на основе холин хлорида с органическими кислотами в качестве доноров водородной связи.

Научная новизна исследования

Впервые предложены и валидированы QSPR-модели для прогнозирования селективности ИСЭ к гидрофильным анионам (карбонат-ионы) на основе ограниченных обучающих наборов данных.

Предложены новые подходы к моделированию чувствительности ИСЭ к катионам тяжелых металлов (Ca^{2+} , Cd^{2+} , Pb^{2+}) на основе подструктурных молекулярных фрагментов с использованием ограниченных обучающих наборов данных, модели валидированы на независимых тестовых наборах соединений.

Впервые предложены и проверены QSPR-модели для прогнозирования физико-химических свойств ГЭР на основе холин хлорида с органическими кислотами в качестве доноров водородной связи. Модели валидированы на независимых тестовых наборах соединений.

Для всех QSPR-моделей выявлены ключевые молекулярные дескрипторы, оказывающие значимое влияние на свойства соединений, что способствует развитию подходов к получению материалов с заданными свойствами.

Показана эффективность QSPR подходов при работе с малыми обучающими выборками и предложены способы валидации QSPR-моделей в условиях ограниченной доступности соединений с экспериментально установленными целевыми свойствами.

Практическая значимость исследования

Демонстрация возможности успешного применения QSPR-подхода для прогнозирования свойств объектов аналитической химии, для которых характерны ограниченные наборы данных.

Разработка QSPR-моделей для направленного дизайна ионофоров, обеспечивающих заданные характеристики (селективность к карбонат-ионам и чувствительность к катионам тяжелых металлов) ионоселективных электродов, существенно облегчает поиск новых лигандов для ИСЭ.

Разработка QSPR-модели для прогнозирования физико-химических свойств глубоких эвтектических растворителей на основе холин хлорида с органическими кислотами позволяет подбирать ГЭР с необходимыми свойствами под конкретную задачу, минуя трудоемкие этапы экспериментального исследования их свойств.

На защиту выносятся следующие положения

QSPR-модели, построенные на основе подструктурных молекулярных фрагментов, описывающих ионофоры для ИСЭ, с использованием ограниченных обучающих наборов данных, обеспечивают прогнозирование селективности потенциометрических сенсоров к карбонат-аниону с погрешностью не более 1,5 $\log K^{\text{сел}}(\text{HCO}_3^- / \text{Cl}^-)$.

1. QSPR-модели, построенные на основе подструктурных молекулярных фрагментов, описывающих фосфорил ацетамидные ионофоры, обеспечивают прогнозирование чувствительности ионоселективных сенсоров к катионам тяжелых металлов (Ca^{2+} , Cd^{2+} , Pb^{2+}) с погрешностью не более 8 мВ/дек.

2. QSPR-модели, построенные на основе подструктурных молекулярных фрагментов и полуэмпирических дескрипторов, рассчитанных с помощью квантово-химического метода PM3, с использованием ограниченных обучающих наборов данных, обеспечивают прогнозирование плотности, вязкости и электропроводности ГЭР на основе холин хлорида с органическими кислотами в качестве доноров водородной связи. При этом погрешности в оценке не превышают 0,065 г/см³ для плотности, 4,6 мСм/см для проводимости, 3,6 мПа·с для вязкости.

Основные результаты работы

Результаты были представлены на следующих научных мероприятиях: 1-ая Международная конференция The 1st International Electronic Conference on Chemical Sensors and Analytical Chemistry; XII Международная конференция молодых ученых «Mendeleev 2021», г. Санкт-Петербург; 13-ый Международный симпозиум WSC-13, г. Москва; 2-ая Международная конференция The 2nd International Electronic Conference on Chemical Sensors and Analytical Chemistry; а также в виде трёх публикаций в рецензируемых международных журналах, индексируемых Scopus и Web Of Science [1–3]:

1. Vladimirova, N., Polukeev, V., Ashina J., Babain V., Legin, A., Kirsanov, D. Prediction of Carbonate Selectivity of PVC-Plasticized Sensor Membranes with Newly Synthesized Ionophores through QSPR Modeling (2022) *Chemosensors*, 10(2), 43. DOI: 10.3390/chemosensors10020043

2. Vladimirova N., Puchkova E., Dar'in D., Turanov A., Babain V., Kirsanov D. Predicting the Potentiometric Sensitivity of Membrane Sensors Based on Modified Diphenylphosphoryl Acetamide Ionophores with QSPR Modeling (2022) *Membranes*, 12(10), 953. DOI: 10.3390/membranes12100953

3. Vladimirova N., Bochko T., Shishov A., Dmitry Kirsanov. Predicting the properties of deep eutectic solvents based on choline chloride and carboxylic acids and

their mixtures with water using QSPR approach (2024) *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 692. DOI: 10.1016/j.colsurfa.2024.133961

Основные научные результаты:

1. Vladimirova N., Puchkova E., Dar'in D., Turanov A., Babain V., Kirsanov D. Predicting the Potentiometric Sensitivity of Membrane Sensors Based on Modified Diphenylphosphoryl Acetamide Ionophores with QSPR Modeling (2022) *Membranes*, 12(10), 953. DOI: 10.3390/membranes12100953

В данном исследовании была изучена применимость метода QSPR для прогнозирования потенциометрической чувствительности пластифицированных полимерных мембранных сенсоров на основе химических структур ионофоров, описанных подструктурными молекулярными дескрипторами, в качестве исходных данных модели. Модель QSPR была обучена на литературных данных о чувствительности к тяжёлым металлам ранее изученных структурно схожих ионофоров и показала высокую прогностическую способность для чувствительности ИСЭ к Cu^{2+} , Cd^{2+} и Pb^{2+} . Прогнозы модели для четырёх новых ионофоров на основе дифенилфосфорилацетамида были сопоставлены с потенциометрическими экспериментальными данными для этих ионофоров, было отмечено удовлетворительное согласие.

Соискатель внес значительный вклад в данное исследование, проведя сбор и анализ литературных данных, формализацию структур ионофоров, построение QSPR-модели, анализ полученных результатов и интерпретацию вклада различных дескрипторов в чувствительность сенсоров. Соискатель также активно участвовал в подготовке научной публикации.

2. Vladimirova N., Bochko T., Shishov A., Kirsanov D. Predicting the properties of deep eutectic solvents based on choline chloride and carboxylic acids and their mixtures with water using QSPR approach (2024) *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 692. DOI: 10.1016/j.colsurfa.2024.133961

В работе исследован подход QSPR для прогнозирования свойств ГЭР. На основе структур органических кислот, входящих в состав ГЭР, построены регрессионные модели. Модели связывали свойства растворителей и молекулярные дескрипторы соответствующих карбоновых кислот, представленные подструктурными фрагментами и квантово-химическими дескрипторами. Были установлены статистически значимые корреляции между дескрипторами кислот плотностью, вязкостью и электропроводностью ГЭР. Было проанализировано влияние различных дескрипторов на эти свойства на основе регрессионных коэффициентов. Учитывая достаточно ограниченный набор данных, с помощью вложенной перекрёстной валидации была проведена оценка статистической достоверности регрессионных моделей.

Соискатель принимал самое активное участие в данном исследовании, выполнил сбор и анализ данных, формализацию структур карбоновых кислот с помощью подструктурных молекулярных фрагментов и квантово-химических дескрипторов, построение и валидацию QSPR-моделей, а также интерпретацию полученных результатов. Соискатель также подготовил первоначальный черновик научной публикации.

3. Vladimirova, N., Polukeev, V., Ashina J., Babain V., Legin, A., Kirsanov, D. Prediction of Carbonate Selectivity of PVC-Plasticized Sensor Membranes with Newly Synthesized Ionophores through QSPR Modeling (2022) Chemosensors, 10(2), 43. DOI: 10.3390/chemosensors10020043

В данной работе исследован потенциал подхода QSPR для прогнозирования потенциометрической селективности пластифицированных полимерных мембранных сенсоров на основе новых лигандов. В качестве объекта исследования рассмотрены сенсоры с селективностью к карбонату. Используя данные по селективности для 40 ионофоров, имеющих в литературе, и их подструктурные молекулярные фрагменты в качестве дескрипторов, была построена QSPR-модель, которая продемонстрировала достаточную точность в

прогнозировании селективности для новых лигандов, имеющих сходные молекулярные фрагменты с лигандами, используемыми для обучения модели.

Соискатель внес значительный вклад в эту работу, проведя сбор и анализ литературных данных, формализацию структур ионофоров, построение QSPR-модели, анализ результатов моделирования, в том числе установление вклада в модель подструктурных молекулярных дескрипторов, и экспериментальное определение селективности сенсоров с новыми ионофорами в составе. Соискатель также активно участвовал в подготовке научной публикации.

Глава 1. ЛИТЕРАТУРНЫЙ ОБЗОР

1.1. Поиск количественных соотношений структура-свойство: теоретические основы, методология и применение в химии

Для современных химических исследований актуальной задачей является поиск соединений с заранее заданными, желаемыми свойствами. Традиционный экспериментальный подход, в ходе которого предлагается гипотеза, а затем выполняется серия экспериментов, хоть и обладает долгой историей и является основой понимания химических явлений, часто сталкивается с ограничениями, когда речь заходит об эффективности отбора структур с заданными свойствами. Во-первых, цикл исследований - формулирование гипотез, проведение экспериментов, наблюдение за результатами и уточнение процедур часто приводит к долгому достижению желаемого результата. Каждый шаг, основанный на предыдущих знаниях, теоретических основах и выдвинутых новых предположениях, требует строгого соблюдения методики эксперимента, анализа, а затем корректировки с учётом последних результатов. Кроме того, каждая серия экспериментов требует материальных затрат и труда экспериментаторов. Наконец, проведённый синтез не гарантирует, что полученное соединение или материал будет иметь желаемые свойства.

С развитием вычислительной техники всё более важным дополнением к экспериментальной работе становятся вычислительные подходы. Эти вычислительные методы, включая квантовые расчёты, моделирование молекулярной динамики и поиск количественных соотношений структура-свойство, позволяют решать как теоретические задачи вроде исследования механизмов взаимодействия веществ, так и практические задачи вроде прогнозирования свойств неисследованных ранее соединений. Такой расчёт даёт возможность предварительной оценки новых соединений и материалов, который

осуществляется быстрее и со значительно меньшими материальными затратам, чем эксперимент, проведённый в лаборатории.

QSPR – вычислительный подход, позволяющий установить связь между химической структурой молекулы и её свойствами. Общая идея этого подхода заключается в расчёте с помощью различных вычислительных методов математической зависимости (модели) между структурой соединения и её свойством, и дальнейшем использовании этой модели для расчёта свойств новых, не используемых при составлении модели молекул. Вычисление осуществляется на основе параметров, представляющих информацию о свойствах соединений в числовом виде – дескрипторов. Общая схема подхода представлена на Рисунке 1, где отображается, что расчёт связи структуры соединения и его свойства осуществляется через стадии сбора дескрипторов и дальнейшего их статистического анализа.



Рисунок 1. Схема методологии, используемой в работах, применяющих подход QSPR

Идеи, лежащие в основе QSPR, кажутся интуитивно понятными, и они обсуждались задолго до официального появления этого термина. Первая работа, в которой эти идеи наиболее явно сформулированы – статья Александра Крама Брауна и Томаса Р. Фрейзера [4], в которой утверждается, что должна существовать связь между химической структурой и физиологическим действием вещества. В таком случае, если структура изменяется на ΔC , а свойства при этом изменяются на $\Delta \Phi$, тогда становится возможным найти зависимость между C и Φ .

Официально термин «QSPR» в 1960-х годах предложил Корвин Ханш [5]. Работы Ханша и его группы относятся к области биохимии [6,7], где подход QSPR часто называют QSAR – Quantitative Structure-Activity Relationships (поиск количественных соотношений структура-активность).

Со времени работ Ханша и по настоящее время именно расчёт активности соединений представляет собой наиболее широкую область применения QSPR, в которой отдельно нужно выделить такое направление как поиск новых лекарств. Например, в одной из недавних работ [8] была предложена структура производного хиноксалина, нового лекарства от лешманиоза. Новое вещество в ходе этой работы было не только предложено с помощью анализа дескрипторов, программы OSIRIS Property Explorer и правил Вебера и Липинского, но его свойства также были спрогнозированы моделью QSPR. После синтеза соединения с предложенной структурой и его исследования, оказалось, что спрогнозированное свойство – показатель концентрации полумаксимального ингибирования близко к полученному экспериментально (процентное отклонение спрогнозированного значения от полученного экспериментальным путём составило менее 4%).

С помощью подхода QSAR были оценены структуры новых кумарин-изоксазол-сульфонамидных гибридных соединений в качестве антибиотиков [9], а также были изучены производные 1H-пиразола новой вариацией QSAR, 5D-QSAR, для выявления наилучшего ингибитора рецептора эпидермального фактора роста, такие ингибиторы применяются при лечении рака [10].

Почему именно поиск новых лекарств представляет собой столь значительную область применений QSPR? Всё дело в особенностях подхода: во-первых, важнейшее преимущество QSPR – значительное снижение затрат, связанных с разработкой новых материалов и соединений. Традиционные экспериментальные подходы к разработке новых материалов зачастую оказываются ресурсоемки и требуют значительных финансовых, трудовых и временных затрат. Однако модели, рассчитанные с помощью QSPR, позволяют прогнозировать свойства соединений до начала эксперимента. Особенно это важно в случае, когда измерения продолжительны или сложно осуществимы с технической точки зрения.

Кроме того, QSPR играет важную роль в сокращении сроков разработки новых материалов. Расчёт свойств соединений до эксперимента ускоряет цикл разработки, синтеза и испытаний благодаря выявлению перспективных кандидатов и отсеву неподходящих. Это особенно полезно в быстроразвивающихся отраслях, где скорость выхода на рынок имеет решающее значение.

Наконец, в области разработки продуктов, требующих тестов на животных, способность QSPR предоставлять прогностические данные уменьшает количество испытаний, что важно с этической точки зрения.

Благодаря всем этим преимуществам QSPR успешно применяется во множестве областей химии, ранее упомянутый поиск лекарств – только один из примеров применения.

Так, в области химии окружающей среды (экологической химии) QSPR используется для прогнозирования поведения химических веществ с точки зрения экологии, включая биоразлагаемость [11,12], токсичность для водных организмов [13,14], потенциал к биоаккумуляции [15,16] и оценку стойкости загрязнений [17]. Учёт этих факторов способствует разработке более безопасных химических веществ и соблюдению нормативных требований. Похожие цели преследуют

исследования, применяющие QSPR в области агрохимии, где прогнозируют токсичность пестицидов [18,19], гербицидов [20,21] и фунгицидов [22,23], помогая разрабатывать соединения с повышенной эффективностью и сниженным воздействием на окружающую среду.

Подход QSPR широко применяется для прогнозирования свойств новых материалов, включая полимеры [24,25], наноматериалы [26,27] и композиты [28], эти материалы затем применяются в таких областях как электроника, накопление энергии и фотовольтаика. Много исследований посвящено разработке новых катализаторов, поскольку полученные модели QSPR помогают рассчитать их активность [29,30].

В пищевой химии модели QSPR позволяют спрогнозировать вкус [31,32], аромат [33] и стабильность пищевых добавок и консервантов [34], способствуя разработке новых продуктов питания и повышая их безопасность.

В косметической промышленности модели QSPR прогнозируют абсорбцию химических веществ через кожу [35], комедогенность [36] и риск раздражения [37] от новых компонентов, давая возможность для разработки более безопасных и эффективных косметических продуктов.

В химической инженерии QSPR помогает в разработке и оптимизации химических процессов путем прогнозирования термодинамических свойств и реакционной способности исходных веществ и продуктов [38,39], значительное количество работ посвящено прогнозированию свойств ионных жидкостей [40,41].

В нефтяной и топливной промышленности подход QSPR используется для прогнозирования свойств горюче-смазочных материалов [42–44], таких как вязкость, температура вспышки и эффективность сгорания, что помогает в разработке более эффективных и экологически чистых составов топлива.

1.2. Дескрипторы

В подходе QSPR концепция дескрипторов играет ключевую роль и связывает структуру молекулы и ее физические, химические или биологические свойства. В общем смысле дескрипторы – это числовые значения, которые количественно описывают какой-либо аспект молекулы, чаще всего молекулярную структуру, и служат основой для построения прогностических моделей. Разнообразие и выбор этих дескрипторов имеют решающее значение для точности и применимости моделей QSPR, поскольку напрямую влияют на прогностическую способность модели. Вне зависимости от природы дескрипторов, к ним предъявляются следующие обязательные требования [45]: независимость от маркировки и нумерации атомов, независимость от вращения молекулы, определение однозначным алгоритмом, чётко определенная применимость к молекулярным структурам. Кроме того, желательно, чтобы дескрипторы были простыми и могли применяться к широкому набору молекул.

Большую часть дескрипторов можно отнести к следующим категориям: структурные, геометрические (3D), электростатические и квантовые.

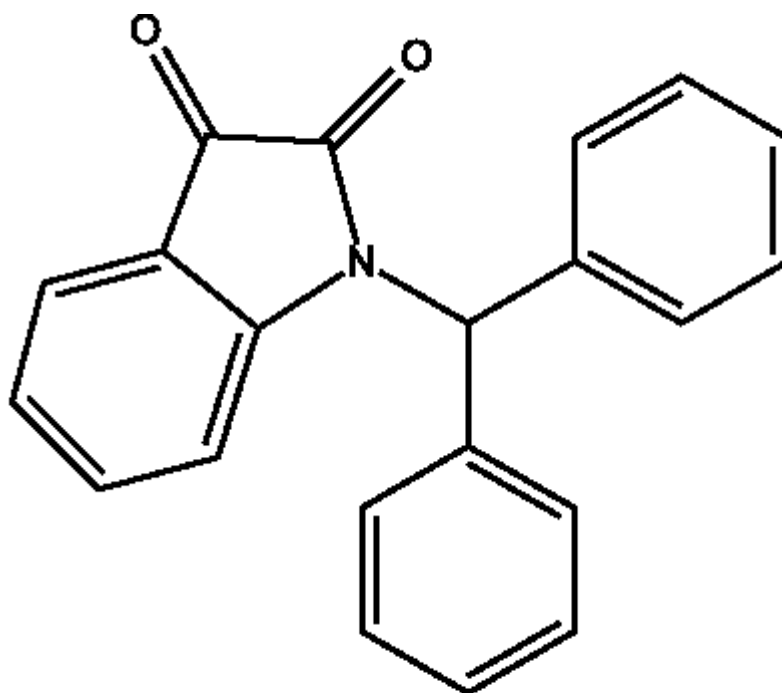
1.2.1. Структурные дескрипторы

Структурные дескрипторы -- это наиболее простые дескрипторы, количественно определяющие структурные характеристики молекул без углубления в их электронные или пространственные характеристики. К структурным дескрипторам относятся количество атомов определенного типа, количество связей и наличие определенных функциональных групп. Их легко рассчитать, и они предоставляют базовую, но важную информацию о размере и составе молекулы. Для получения структурных дескрипторов существуют различные методы и инструменты, каждый из которых обладает уникальными

преимуществами и возможностями применения.

Среди методов выделяются:

- SMILES (Simplified Molecular-Input Line-Entry System, Упрощенная система линейного ввода молекулярных данных): SMILES - это нотация, представляет химическую структуру в виде строки текста. Она описывает структуру на основе последовательности атомов и типов связей между ними. Строки SMILES могут быть легко преобразованы в числовые дескрипторы с помощью различных программных средств, включая широко используемый ChemDraw, что делает их очень удобными для исследований с применением подхода QSPR. На Рисунке 2 представлена структура молекулы 1-бензгидрилиндолин-2,3-диона и её запись в SMILES.



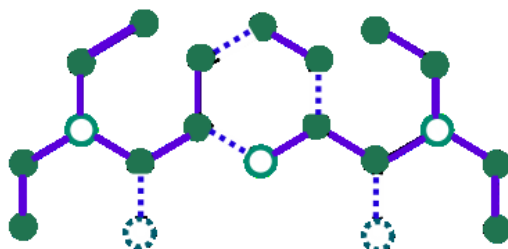
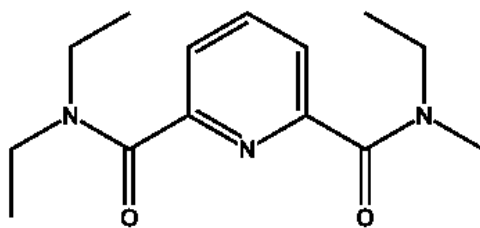
1-бензгидрилиндолин-2,3-дион

O=C(C1=CC=CC=C12)N2C(C3=CC=CC=C3)C4=CC=CC=C4

Рисунок 2. Пример записи структуры молекулы с помощью SMILES

- Молекулярные графы – подход, представляющий молекулу в виде графа, где каждая вершина обозначается соответствующим химическим элементом (например, С = углерод, Н = водород), а каждое ребро - типом ковалентной связи (одинарная -, двойная =, тройная, ароматическая). Атомы водорода, как правило, опускаются в таком представлении молекулы, так как валентность каждого атома известна. Применяя этот подход, можно получить различные дескрипторы на основе графов, такие как матрицы смежности или топологические индексы. На Рисунке 3 представлена молекула N2,N2,N6,N6-тетраэтилпиридин-2,6-дикарбоксамида в виде графа. Для нахождения дескрипторов на основе графов используются различные как бесплатные, так и коммерческие программы. В числе наиболее популярных программ проприетарный (имеющий платную лицензию) DRAGON и находящиеся в открытом доступе PaDEL и ISIDA. Последняя разработана доктором химических наук, ведущим научным сотрудником лаборатории новых физико-химических проблем ИФХЭ РАН Соловьёвым Виталием Петровичем.

**N2,N2,N6,N6-тетраэтилпиридин
-2,6-дикарбоксамид**



Вершины	Символ	Валентность
	O	2
	N	3
	C	4



Рёбра	Химическая связь
	одинарная
	двойная

Рисунок 3. Представление молекулы N2,N2,N6,N6-тетраэтилпиридин-2,6-дикарбоксиамида в виде графа. Различные типы узлов в соответствии с атомными элементами и различные типы ребер в зависимости от химической связи в молекуле

- Молекулярные отпечатки пальцев (molecular fingerprints) – это двоичные или численные векторы, которые отражают наличие или отсутствие (а иногда и количество) определенных субструктур или паттернов в молекуле. Существует несколько типов отпечатков, включая отпечатки на основе пути, отпечатки с расширенным соединением (ECFP, Extended-Connectivity Fingerprints) [46] и

ключи молекулярной системы доступа (MACCS, Molecular ACCess System). На Рисунке 4 представлены записи структур с помощью MACCS и ECFP. Одна ячейка отпечатка называется битом. Бит MACCS равен 1, если молекула содержит соответствующую биту подструктуру; в противном случае он равен 0. ECFP ищет относительное положение каждого атома в молекуле с помощью слоя связей l и записывает структурную информацию в битовый массив.

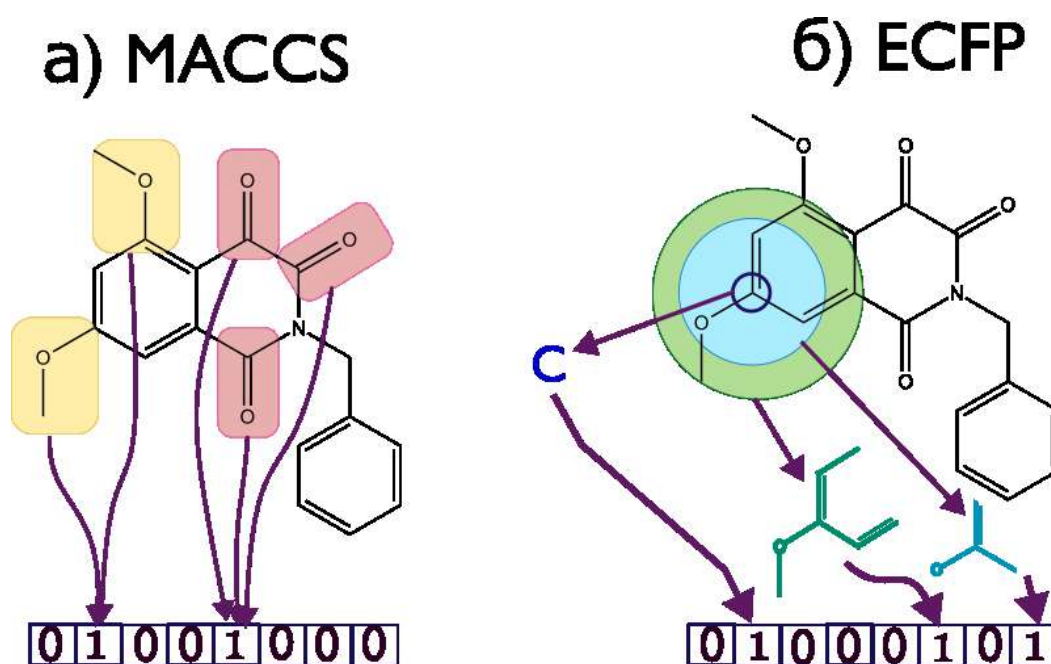


Рисунок 4. Представление молекулярных структур с помощью (а) MACCS и (б) ECFP

1.2.2. Геометрические (3D) дескрипторы

Геометрические (3D) дескрипторы – дескрипторы, включающие в себя трехмерную информацию о молекуле, такую как форма, объем и площадь поверхности. Геометрические дескрипторы важны для понимания свойств, зависящих от пространственного расположения атомов, и часто применяются для объектов исследования, представляющих собой большие молекулы, такие как протеины. Однако модели, обученные на наборах данных, составленных

исключительно из 3D дескрипторов, часто уступают в прогностической способности моделям, составленным исключительно из структурных дескрипторов, поэтому зачастую геометрические дескрипторы используют как дополнение к структурным дескрипторам [47]. Геометрические дескрипторы можно разделить на несколько групп, каждая из которых отражает различные аспекты молекулярной геометрии и ее влияние на поведение молекул.

1. Дескрипторы молекулярной формы. Эти дескрипторы количественно характеризуют общую форму молекулы, которая может влиять на ее способность взаимодействовать с биологическими мишенями или встраиваться в рецепторные участки. Например, эксцентриситет, асферичность и индексы формы, которые описывают, насколько молекула вытянута, плоская или сферическая.
2. Дескрипторы объема и площади поверхности. В эту группу входят дескрипторы, измеряющие площадь поверхности и объем молекулы, такие как площадь поверхности, доступной для растворителя, и молекулярный объем. Эти дескрипторы важны для предсказания растворимости, проницаемости и других свойств, связанных с тем, как молекула взаимодействует с окружающей средой.
3. Дескрипторы стерических эффектов. Эти дескрипторы количественно определяют пространственное расположение атомов в молекуле и возникающие при этом стерические эффекты, которые могут влиять на реакционную способность молекул и их взаимодействие. Отдельно можно выделить параметр Стеримол [48], который учитывает размеры атомов в молекуле и позволяет оценить стерическую доступность.
4. Дескрипторы момента инерции. Эти дескрипторы связаны с распределением массы внутри молекулы и ее вращательной инерцией

вокруг различных осей. Они предоставляют информацию о балансе молекулы и распределении массы в трехмерном пространстве.

5. 3D индексы. 3D индекс Винера (W3D) – наиболее известный топологический дескриптор, используемый для количественной оценки разветвленности молекул и определяемый как сумма длин кратчайших путей между всеми парами вершин (атомов) в молекуле [49]. W3D учитывает геометрические расстояния между парами атомов, а не только их топологические расстояния. Эта модификация числа Винера позволяет W3D точнее характеризовать пространственную конфигурацию молекулы. 3D Индекс Балабана (J3D) — еще один топологический дескриптор, измеряющий степень разветвленности и цикличности внутри молекулы. Как и индекс Винера, J3D рассчитывается на основе геометрических, а не топологических расстояний между атомами [50].
6. Автокорреляционные 3D дескрипторы. В отличие от предыдущих дескрипторов, 3D индексов, дескрипторы 3D-автокорреляции количественно оценивают не только расстояние между парами атомов, но и вдобавок к ним распределение определенных молекулярных свойств (таких как заряд, масса или гидрофобность) по структуре молекулы [51].
7. Взвешенные холистические инвариантные молекулярные дескрипторы (Weighted Holistic Invariant Molecular, WHIM). Дескрипторы WHIM отражают информацию о форме, размере, симметрии и распределении атомов в молекуле и рассчитываются на основе (x, y, z)-координат атомов. Дескрипторы WHIM разработаны таким образом, чтобы быть постоянными, вне зависимости от положения или поворота молекулы [52].

1.2.3. Квантово-химические дескрипторы

Квантово-химические дескрипторы играют важную роль в вычислительной химии, позволяя понять свойства и поведение молекул с помощью основанных на квантовой механике расчётов. Для получения этих дескрипторов используются различные вычислительные методы, включая *ab initio*, теорию функционала плотности (density functional theory, DFT) и полуэмпирические подходы.

Методы *Ab initio* направлены на решение уравнения Шредингера для молекулярных систем. Расчёты этими методами выполняются без обращения к эмпирическим параметрам или подгонки к экспериментальным данным, и единственными исходными данными для расчетов *ab initio* являются физические константы. Это делает эти методы высокоточными, однако крайне долгими и требующими значительных вычислительных мощностей.

Теория функционала плотности предлагает другой подход, основанный на электронной плотности системы, а не на её волновой функции. DFT предполагает, что все свойства системы могут быть определены из её электронной плотности, вычисления этим методом являются менее ресурсозатратными.

Полуэмпирические методы упрощают вычислительный процесс за счет включения эмпирических данных и упрощений в квантово-механические расчёты. Такой подход значительно сокращает необходимые вычислительные ресурсы. Полуэмпирические методы, такие как Parametric Method 3 (PM3), обеспечивают баланс между эффективностью вычислений и точностью, необходимой для эффективного изучения свойств молекул. Квантово-химические дескрипторы включают в себя такие параметры как:

- полная энергия молекулы, полученная в результате квантово-механических расчетов, включает в себя все кинетические и потенциальные энергии электронов и ядер.

- теплота образования, характеризующая изменение энергии, связанное с образованием молекулы из составляющих ее атомов в их стандартных состояниях.
- энергии отталкивания и притяжения, количественно определяющие электростатическое отталкивание между электронами и притяжение между электронами и ядрами, соответственно.
- энергия протонирования, характеризующая изменение энергии, связанное с присоединением протона к молекуле.
- электронная плотность, отображающая распределение электронной плотности в молекуле, выделяя области с большим и меньшим количеством электронов.

1.2.4. Электростатические дескрипторы

Электростатические дескрипторы количественно определяют распределение и влияние электрических зарядов внутри молекулы. Характеристики, описанные этими дескрипторами, влияют на такие свойства молекулы, как растворимость, сродство связывания и реакционная способность. Для расчёта электростатических дескрипторов молекул используют те же программы, что и для расчёта квантовых дескрипторов: Gaussian, HyperChem, Open3DGRID.

Среди электростатических дескрипторов можно выделить следующие: эмпирические атомные частичные заряды Гастайгера-Марсили -- полуэмпирические заряды, рассчитанные с помощью метода выравнивания электроотрицательности и дающие приближенное представление о распределении заряда между атомами молекулы. Эмпирические атомные парциальные заряды Зефирова – заряды, полученные на основе эмпирического метода, учитывающего электронную структуру молекулы.

Атомные частичные заряды Малликена – заряды, полученные в результате квантово-механических расчетов, отражают распределение электронной плотности вокруг атомов [53].

Минимальный и максимальный атомные частичные заряды представляют собой наиболее отрицательный и наиболее положительный заряды в молекуле, указывая на области высокого и низкого сродства к электрону соответственно.

Параметры полярности – дескрипторы, количественно характеризующие общую полярность молекулы и отражающие её дипольный характер, и потенциал для полярных взаимодействий. Дипольный момент -- векторная величина, измеряющая разделение положительных и отрицательных зарядов в молекуле и указывающая на ее общую полярность. Молекулярная поляризуемость – дескриптор, описывающий, насколько легко электронное облако молекулы может быть искажено внешним электрическим полем, влияющим на ее взаимодействие со светом и другими молекулами. Молекулярная гиперполяризуемость – поляризуемость высшего порядка, которая вносит вклад в нелинейно-оптические свойства. Локальная полярность молекулы характеризует изменение полярности в различных областях молекулы.

Средняя энергия ионизации отражает энергию, необходимую для удаления электрона из молекулы, связанную с её ионизационным потенциалом и стабильностью. Минимальный и максимальный электростатический потенциал на поверхности молекулы указывают на области с самым низким и самым высоким электростатическим потенциалом, что важно для понимания того, как молекулы взаимодействуют с заряженными частицами. Общая вариация поверхностного электростатического потенциала – количественная оценка изменчивости электростатического потенциала по всей поверхности молекулы. Параметр электростатического баланса описывает баланс между положительными и отрицательными участками внутри молекулы, который влияет на ее общее электростатическое поведение.

Кроме того, отдельно можно выделить дескрипторы заряженной частичной площади поверхности (Charged partial surface area, CPSA). Дескрипторы CPSA созданы для описания полярных межмолекулярных взаимодействий. Дескрипторы CPSA оказались применимыми, например, при изучении токсичности водных растворов, выступая альтернативой мерам уровня энергии LUMO для описания глобальной и локальной электрофильности в случае нековалентных молекулярных взаимодействий [54].

1.3. Математические методы для обработки данных

Для установления количественных связей между молекулярной структурой соединений и их свойствами в подходе QSPR используются различные математические методы, позволяющие выявить сложные взаимосвязи. Основная цель этих методов, в частности, основанных на линейной регрессии – моделирование количественных свойств таких как селективность, плотность или биологическая активность.

Среди методов регрессии, используемых в QSPR, преобладают множественная линейная регрессия (MLR) и частичные наименьшие квадраты (PLS). Каждый метод направлен на получение математического уравнения регрессии, связывающего молекулярные дескрипторы со свойствами. В общем случае линейной регрессии это уравнение выглядит следующим образом

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n + \varepsilon, \quad (1)$$

Где y - зависимая переменная, представляющая моделируемое свойство,
 $a_0, a_1 \dots a_n$ - коэффициенты регрессии, определяющие изменение Y в результате изменения каждой переменной X на одну единицу при условии, что все остальные переменные X остаются неизменными,

x_1, x_2, \dots, x_n - независимые переменные, которые при применении подхода QSPR представлены молекулярными дескрипторами,

ε представляет собой ошибку, учитывающую отклонение наблюдаемых значений от зависимости, определённой моделью.

Таким образом, основная цель регрессионных методов - найти оптимальные коэффициенты a , которые наилучшим образом описывают связь между дескрипторами и свойствами. Для этого коэффициенты необходимо рассчитать таким образом, чтобы минимизировать ошибку прогнозирования. Итоговую модель характеризуют различными метриками, чтобы оценить её качество.

Однако задача построения модели не ограничивается расчётом коэффициентов. Учитывая, что наборы данных в химических исследованиях зачастую невелики - как правило, они состоят из десятков точек данных, - риск переобучения модели очень велик. Переобучение происходит, когда модель описывает случайный шум, а не составляющий регрессию тренд. Для уменьшения этого вклада в модель используются такой метод проверки, как перекрёстная проверка. Перекрёстная проверка предполагает разделение данных на несколько подмножеств. При этом одни из них используются для обучения модели, а другие - для её тестирования. Такой подход позволяет убедиться в том, что модель хорошо обобщается на новые, неизвестные данные, а не просто повторяет зависимости обучающего набора.

1.3.1. Метрики и валидация

Оценка качества моделей важна для обеспечения их прогностической способности. Обычно такая оценка проводится с помощью различных статистических метрик, которые количественно определяют, насколько хорошо модели прогнозируют свойства. Прежде чем перейти непосредственно к

рассмотрению этих метрик, стоит обратить внимание на понятие остатков, как основополагающей идеи для понимания этих показателей.

Остатки - это разница между наблюдаемыми значениями целевой переменной и прогнозами, сделанными моделью.

Остаток = фактическое значение – спрогнозированное значение.

На Рисунке 5 проиллюстрирована линейная регрессия, построенная на одной переменной (x) и прогнозирующая зависимую переменную (y). Красные точки представляют фактические наблюдения, фиолетовая прямая соответствует линейной регрессионной модели, зелёные линии выделяют остатки для каждого наблюдения, визуально отображая ошибки в прогнозе.

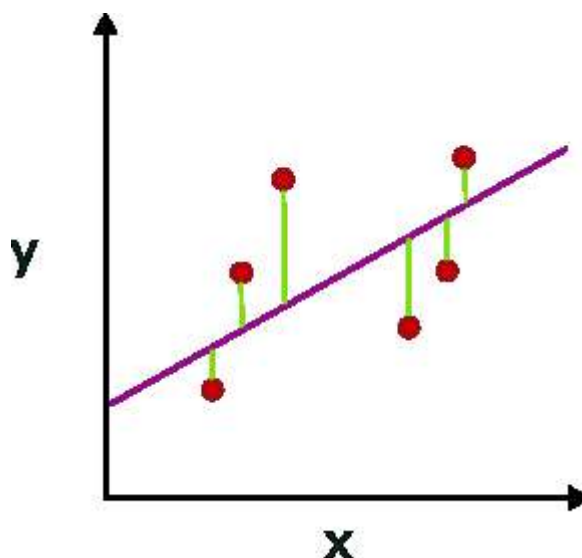


Рисунок 5. Пример остатков для линейной модели с одним признаком

Это визуальное представление – не просто метод, позволяющий увидеть, насколько модель справляется с поставленной задачей прогнозирования, но также способ, помогающий выявить закономерности ошибок, которые могут указывать на определённые проблемы модели.

Основываясь на концепции остатков, средняя квадратичная ошибка (Mean Squared Error, MSE) определяет среднее значение этих квадратичных разностей, обеспечивая комплексную меру ошибки прогнозирования модели:

$$MSE = \frac{1}{n} \sum_{i=1}^n \text{остаток}^2, \quad (2)$$

Корень средней квадратичной ошибки (Root Mean Squared Error, RMSE) обеспечивает меру ошибки в тех же единицах, что и прогнозируемое свойство: $RMSE = \sqrt{MSE}$.

Коэффициент детерминации, или R^2 -- ещё одна важная метрика, полученная из остатков. Он измеряет долю дисперсии зависимой переменной и даёт представление о силе связи между входными и выходными данными модели:

$$R^2 = 1 - \frac{\sum \text{остаток}^2}{\sum (y_i - \bar{y})^2}, \quad (3)$$

Из уравнения следует, что максимальное значение $R^2 = 1$, тогда чем больше значение R^2 , тем лучше модель, тем точнее модель воспроизводит наблюдаемые значения зависимой переменной.

При расчёте модели, важно оценивать качество прогнозирования моделью неизвестных ей данных. Самый простой метод такой оценки - разделение исходного набора данных на две части: обучающую и тестовую выборки. Модель обучается на обучающем наборе, а затем оценивается с помощью метрик, описанных выше, на тестовом наборе, который выступает в качестве новых, неизвестных модели данных. Этот метод максимально прост, но исключает тестовый набор из данных для обучения, что может быть существенным ограничением при работе с небольшими наборами данных.

Чтобы максимально эффективно использовать имеющиеся данные, особенно если они ограничены, метод разбиения данных получил развитие в

перекрёстную проверку (кросс-валидацию). Перекрёстная проверка предполагает многократное разбиение набора данных на обучающий и тестовый наборы, что позволяет использовать каждое подмножество данных как для обучения, так и для тестирования. Наиболее распространенной формой этой техники является K-кратная кросс-валидация.

При K-кратной кросс-валидации набор данных делится на K равных частей. Модель обучается на K-1 частях, а оставшаяся часть используется в качестве тестового набора. Этот процесс повторяется K раз, причем каждая из K частей используется в качестве тестового набора ровно один раз. Окончательная производительность модели обычно представляет собой среднее арифметическое оценок производительности модели на этих K тестовых наборах. Этот метод не только повышает устойчивость оценки модели, но и более эффективно использует данные, обеспечивая вклад каждой точки данных как в обучение, так и в проверку.

Экстремальным вариантом K-кратной кросс-валидации является полная кросс-валидация (full cross-validation, leave-one-out cross-validation, LOO CV), где K равно количеству образцов в наборе данных. Это означает, что каждая часть содержит ровно одну точку данных. Модель обучается на всех данных, кроме одного образца, и тестируется на этом исключённом образце. Метод LOO CV особенно полезен для очень маленьких наборов данных, но может потребовать больших вычислительных затрат при работе с большими наборами данных, что делает его менее практичным в таких сценариях.

1.3.2. Множественная линейная регрессия

Множественная линейная регрессия (Multiple Linear Regression, MLR) является одним из основополагающих и до сих пор наиболее часто применяемых

статистических методов в QSPR. Его популярность обусловлена простотой, интерпретируемостью и понятным способом моделирования взаимосвязи между зависимой переменной (интересующим свойством) и несколькими независимыми переменными (молекулярными дескрипторами). Будучи одним из самых первых методов, использованных в исследованиях QSPR, MLR является предшественником более сложных вычислительных методов в этой области. основополагающая идея этого метода заключается в предположении, что между интересующим свойством и молекулярными дескрипторами существует линейная зависимость.

С формальной точки зрения модель MLR описывается общим уравнением линейной регрессии (1), приведённом ранее. И цель MLR – найти набор коэффициентов (a_0, a_1, \dots, a_n), которые наилучшим образом соответствуют наблюдаемым данным – соответствует общей цели линейной регрессии. Как правило, нахождение коэффициентов реализуется с помощью метода наименьших квадратов.

Из-за своей простоты метод MLR эффективно работает только на очень простых объектах исследований из-за следующих существенных ограничений: количество дескрипторов не должно превышать количество образцов и между дескрипторами не должно быть линейной зависимости. Кроме того, нужно учитывать, что метод предполагает существование линейной связи между каждым дескриптором и зависимой переменной.

1.3.3. Проекция на латентные структуры

Регрессия с помощью метода частичных наименьших квадратов (Partial Least Squares, PLS), также известного как проекция на латентные структуры (projection on latent structures), является ключевым статистическим методом в области QSPR, особенно ценным за его эффективность при работе со сложными, многомерными данными. Метод PLS разработан Сванте Вольдом специально для применения в хемометрике [55] и незаменим в исследованиях QSPR, где

имеющиеся экспериментальные данные могут быть скудными. В отличие от ранее описанного метода MLR, метод PLS позволяет моделировать зависимость с помощью латентных переменных (latent variable, LV) в наборах переменных, где количество дескрипторов (независимых переменных) значительно превышает количество образцов и где переменные могут быть линейно зависимы. Эти особенности метода незаменимы для нахождения зависимости в случаях, когда дескрипторы представляют собой частоты спектра [56].

Регрессия PLS направлена на поиск зависимости между двумя матрицами, X (независимые переменные, представленные в QSPR молекулярными дескрипторами) и Y (отклики, например, физико-химические свойства), путём проецирования независимых переменных на новые латентные переменные. Исходные данные представляют собой пространство с размерностью, соответствующей количеству дескрипторов, однако для нахождения линейной зависимости мы уменьшаем исходную размерность проекцией имеющихся данных на новые оси – латентные переменные.

Процесс нахождения этих переменных следующий: первая стадия – нормализация и центрирование данных дескрипторов. Это гарантирует, что каждый дескриптор преобразован с учётом значений таким образом, чтобы переменные с большими числовыми диапазонами не имели большего веса в анализе. Тогда анализ будет сосредоточен на распределении и ковариации дескрипторов, а не на их абсолютных значениях.

После этого, направления максимальной ковариации определяются с помощью итерационного процесса: алгоритм начинается с нахождения в пространстве дескрипторов направления, которое имеет наибольшую ковариацию с зависимой переменной. Далее строится латентная переменная как линейная комбинация дескрипторов, взвешенная по их вкладу в это направление максимальной ковариации.

Процесс продолжается последовательным нахождением латентных переменных, каждая из которых представляет собой направление в пространстве дескрипторов, отражающее следующий наивысший уровень ковариации с зависимой переменной, ортогональное к уже рассмотренным направлениям. Эта ортогональность гарантирует, что каждая новая латентная переменная отражает уникальную информацию, которая еще не учтена предыдущими латентными переменными. На Рисунке 6 представлен результат нахождения латентных переменных для набора образцов.

Окончательная модель PLS строится с использованием этих латентных переменных в качестве исходных данных. Количество латентных переменных, которые необходимо включить в модель, обычно определяется с помощью валидации, чтобы сбалансировать передачу информации модели и риск её переобучения.

Таким образом, кроме матрицы независимых переменных X и матрицы свойств Y , для нахождения зависимости используются также матрицы счетов и нагрузок (Рисунок 7): матрица счетов (T) представляет собой проекцию исходных данных на латентные переменные, а матрица нагрузок (P): нагрузки показывают, какой вклад вносит каждый дескриптор в каждую латентную переменную.

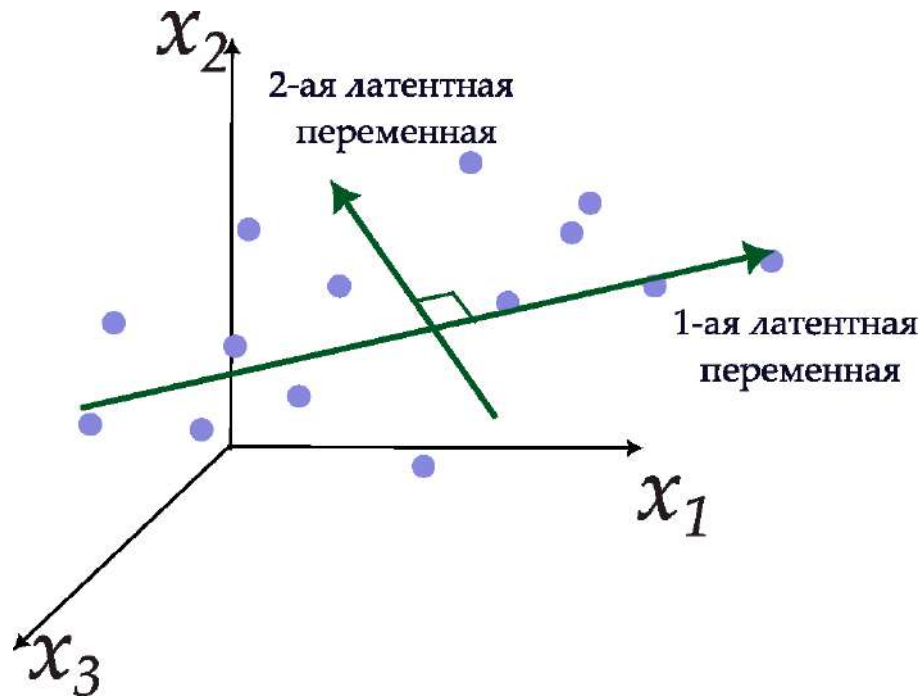


Рисунок 6. Иллюстрация итеративного процесса построения латентных переменных в PLS-регрессии. Исходные оси представляют собой нормализованные и центрированные дескрипторы. Облако точек отображает распределение объектов в пространстве дескрипторов. Латентные переменные найдены как направления максимальной ковариации с зависимой переменной, ортогональные друг другу

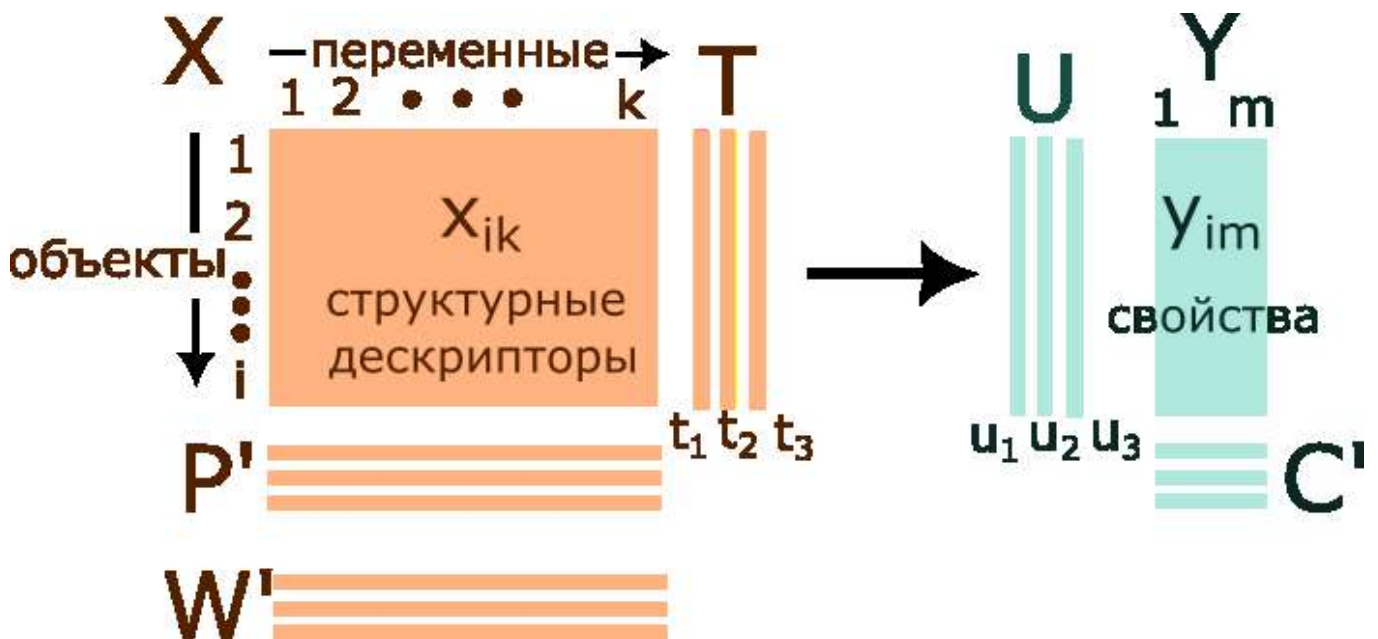


Рисунок 7. Данные PLSR могут быть представлены в виде двух таблиц, матриц, X и Y

Основная идея PLS может быть описана следующими уравнениями:

$$X = TP^T + E, \quad (4)$$

$$Y = UQ^T + F, \quad (5)$$

где T и U - матрицы счетов для X и Y, соответственно, представляющие проекции X и Y на латентные переменные.

P и Q - матрицы нагрузок, показывающие, какой вклад вносит каждая исходная переменная в X и Y в компоненты.

E и F - матрицы остатков, содержащие информацию, не объясненную моделью.

Таким образом, уравнение для нахождения модели PLS:

$$Y = XB + F, \quad (6)$$

где B - матрица коэффициентов регрессии, рассчитанная для максимизации ковариации между X и Y.

Поскольку модель позволяет не только спрогнозировать свойства, но и понять взаимосвязь между структурой и свойствами с помощью интерпретации счетов и нагрузок, то PLS-регрессия является широко применимым методом в QSPR, позволяющим исследователям раскрывать взаимосвязи между молекулярными дескрипторами и искомыми свойствами даже в условиях сложных сценариев получения данных [57,58].

1.3.4. Метод опорных векторов

Регрессия методом опорных векторов (support vector regression, SVR) представляет собой подход, известный своей эффективностью при работе со сложными и нелинейными взаимосвязями. Возникнув на основе принципов метода опорных векторов (support vector machine, SVM), используемого для классификации, SVR расширяет применение метода для решения задач регрессии.

В отличие от традиционных методов регрессии, которые минимизируют ошибку между прогнозируемыми и фактическими значениями, SVR фокусируется на подгонке ошибки в пределах заданного порога, что делает её полезной при работе с наборами данных с шумом [59].

Основная идея SVR заключается в поиске гиперплоскости в многомерном пространстве дескрипторов, которая наилучшим образом соответствует данным. Понятие гиперплоскости является центральным для понимания метода SVR и её предшественника – метода SVM. Гиперплоскость – (n-1)-мерная подплоскость в n-мерном пространстве (линия в двумерном пространстве, плоскость в трёхмерном), которая используется для разделения точек данных в пространстве с целью нахождения классификации или регрессии.

Цель SVR – построение модели, которая предсказывает выходное значение y для входного вектора X с отклонением не более ϵ для каждой точки данных.

Уравнение гиперплоскости в SVR выглядит следующим образом:

$$f(x) = w^T X + b, \quad (7)$$

где w – вектор весов дескрипторов, X – матрица дескрипторов, а b – смещение гиперплоскости относительно начала координат.

1.3.5. Нейронные сети

Нейронные сети (Neural Networks) или искусственные нейронные сети (Artificial Neural Net) – самый сложный из рассматриваемых класс моделей машинного обучения. Главная привлекательность нейронных сетей заключается в способности моделировать сложные, нелинейные взаимосвязи, которые зачастую слишком сложны для традиционных статистических методов. Построение таких взаимосвязей достигается имитацией структуры и функций человеческого мозга, используя слои взаимосвязанных узлов или «нейронов» для обработки данных с помощью ряда преобразований и связей.

Нейроны принимают сигналы от одного или нескольких входов и

преобразуют их в сигнал на выходе. Каждый нейрон можно описать как совокупность двух составляющих (Рисунок 8) – сумматор (Σ) и функция активации (f). Сумматор получает входные сигналы и возвращает их взвешенную сумму, которая затем преобразуется нелинейной функцией активации в сигнал выхода. В качестве функции активации могут использоваться различные функции, такие как сигмоида, гиперболический тангенс (\tanh), ReLU (Rectified Linear Unit) и другие. Выбор конкретной функции зависит от задачи и архитектуры нейронной сети. С математической точки зрения эту схему можно описать следующим образом:

$$y = f(w \cdot x + b), \quad (8)$$

Где y - выходной сигнал, x - входной вектор, w - весовой вектор, b - смещение, f - функция активации

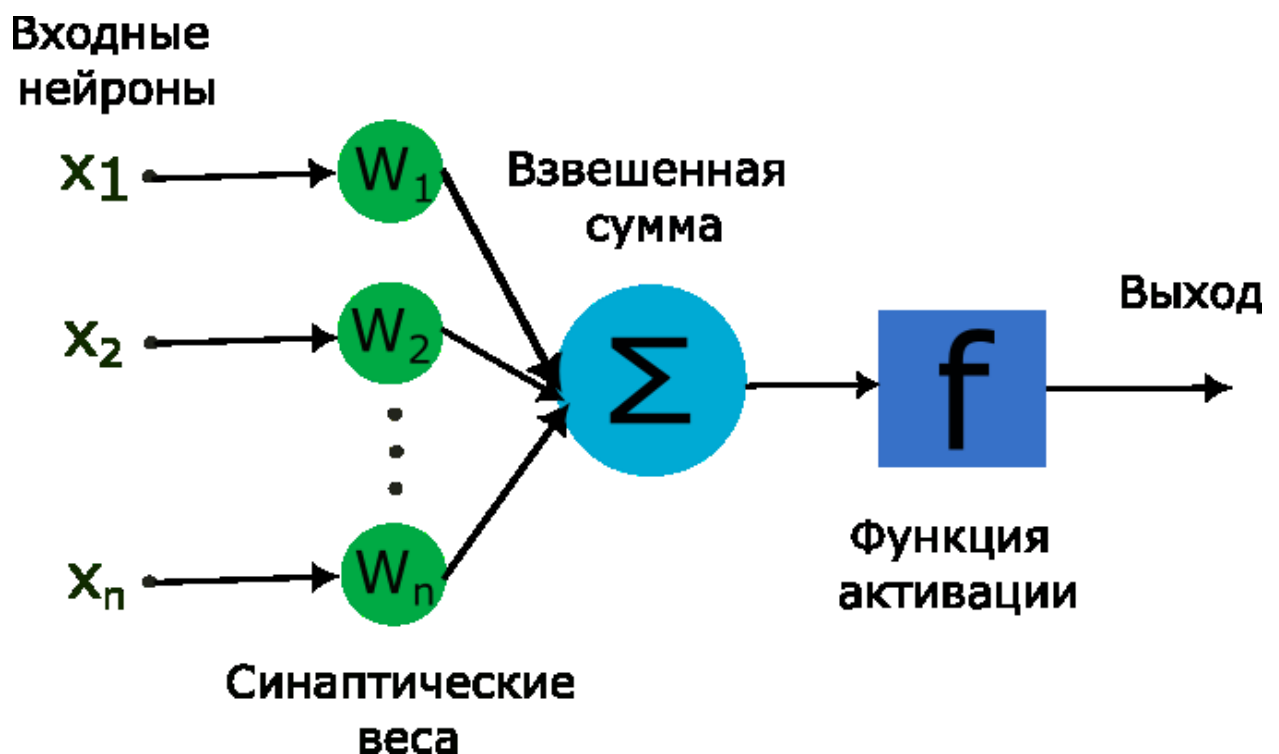


Рисунок 8. Схема нейрона искусственной нейронной сети

Фундаментальная идея NN предполагает расположение нейронов в слоях: входной слой, один или несколько скрытых слоев и выходной слой (Рисунок 9). Каждый нейрон в одном слое соединяется с нейронами в следующем слое. Тогда

процесс обучения нейронной сети, в ходе которого минимизируется разница между прогнозируемым и фактическим результатом, представляет собой изменение смещения и взвешенных сумм в сумматорах нейронов.

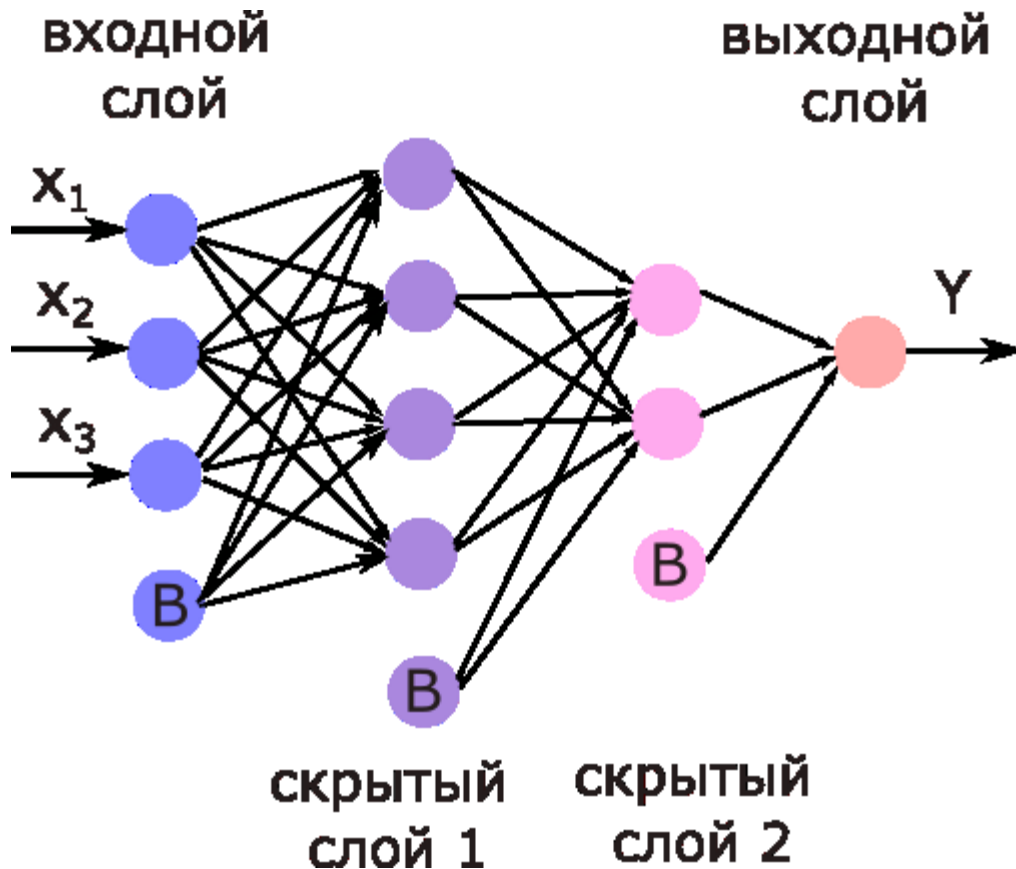


Рисунок 9. Многослойная нейронная сеть с тремя входами ($X_i = 1, 2, 3$), двумя скрытыми слоями и одним выходом (Y). B – нейроны смещения

Нейронные сети, благодаря своей сложной структуре, позволяют установить зависимости в наборах данных, где дескрипторы нелинейно связаны со свойством. Однако NN имеют и недостатки – они легко переобучаемы, особенно при большом количестве дескрипторов, и сложно интерпретируемы, что затрудняет понимание влияния дескрипторов на свойства.

1.4. Применение QSPR в аналитической химии

Несмотря на значительный прогресс в области QSPR, разработка точных и надежных моделей для прогнозирования свойств по-прежнему сталкивается с трудностями, связанными с ограниченным объемом и разнообразием экспериментальных данных. Вариативность экспериментальных условий и сложность химических систем, особенно в случае многокомпонентных смесей, таких как глубокие эвтектические растворители, затрудняют сбор однородных наборов данных, необходимых для построения обобщающих моделей.

Тем не менее, QSPR-моделирование успешно применяется в ряде областей аналитической химии, где удалось собрать достаточное количество данных из экспериментов, проведенных в схожих условиях. К таким областям относятся разработка ионоселективных сенсоров, оптимизация экстракции целевых компонентов из натуральных продуктов с помощью ГЭР и разработка новых хроматографических фаз.

Одним из важнейших применений QSPR является разработка ионоселективных сенсоров. В исследовании Соловьева и др. (2019) [60], применили метод QSPR, используя структурные фрагменты химических соединений в качестве дескрипторов, для прогнозирования потенциометрической чувствительности полимерных мембранных сенсоров к ионам тяжелых металлов, таких как Cu^{2+} , Zn^{2+} , Cd^{2+} и Pb^{2+} . Эта работа демонстрирует, что, несмотря на ограниченный набор данных, модели QSPR могут достичь значительной прогностической силы, о чём свидетельствуют низкое значение RMSE (около 5 мВ/дек) и высокий коэффициент детерминации (R^2 около 0,8) при перекрёстной проверке. Это говорит о том, что QSPR может эффективно дополнять традиционные экспериментальные подходы, прогнозируя, как различные лиганды в ПВХ-пластифицированных мембранах будут реагировать на конкретные ионы.

В дальнейшем Мартышко и др. (2020) изучили селективность Mg^{2+}/Ca^{2+} сенсорных мембран с использованием различных амидных лигандов [61]. Это исследование демонстрирует, как модели QSPR могут облегчить ранние стадии разработки сенсоров, выявляя перспективных кандидатов до проведения синтеза и измерений.

QSPR-моделирование позволяет не только прогнозировать чувствительность сенсоров, но и разрабатывать более селективные и эффективные сенсорные системы. Например, в исследовании Паурбашира (2023) используется комбинация генетических алгоритмов и машин опорных векторов для прогнозирования селективности к карбонатам для сенсоров с ПВХ-пластифицированной мембраной [62].

Аналогичным образом Киани-анбуи с коллегами применили QSPR для моделирования коэффициентов селективности электродов к лантану [63]. Их подход предполагает использование теории функционала плотности для расчёта дескрипторов, что позволяет более детально и точно спрогнозировать характеристики сенсора.

Новое применение QSPR проиллюстрировано в работе Фан Ти Дием-Тран и др. (2023), где модели QSPR были разработаны для прогнозирования стабильности и потенциометрической чувствительности комплексов, образованных ионами тяжелых металлов с органическими флуоресцентными соединениями [64]. В данном исследовании особое внимание уделялось кумариноподобным лигандам, которые известны своими флуоресцентными свойствами. Модели QSPR показали отличную производительность, с квадратичными коэффициентами корреляции (R^2) до 1,00, демонстрируя точность метода в прогнозировании взаимодействий лигандов с такими металлами, как Cu^{2+} , Cd^{2+} и Pb^{2+} . Проанализировав виртуальную базу данных кумариноподобных структур, исследователи выявили несколько новых лигандов с многообещающими чувствительностью и константами стабильности. Последующие квантово-химические расчеты и токсикологическое

профилирование помогли оценить их пригодность в качестве флуоресцентных хемосенсоров для тяжелых металлов в промышленных сточных водах.

Глубокие эвтектические растворители – новый тип растворителей, представленный Эбботтом и коллегами около двух десятилетий назад [65]. ГЭР обычно состоят из донора водородных связей (hydrogen bond donor, HBD) и акцептора водородных связей (hydrogen bond acceptor, HBA), которые вместе образуют эвтектическую смесь с температурой плавления значительно ниже, чем у каждого из компонентов в отдельности. Эта особенность позволяет ГЭР оставаться жидкими при комнатной температуре, что делает их идеальными для различных химических применений благодаря повышенной растворяющей способности, пониженной летучести и регулируемой вязкости.

Поскольку ГЭР – многокомпонентные системы, как правило, при применении к ним подхода QSPR, моделируют какой-либо один из компонентов.

Лемауи и др. впервые применили QSPR для прогнозирования свойств глубоких эвтектических растворителей, используя модели QSPR и сигма-профили COSMO-RS в качестве молекулярных дескрипторов [66,67]. Прогнозируемые значения электропроводности ГЭР, полученные с помощью разработанных моделей, хорошо согласовывались с экспериментальными данными.

Салахшури и др. расширили применение QSPR ГЭР, разработав гибридную модель, сочетающую QSPR и регрессию гауссова процесса, для прогнозирования растворимости CO_2 в различных составах ГЭР [68]. Это исследование продемонстрировало потенциал моделей QSPR как инструмента для точного прогнозирования взаимодействий между компонентами ГЭР и растворенными газами, что может быть использовано для создания и оптимизации систем растворителей для улавливания углерода.

Кроме того, модели QSPR нашли применение в оптимизации эффективности экстракции целевых компонентов из натуральных продуктов с помощью ГЭР. Например, Ху и др. использовали модели QSPR в сочетании с моделями COSMO-

SAC для исследования влияния состава ГЭР на эффективность экстракции олеокантала из оливкового масла [69]. Благодаря построенным моделям, удалось точно спрогнозировать результаты экстракции и выбрать наиболее эффективные составы ГЭР для данного процесса.

Разработка хроматографических фаз – необходимый процесс для развития хроматографии, чрезвычайно широко используемого аналитического метода. Подход QSPR можно применить для разработки новых фаз, предназначенных для конкретных аналитических целей, тем самым повышая эффективность и специфичность хроматографического разделения сложных смесей. Первые исследования в области прогнозирования времени удерживания были проведены Федоровой и др. в работе [70], где изучили прогнозирование времени удерживания с помощью моделей глубокого обучения в высокоэффективной жидкостной хроматографии (ВЭЖХ). На молекулярных отпечатках и представлениях SMILES были обучены различные архитектуры моделей глубокого обучения. Наилучшие показатели продемонстрировала одномерная свёрточная нейронная сеть (1D CNN), для которой использовали SMILES в качестве входных данных.

Опираясь на исследование Фёдоровой, Обрадович и др. ещё более усовершенствовали прогностические возможности моделей QSPR в своем сравнительном исследовании инструментов расчета дескрипторов для моделирования механизма удерживания изомерных соединений [71]. Это исследование расширило предыдущую работу, применив более широкий набор вычислительных инструментов и методов машинного обучения, включая SVM, MLR и XGBoost, а также такие методы отбора признаков, как генетический алгоритм и шаговая регрессия.

Несмотря на достигнутые успехи, QSPR-моделирование в аналитической химии сталкивается с рядом нерешённых задач, требующих дальнейших исследований.

Во-первых, возможности QSPR для прогнозирования свойств глубоких эвтектических растворителей остаются недостаточно изученными. Существующие подходы, основанные на моделировании отдельных компонентов ГЭР, не всегда позволяют учесть сложные взаимодействия между компонентами, что может ограничивать точность прогнозирования свойств смеси в целом. Разработка новых моделей QSPR, способных прогнозировать свойства ГЭР на основе их состава, является актуальной задачей, решение которой позволит расширить область применения этих растворителей.

Во-вторых, существует необходимость в расширении применения QSPR-моделирования для прогнозирования различных характеристик сенсоров, включая селективность к гидрофильным анионам, таким как карбонат-ионы, и чувствительность к катионам тяжёлых металлов. Ограниченное число доступных ионофоров и сложность достижения высокой селективности и чувствительности представляют собой значительные проблемы в разработке эффективных сенсоров. Учитывая важность определения как анионов, так и катионов тяжёлых металлов в экологическом мониторинге, контроле качества, медицине и других областях, создание новых QSPR-моделей для прогнозирования и оптимизации характеристик сенсоров является крайне актуальной задачей.

1.5. Цели и задачи работы

Анализ применимости подхода QSPR к различным объектам аналитической химии, в сочетании с обзором существующих дескрипторов и методов моделирования, показал, что, несмотря на перспективность данного подхода, количество исследований в этой области ограничено. Это связано с трудностями получения больших однородных наборов данных, необходимых для построения классических QSPR-моделей.

Тем не менее, решение этой задачи представляется крайне важным для развития аналитической химии. Поэтому в данной работе были выбраны два направления исследований, обладающие высоким потенциалом практического применения и позволяющие собрать достаточное количество экспериментальных данных для построения надежных QSPR-моделей: ионоселективные электроды и глубокие эвтектические растворители.

В соответствии с этим, основная цель данного исследования заключается в расширении области применения QSPR в аналитической химии путем разработки моделей для прогнозирования свойств ионоселективных электродов и глубоких эвтектических растворителей. Для достижения этой цели были поставлены следующие три задачи:

1. Разработка и валидация QSPR-моделей для прогнозирования селективности ионоселективных сенсоров к карбонат-ионам (CO_3^{2-}).

Определение анионов, таких как карбонат-ион (CO_3^{2-}), с помощью ионоселективных электродов (ИСЭ) представляет собой сложную задачу в аналитической химии. Разработка селективных лигандов для связывания анионов в целом затруднена по сравнению с катионами, поскольку неорганические анионы характеризуются более низким отношением заряда к радиусу, большим разнообразием геометрий, а их формы существования в растворе сильно зависят от pH.

Кроме того, создание высокоселективных сенсоров для карбонат-ионов затруднено из-за их гидрофильности, что препятствует эффективному связыванию с ионофорами в гидрофобной мембране сенсора.

В настоящее время наиболее успешные коммерческие ионофоры для определения карбонат-ионов основаны на трифторацетофеноне и его производных. Ключевым структурным элементом этих соединений является карбонильная группа с пониженной электронной плотностью на кислороде, обусловленной сильным акцепторным влиянием группы $-\text{CF}_3$. Это позволяет

предположить, что другие соединения, содержащие акцепторные заместители вблизи карбонильной группы, также могут проявлять высокую селективность к карбонат-ионам.

Однако экспериментальная проверка каждого потенциального ионофора требует значительных временных и материальных затрат. Разработка QSPR-моделей, способных прогнозировать селективность ИСЭ к карбонат-ионам на основе структурных характеристик ионофоров, позволит значительно ускорить и оптимизировать процесс поиска новых эффективных лигандов. В качестве дескрипторов в таких моделях используются структурные фрагменты молекул ионофоров, а также характеристики пластификатора мембраны, кодируемые с помощью специализированного программного обеспечения.

2. Разработка и проверка QSPR-моделей для прогнозирования чувствительности ионоселективных сенсоров к катионам тяжелых металлов (Ca^{2+} , Cd^{2+} , Pb^{2+}).

Определение тяжёлых металлов (Ca^{2+} , Cd^{2+} , Pb^{2+}) в окружающей среде, пищевых продуктах и биологических образцах является важной задачей для экологического мониторинга, контроля качества и медицинской диагностики. Ионоселективные электроды представляют собой удобный и доступный инструмент для определения концентрации ионов металлов, однако их применение в реальных условиях часто ограничено недостаточной чувствительностью.

Одним из перспективных направлений улучшения характеристик ИСЭ является разработка новых ионофоров – веществ, обеспечивающих селективное связывание ионов металлов в сенсорной мембране. Дифенилфосфорилацетамиды представляют собой новый класс лигандов, обладающих высокой экстракционной способностью по отношению к ионам металлов, что делает их перспективными кандидатами для создания новых ионофоров [72].

Однако экспериментальная проверка каждого потенциального ионофора является дорогостоящим и трудоемким процессом. Разработка QSPR-моделей, позволяющих прогнозировать чувствительность ИСЭ на основе структурных характеристик ионофоров, позволит значительно ускорить и упростить процесс поиска новых эффективных лигандов. В качестве дескрипторов в таких моделях используются структурные фрагменты молекул ионофоров, кодируемые с помощью специализированного программного обеспечения.

3. Разработка и проверка QSPR-моделей для прогнозирования физико-химических свойств ГЭР с органическими кислотами в качестве доноров водородной связи.

Глубокие эвтектические растворители представляют собой новый класс растворителей, обладающих уникальными свойствами, такими как низкая летучесть, высокая растворяющая способность и регулируемая вязкость. Эти свойства делают ГЭР перспективными для применения в различных областях, включая экстракцию, электрохимию, катализ и разработку новых материалов. Однако поиск оптимальных составов ГЭР для конкретных применений затруднен из-за большого числа возможных комбинаций компонентов и сложности экспериментального определения их свойств.

Данная задача заключается в создании моделей QSPR, позволяющих прогнозировать физико-химические свойства ГЭР, такие как плотность, проводимость, вязкость и коэффициент преломления, на основе структурных характеристик органических кислот, выступающих в качестве доноров водородной связи в составе ГЭР.

В качестве дескрипторов используются не только структурные фрагменты молекул органических кислот, но и полуэмпирические дескрипторы, такие как энергии НОМО и LUMO, дипольный момент и теплоемкость, рассчитанные с помощью квантово-химического метода PM3. Это позволит учесть электронные свойства молекул и повысить точность прогнозирования.

Выбор ГЭР с органическими кислотами в качестве доноров водородной связи обусловлен их широким использованием и структурным разнообразием, что позволяет варьировать свойства ГЭР в широких пределах. Плотность, проводимость, вязкость и коэффициент преломления являются ключевыми физико-химическими свойствами ГЭР, которые определяют их применимость в различных областях, а именно: плотность и вязкость влияют на эффективность экстракции и разделения компонентов смеси, проводимость является критическим параметром для электрохимических процессов, таких как электролиз и электроосаждение, физико-химические свойства ГЭР могут влиять на активность и селективность катализаторов.

Разработка таких моделей позволит оптимизировать состав ГЭР для конкретных применений, и ускорить процесс разработки новых ГЭР с улучшенными характеристиками.

Глава 2. ПРИМЕНЕНИЕ QSPR-МОДЕЛИРОВАНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ СЕЛЕКТИВНОСТИ КАРБОНАТНЫХ ИОНОСЕЛЕКТИВНЫХ ЭЛЕКТРОДОВ

В данной главе исследуется возможность применения QSPR для прогнозирования потенциометрической селективности к гидрофильным анионам пластифицированных полимерных мембранных сенсоров на основе новых лигандов. Разработка таких сенсоров является важной задачей в области экологического мониторинга, контроля качества воды и пищевых продуктов, а также в медицинской диагностике.

Несмотря на значительный прогресс в разработке ионоселективных электродов, создание сенсоров с высокой селективностью к карбонат-ионам остается сложной задачей. Это связано с гидрофильностью карбонат-иона и ограниченным числом доступных ионофоров, способных эффективно связывать его в гидрофобной мембране сенсора.

QSPR-моделирование представляет собой перспективный подход к решению этой проблемы, позволяющий прогнозировать селективность сенсоров на основе структурных характеристик ионофоров. В данной главе рассмотрено применение QSPR-моделирования для прогнозирования селективности карбонатных ИСЭ на основе набора данных, включающего 40 ионофоров, описанных в литературе. В качестве дескрипторов использованы структурные фрагменты молекул ионофоров. Кроме того, для расширения набора данных и учёта влияния матрицы сенсора на его отклик, добавлены данные о селективности хлоридных ионофоров и характеристики пластификатора мембраны.

Анализ полученных QSPR-моделей позволяет оценить их прогностическую способность и выявить ключевые структурные особенности ионофоров, ответственные за селективность к карбонат-ионам, что может быть использовано для направленного поиска новых лигандов с улучшенными характеристиками и разработки высокоселективных карбонатных сенсоров.

2.1. Экспериментальная часть

2.1.1. Выбор дескрипторов

Следуя методологии, приведённой в разделе 1.1, первая стадия эксперимента – сбор данных для дальнейшей обработки и построения модели. Набор данных для моделирования QSPR включал 40 структур с известными селективностями $\log K^{\text{сел}}(\text{HCO}_3^-/\text{Cl}^-)$, описанными в литературе (структуры приведены в Таблице 1 в Приложении А) [73–79]. Поскольку количество работ, в которых сообщается об успешной разработке карбонатных потенциометрических сенсоров, невелико, то с целью расширения набора данных за счёт дополнительных образцов, были также добавлены несколько ионофоров, демонстрирующих низкую карбонатную селективность, поскольку данные вещества являлись хлоридными ионофорами. В этих случаях приведенные значения $\log K^{\text{сел}}(\text{Cl}^-/\text{HCO}_3^-)$ были преобразованы в $\log K^{\text{сел}}(\text{HCO}_3^-/\text{Cl}^-)$ в соответствии с уравнением Никольского-Эйзенмана:

$$E = E_I^0 + \frac{RT}{z_I F} \ln \left(a_I + \sum K_{IJ} a_J^{z_I/z_J} \right), \quad (9)$$

где a_I , a_J , z_I , z_J – активности и заряды основного и мешающего ионов, соответственно, а K_{IJ} – коэффициент селективности. Значение коэффициента селективности может быть найдено в соответствии с выражением:

$$K_{IJ} = a_I / a_J^{z_I/z_J}, \quad (10)$$

тогда можно вычислить логарифм коэффициента селективности K_{IJ} при условии, что заряды ионов I и J равны. Таким образом, в нашем случае $\log K^{\text{сел}}(\text{Cl}^-/\text{HCO}_3^-) = -\log K^{\text{сел}}(\text{HCO}_3^-/\text{Cl}^-)$

Единообразие экспериментальных условий очень важно для получения качественного набора данных для моделирования QSPR. При сборе данных из литературных источников мы ограничились теми работами, в которых в качестве полимера для матрицы мембраны сенсора использовался поливинилхлорид (ПВХ). Ограниченное количество исследований в области селективности карбонатных сенсоров затрудняет составление подходящего для моделирования набора данных, в котором использовались бы одни и те же полимеры мембраны и пластификаторы. Поэтому мы дополнительно рассмотрели те работы, в которых в качестве пластификаторов использовались *o*-нитрофенил-октиловый эфир (*o*-nitrophenyl octyl ether, NPOE), диоктил адипат (dioctyl adipate, DOA) или бис(2-этилгексил) себацинат (bis(2-ethylhexyl) sebacate, DOS). При этом для учёта различий в свойствах сенсоров, которые обусловлены разными пластификаторами в составах мембран, в набор данных в качестве дескриптора была добавлена диэлектрическая проницаемость этих пластификаторов. В качестве анионообменника во всех литературных источниках использовался тридодecilметиламмоний (tridodecylmethylammonium, TDMA).

Кроме того, мы учитывали условия проведения измерений селективности сенсоров и рассматривали только литературные данные, полученные в диапазоне pH 7,0-8,6. Это было сделано, чтобы убедиться, что HCO_3^- является доминирующей ионной формой в растворах.

Был также оценён метод измерения селективности – все значения селективности $\log K^{\text{сел}}(\text{HCO}_3^-/\text{Cl}^-)$ в отобранных литературных источниках были получены методом отдельных растворов, за исключением восьми веществ из [77], где использовался метод приведённых потенциалов. Однако было решено оставить эти восемь образцов в наборе данных, так как эти два метода обычно дают сопоставимые результаты. Селективность $\log K^{\text{сел}}(\text{HCO}_3^-/\text{Cl}^-)$ в литературном наборе данных варьировалась от -5,8 до 6,2. Среднее значение составило -1,4, а медиана – 2,6.

2.1.2. Ионоселективные сенсоры с новыми лигандами в составе

Для изучения возможности прогнозирования селективности карбонатных сенсоров с помощью полученной QSPR-модели, основанной на литературных данных, были разработаны и исследованы сенсоры на основе новых лигандов.

Структуры лигандов представлены в Таблице 1. Выбор этих соединений для синтеза был обусловлен тем, что большинство известных карбонатных ионофоров основаны на трифторацетофеноне и его производных. Карбонильная группа с пониженной электронной плотностью на кислороде, обусловленной сильным акцепторным влиянием группы $-CF_3$, является ключевым фрагментом, ответственным за связывание карбонат-иона. Вещество 1, коммерчески доступное, было приобретено у компании Merck (Дармштадт, Германия) и использовалось без дополнительной очистки. Лиганды 2-4 были синтезированы для данного исследования коллегами из Вектона. Все четыре вещества содержат различные акцепторные заместители вблизи карбонильной группы, что позволяет предположить их потенциальную способность к эффективному связыванию карбонатов.

Для приготовления ПВХ-пластифицированных сенсорных мембран были использованы четыре новых лиганда. Составы мембран приведены в Таблице 2. Каждая мембрана содержала 50 ммоль/кг ионофора и 10 ммоль/кг TDMA-NO₃ в качестве анионообменника. Соотношение между ПВХ и растворителем-пластификатором составляло 1:2, а общий вес мембраны - 300 мг. В качестве растворителя-пластификатора использовали о-нитрофенилоктиловый эфир (NPOE).

Таблица 1. Структуры новых лигандов, исследованных в эксперименте

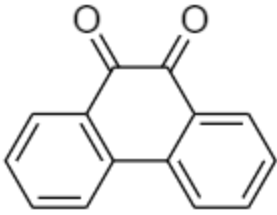
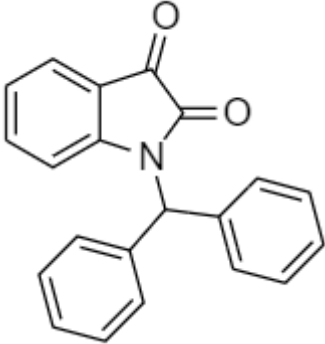
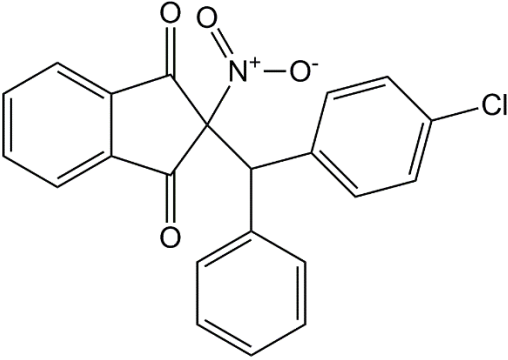
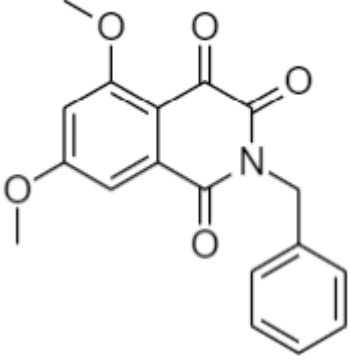
№	Структура лиганда	Название IUPAC
1		Фенантренхинон S1
2		1-(дифенилметил)-1H-индол-2,3-дион S2
3		2-((4-хлорфенил)(фенил)метил)-2-нитро-1H-индан-1,3-дион S3
4		5,7-диметокси-2-бензил-1,3,4(2H)-изохинолинтрион S4

Таблица 2. Составы сенсорных мембран, масс. %

Сенсор	PVC	TDMA NO ₃	Пластификатор	Лиганд
S1	32,79	0,60	65,57	1,04
S2	32,61	0,60	65,22	1,57
S3	32,48	0,60	64,96	1,96
S4	32,59	0,60	65,18	1,63

Полимерные мембраны для сенсоров были приготовлены в соответствии со стандартным протоколом. Навески всех компонентов мембраны растворяли в 3 мл свежеперегнанного тетрагидрофурана (ТГФ) в стеклянном стакане с помощью магнитной мешалки. Полученные растворы заливали в тефлоновые стаканы диаметром 20 мм и оставляли сушиться на 48 ч. Из исходных мембран вырезали диски диаметром 8 мм и прикрепляли их к корпусам ПВХ-датчиков с помощью смеси ПВХ и циклогексанона. Толщина подготовленных мембран составляла 0,4 мм. После высыхания клея внутренние части полученных электродов заполняли смесью 0,01 М NaHCO₃ и 0,001 М NaCl. Хлорид-анион необходим для функционирования Ag/AgCl электрода, а бикарбонат - для обеспечения его постоянного содержания в фазе мембраны, требуемого формализмом Никольского. Воду для приготовления всех водных растворов получали из дистиллятора GFL 2102 (GFL Burgwedel, Германия). Проводимость монодистиллята составляет 2,2 с/см при 25 °С. Наконец, в сенсоры были встроены внутренние Ag/AgCl электроды. Для каждого состава мембраны было изготовлено по три сенсора.

Потенциометрические измерения проводили с помощью многоканального цифрового мВ-метра ХАН-32 (ООО «Сенсорные системы», Санкт-Петербург, Россия), подключенного к ПК для сбора данных через USB-порт. Измерения проводили по отношению к стандартному эталонному Ag/AgCl электроду ЭСр-10101 («Измерительная техника», Москва, Россия). Для контроля значений pH в

растворах образцов использовали стеклянный pH-датчик PY-41 (GOnDO Electronic Co., Ltd., Тайбэй, Тайвань).

Чувствительность сенсоров изучали в водных растворах неорганических солей (сульфат натрия, хлорид натрия, нитрат натрия, бикарбонат натрия, дигидрофосфат натрия) в диапазоне концентраций от 10^{-6} до 10^{-2} М. Значения чувствительности рассчитывали как наклоны линейных участков кривых отклика сенсоров (10^{-4} — 10^{-2} М). Сенсоры промывали несколькими порциями дистиллированной воды до, после и между измерениями до достижения постоянных значений потенциала.

Коэффициенты селективности получали методом отдельных растворов, который также известен как метод бионных потенциалов [80].

E_i и E_j рассчитывались для растворов основного и мешающего ионов с концентрацией 10^{-3} М, соответственно.

2.1.3. QSPR моделирование

В качестве дескрипторов для описания молекулярных структур лигандов использовались структурные дескрипторы в виде молекулярных графов, представленные подструктурными молекулярными фрагментами (substructural molecular fragments, SMF).

Для получения SMF использовалась программа «MolFrag», входящая в состав программного комплекса «ISIDA QSPR» [81]. SMF в ISIDA может быть описана двумя способами: либо как последовательности атомов и/или связей (топологический путь), либо как выбранный («дополненный») атом (атомоцентрированные фрагменты) с его окружением, которое может быть атомами, связями или обоими. Оба способа проиллюстрированы на Рисунке 10.

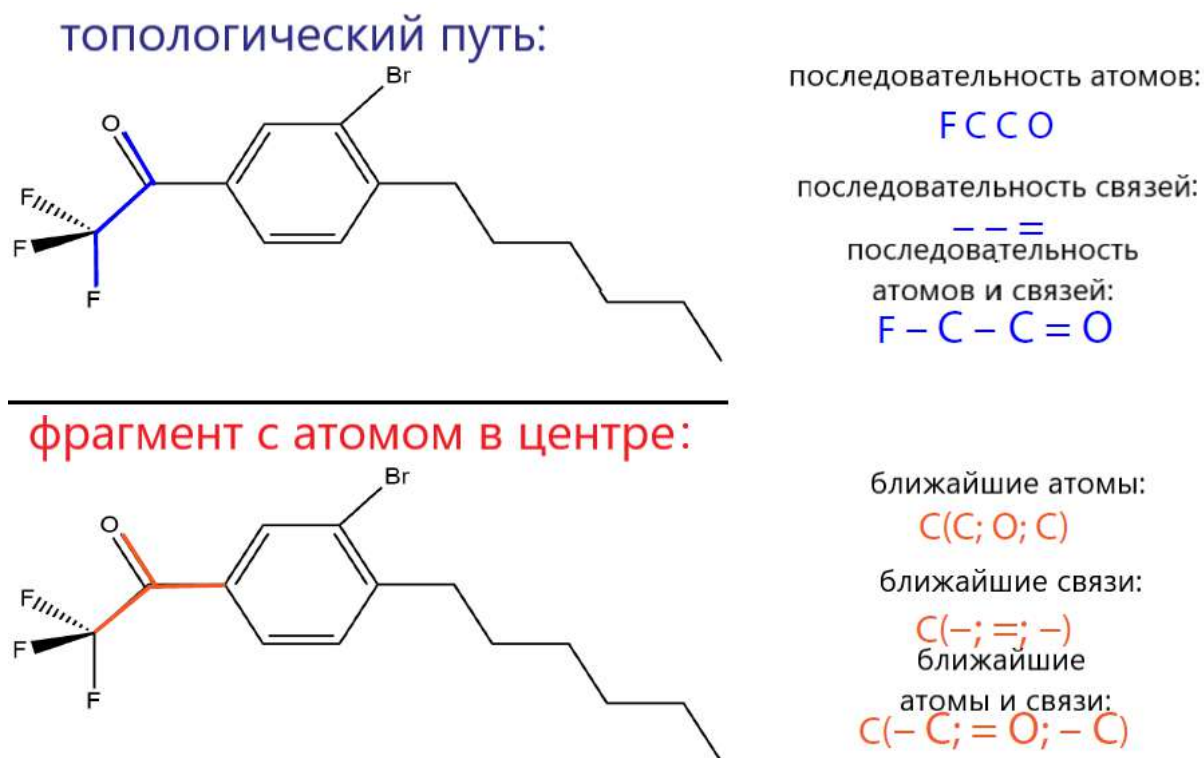


Рисунок 10. Два подхода к получению подструктурных молекулярных фрагментов

В данном исследовании в качестве структурных дескрипторов для описания молекул были выбраны последовательности атомов и связей, которые удовлетворяли следующим условиям: дескриптор представляет собой кратчайший путь от одного атома к другому; его длина составляет от 4 до 10 атомов. Последнее ограничение обусловлено тем, что короткие фрагменты не несут значительной информации о структуре молекулы, представлены в большем количестве относительно более длинных дескрипторов, присутствуют в большинстве соединений. Тогда как дескрипторы с длиной более 10 атомов, напротив, представлены в крайне малом количестве соединений из набора данных, зачастую только в одном.

В результате генерации дескрипторов была получена матрица 40 * 1103, соответствующая количеству литературных соединений * рассчитанные

структурные дескрипторы, которая была дополнена еще одним столбцом, содержащим диэлектрические константы пластификаторов, использованных при приготовлении сенсорной мембраны. Полученная матрица была использована для построения регрессионной модели, связывающей эти дескрипторы со значениями селективности $\log K^{\text{сел}}(\text{HCO}_3^-/\text{Cl}^-)$ соответствующих потенциометрических сенсоров.

Для построения многомерных регрессий применялся алгоритм PLS. Модели PLS рассчитывались в программном пакете The Unscrambler 9.7 (CAMO, Норвегия). Для оптимизации количества латентных переменных и количества дескрипторов в модели применялась полная кросс-валидация с использованием в качестве критерия значения среднеквадратичной ошибки кросс-валидации.

Перед моделированием значения дескрипторов в матрице X были автомасштабированы (из каждого элемента вычиталось среднее значение по столбцам и результат делился на стандартное отклонение по столбцам). Полученная модель была использована для предсказания значений $\log K^{\text{сел}}(\text{HCO}_3^-/\text{Cl}^-)$ для новых лигандов.

2.2. Обсуждение результатов

Матрица молекулярных дескрипторов, полученная для 40 лигандов, была связана с их значениями $\log K^{\text{сел}}(\text{HCO}_3^-/\text{Cl}^-)$ с помощью PLS-регрессии. Количество переменных в исходной модели было оптимизировано с помощью анализа значений регрессионных коэффициентов модели. Все переменные, имеющие коэффициенты регрессии в диапазоне $[-5 \cdot 10^{-3}; 5 \cdot 10^{-3}]$, были исключены из рассмотрения, что позволило сократить количество переменных до 585 из первоначальных 1103. Полученная модель QSPR представлена на Рисунке 11. На основании статистических показателей модели (значения RMSECV и R^2) можно сделать вывод, что полученная зависимость подходит для полуколичественного

прогнозирования потенциометрической селективности ионофоров. Значения $\log K^{\text{сел}}(\text{HCO}_3^-/\text{Cl}^-)$ варьируются от -6 до +6 в смоделированном наборе данных, а значение RMSECV составляет 1,5. Таким образом, модель позволяет различать слабо-, средне- и высокоселективные карбонатные лиганды. Это представляется многообещающим результатом, учитывая широкое химическое разнообразие лигандов, сравнительно небольшой (в масштабах подхода QSPR) набор данных, использование литературных данных, полученных в неодинаковых условиях, и общую простоту подхода.

Анализ коэффициентов регрессии PLS-модели позволяет судить о важности тех или иных дескрипторов и их вкладе в значения селективности. Наибольший вклад вносят фрагменты, имеющие наибольшие абсолютные значения коэффициентов регрессии. Рассчитанные фрагменты (переменные модели) отфильтровывались по присутствию в химических структурах различающихся лигандов. Для анализа выбирались структурные дескрипторы, присутствующие как минимум в пяти лигандах. Затем оставшиеся фрагменты анализировались на повторяемость в других фрагментах, и из ряда эквивалентных фрагментов выбирались наименьшие. Таким образом, фрагмент C-C-C-O-C=C содержит более короткий фрагмент C-O-C=C, и этот более короткий фрагмент был оставлен для анализа. Наконец, фрагменты были отсортированы по значению соответствующих коэффициентов регрессии, и только фрагменты с абсолютными значениями $b > 0,01$ были рассмотрены. Полученная диаграмма важности переменных представлена на Рисунке 12.

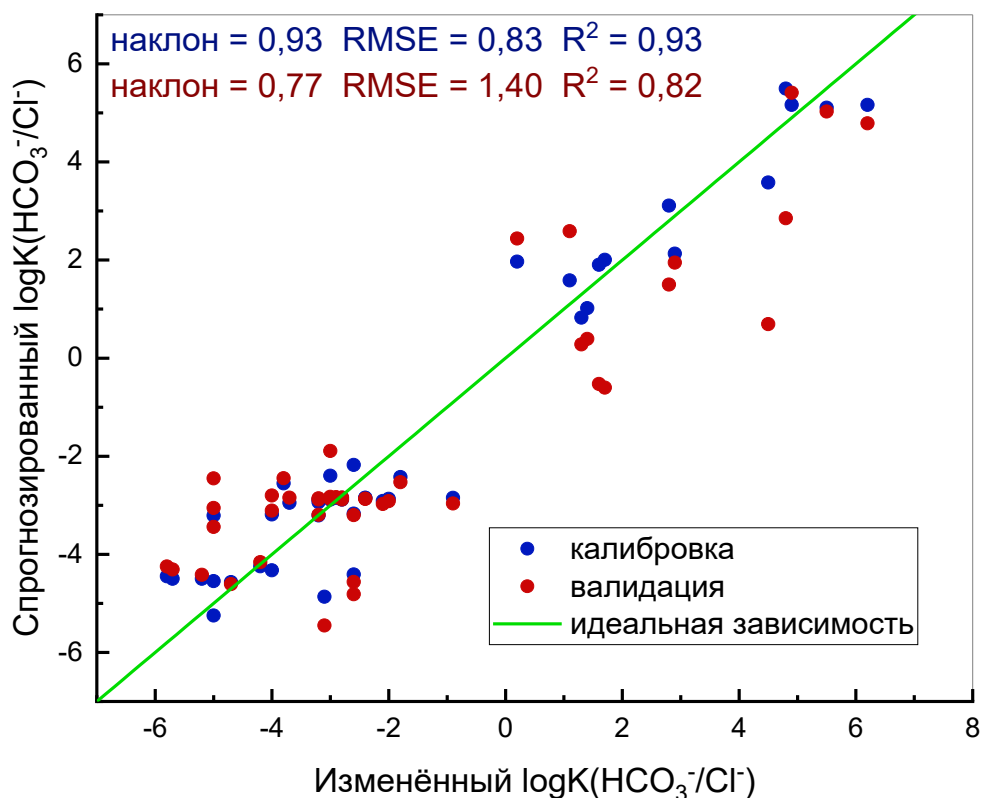


Рисунок 11. График измеренных и спрогнозированных значений кросс-валидированной модели QSPR для прогнозирования селективности мембранных сенсоров. Зелёная линия соответствует идеальному совпадению измеренных и предсказанных значений

Из диаграммы видно, что наибольший отрицательный вклад соответствует фрагменту $\text{C}=\text{CC}=\text{C}-\text{C}-\text{C}-\text{F}$. Моделируемое значение – $\log K^{\text{сел}}(\text{HCO}_3^-/\text{Cl}^-)$, и чем оно более отрицательное, тем выше селективность к HCO_3^- . Это означает, что фрагменты $\text{C}=\text{C}-\text{C}=\text{C}-\text{C}-\text{C}-\text{F}$ способствуют большей селективности к гидрокарбонатному аниону. Следует отметить, что среди 17 фрагментов с отрицательным вкладом есть только 2, которые не содержат фтора (F) или C=O групп. Это фрагменты $\text{C}=\text{C}-\text{C}=\text{C}-\text{C}-\text{O}$ и $\text{C}-\text{C}-\text{N}-\text{C}$, которые также имеют значительный отрицательный вклад среди выбранных фрагментов (15 и 16 место из 17).

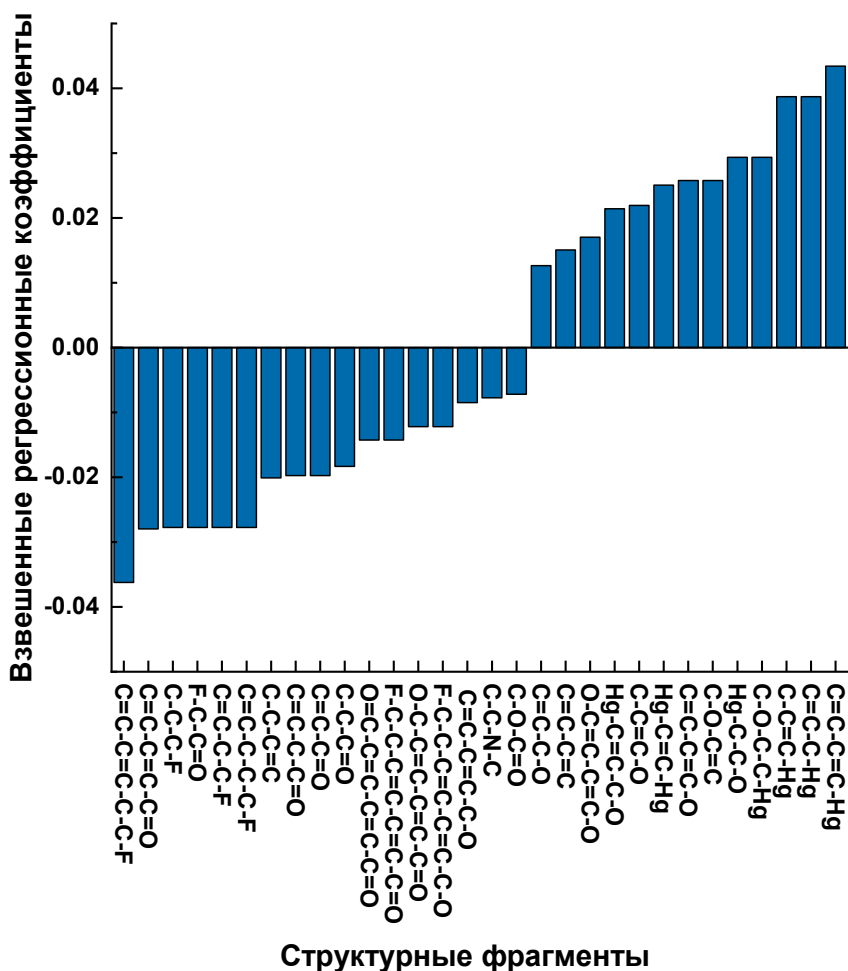


Рисунок 12. Фрагменты с наибольшим вкладом в значения $\log K^{сел} (HCO_3^-/Cl^-)$, где «-» - одинарная связь, а «=» - двойная связь

Десять фрагментов с наибольшим отрицательным вкладом приходятся на фрагмент трифторацетофенона. Трифторацетофенон входит в состав значительного количества существующих карбонатных ионофоров. Отличным примером ионофора с этой группой и высокой селективностью ($\log K^{сел}(HCO_3^-/Cl^-) = -3,1$) является N, N-Диоктил-3, 12-бис(4-трифторацетилбензоилокси)-5-холан-24-амид (карбонатный ионофор VII из каталога Merck). Структура этого соединения приведена на Рисунке 13. Известно, что группа TFA способствует связыванию карбонат-аниона через образование водородных связей [82].

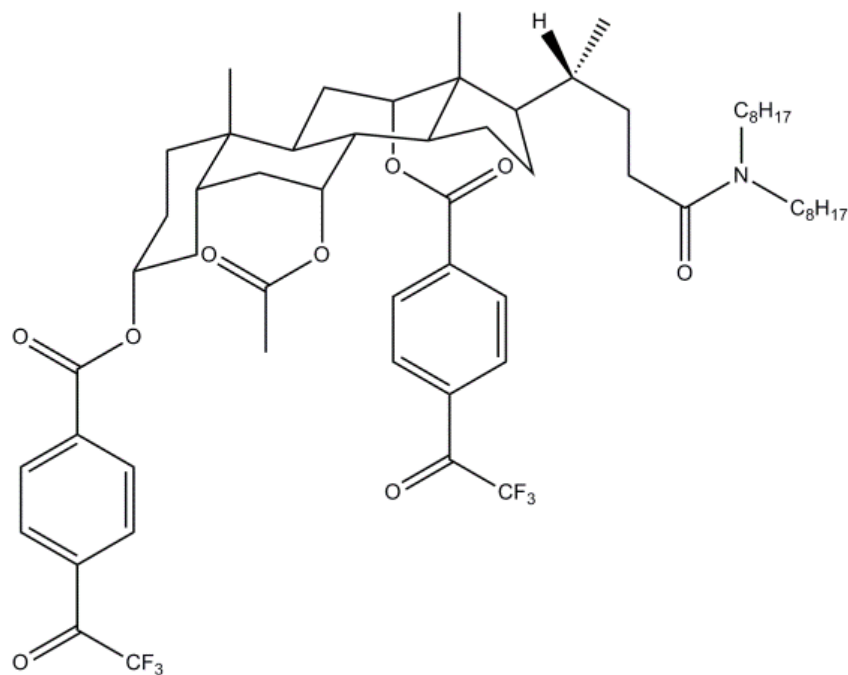


Рисунок 13. Химическая структура *N, N*-Диоктил-3, 12-бис(4-трифторацетилбензоилокси)-5-холан-24-амида

Фрагмент с наибольшим положительным вкладом - C=C-C=C-Hg. Только 6 фрагментов из 13 фрагментов с положительным вкладом не содержат ртути (Hg⁺) в своем составе. В порядке убывания значений вклада — это фрагменты C-O-C=C, C=C-C=C-O, C-C=C-O, O-C=C-C=C-O, C=C-C=C и C=C-C-O. Эти фрагменты принадлежат к ионофорам, предложенным для определения хлоридов, и поэтому обладают низкой селективностью по отношению к карбонату [79]. Примером такого ионофора является $\{\mu\text{-[4,5-диметил-3,6-бис(октилокси)-1,2-фенилен]}\}$ бис(трифторацетато-0) димеркурий, в структуре которого присутствуют все фрагменты с положительным вкладом, что приводит к $\log K^{\text{сел}}(\text{HCO}_3^-/\text{Cl}^-) = 5,5$. Структура этого лиганда приведена на Рисунке 14.

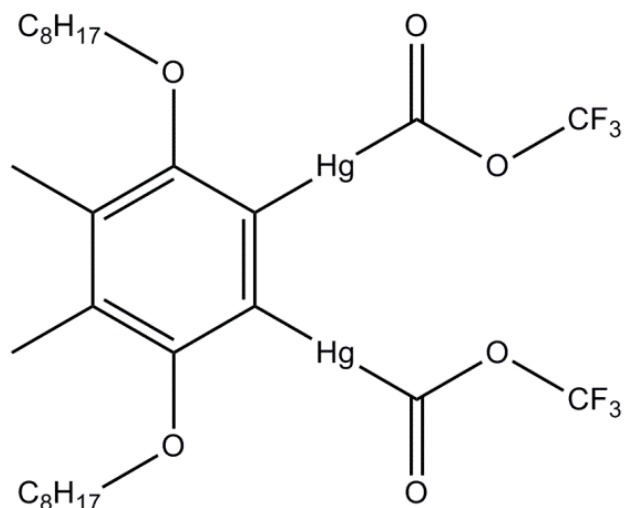


Рисунок 14. Химическая структура $\{\mu\text{-}[4,5\text{-диметил-}3,6\text{-бис(октилокси)-}1,2\text{-фенилен}]\text{бис(трифторацетато-}O)\text{ димеркурия}$

Таким образом, можно утверждать, что построенная QSPR-модель хорошо согласуется с химическими соображениями о структуре лигандов, а значимость различных фрагментов соответствует общехимическим соображениям.

Полимерные сенсорные мембраны, приготовленные с использованием четырех новых лигандов, были изучены на предмет их чувствительности к неорганическим анионам и селективности $\text{HCO}_3^- / \text{Cl}^-$. Типичные кривые отклика новых сенсоров приведены на Рисунке 15, а рассчитанные значения чувствительности в Таблице 3. Следует отметить, что сенсоры S2 и S3 не продемонстрировали достаточной чувствительности к карбонату.

С помощью модели QSPR, полученной на основе литературных данных, была спрогнозирована селективность этих четырёх новых мембранных сенсоров. Для этого каждый из лигандов был описан с помощью одного и того же набора SMF, а полученные данные были приведены в соответствие с отфильтрованным набором из 585 переменных. Полученные таким образом описания структур лигандов были использованы в модели QSPR для расчёта спрогнозированных значений $\log K^{\text{сел}}(\text{HCO}_3^- / \text{Cl}^-)$. Результаты этого прогноза приведены в Таблице 4 вместе с

данными по селективности, полученными в традиционном потенциометрическом эксперименте с использованием метода отдельных растворов.

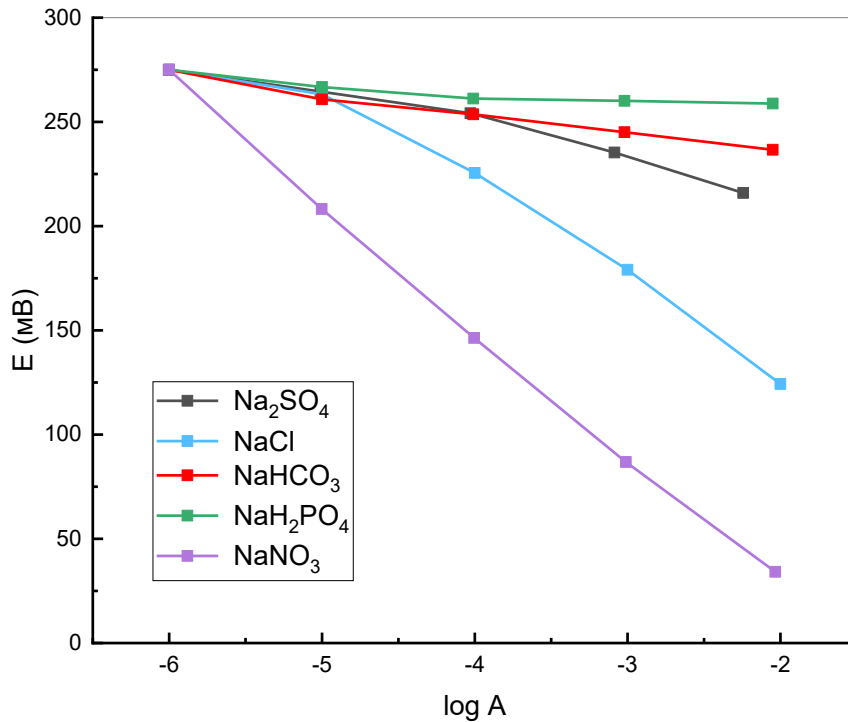


Рисунок 15. Типичные кривые потенциометрического отклика сенсора S2. Показания ЭДС смещены к нулю для большей наглядности

Таблица 3. Значения чувствительности сенсоров в растворе исследуемых анионов, мВ/дек

	SO ₄ ²⁻	Cl ⁻	HCO ₃ ⁻	H ₂ PO ₄ ⁻	NO ₃ ⁻
S1	1,7 ± 0,3	-6,4 ± 1,0	-43,4 ± 3,8	29,6 ± 1,1	-27,5 ± 0,9
S2	-16,6 ± 1,1	-41,0 ± 1,2	-7,2 ± 3,4	-1,5 ± 1,2	-57,8 ± 0,8
S3	-24,0 ± 1,6	-42,5 ± 3,8	-16,4 ± 4,2	-3,5 ± 1,4	-64,9 ± 5,9
S4	-1,2 ± 1,2	-23,8 ± 1,0	-28,7 ± 1,6	12,4 ± 1,4	-47,1 ± 0,7

Таблица 4. Сравнение спрогнозированных и экспериментальных значений $\log K^{\text{сел}}(\text{HCO}_3^-/\text{Cl}^-)$ для новых ионофоров

Сенсор	Экспериментальное значение $\log K^{\text{сел}}(\text{HCO}_3^-/\text{Cl}^-)$	Спрогнозированное значение $\log K^{\text{сел}}(\text{HCO}_3^-/\text{Cl}^-)$
S1	-3,6	-2,5
S2	-0,7	-3,4
S3	-1	-2,9
S4	-2,5	-2,7

Учитывая значение RMSECV модели QSPR, равное 1,4 $\log K^{\text{сел}}(\text{HCO}_3^-/\text{Cl}^-)$, соответствие спрогнозированной и экспериментальной селективностей оказывается удовлетворительным в двух из четырёх случаев (S1, S4). Соединение S1 имеет фрагменты C=C-C=C=O, C-C-C=C, C=C-C-C=O, C=C-C=O, C-C-C=O с отрицательным вкладом и фрагмент C=C-C=C с положительным вкладом. Модель спрогнозировала логарифм коэффициента селективности со значением -2,5, в то время как экспериментально найденный коэффициент селективности составляет -3,6. Соединение S2 содержит фрагмент C-C-N-C, а остальные фрагменты являются теми же фрагментами, что и в первом соединении. Все эти фрагменты, кроме C=C-C=C, вносят отрицательный вклад в логарифм коэффициента селективности. Таким образом, наблюдается заметное отклонение между предсказанным $\log K^{\text{сел}}(\text{HCO}_3^-/\text{Cl}^-)$ -3,4 и экспериментальным значением -0,7. В соединении S3 имеется всего несколько фрагментов, оказывающих значительное влияние на модель. Это фрагменты C-C-C=C с отрицательным вкладом и C=C-C=C с положительным вкладом. Поскольку модель имеет только два структурных фрагмента со значительным влиянием для этого соединения, разница между прогнозом модели (-3,0) и экспериментальными данными (-1,0) также высока. Довольно большое расхождение между прогнозом и

экспериментом для S2 и S3 можно также объяснить тем, что эти сенсоры фактически не показали достаточного отклика на карбонат (Таблица 3).

В состав соединения S4 входят такие фрагменты, как C=C-C=C-C=O, C=C-C=O, C=C-C=O, C-C-C=O, C-C-C=C, O-C-C=C-C=C-C=O и C-C-N-C с отрицательным вкладом и C-O-C=C, C=C-C=C-O, C-C=C-O и C=C-C=C с положительным вкладом. Поскольку количество важных фрагментов достаточно велико, согласованность между экспериментом и прогнозом также довольно высока (-2,7 против -2,5 соответственно).

Таким образом, для корректного прогнозирования свойств новых лигандов с помощью модели QSPR необходимо, чтобы в их структуре присутствовали те же важные фрагменты, что и у лигандов, использованных для построения модели. Этот вывод следует из химических соображений и математических принципов, поскольку разнообразие обучающего набора определяет область применимости модели.

Заключение к Главе 2

На основе литературных данных по 40 ионофорам была разработана QSPR-модель, связывающая структуру карбонатных ионофоров с селективностью ($\log K^{\text{сел}}(\text{HCO}_3^- / \text{Cl}^-)$) соответствующих потенциометрических сенсоров с пластифицированной полимерной мембраной. Анализ коэффициентов регрессии в PLS-модели позволил выявить важные структурные фрагменты, влияющие на селективность.

Для проверки прогностической способности модели были использованы синтезированные для этой работы четыре новых лиганда с различными акцепторными заместителями при карбонильной группе. Два из этих лигандов продемонстрировали потенциометрическую чувствительность к гидрокарбонат-ионам. Для этих лигандов наблюдалось хорошее согласие между

спрогнозированными и экспериментальными значениями селективности, особенно в случае лигандов, имеющих общие важные молекулярные фрагменты с ионофорами, использованными для построения модели.

Полученные результаты свидетельствуют о потенциале подхода QSPR для разработки новых анион-селективных сенсоров, позволяя прогнозировать селективность на основе структурных характеристик ионофоров.

Глава 3. ПРИМЕНЕНИЕ QSPR-МОДЕЛИРОВАНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ ЧУВСТВИТЕЛЬНОСТИ ИОНОСЕЛЕКТИВНЫХ ЭЛЕКТРОДОВ К КАТИОНАМ ТЯЖЕЛЫХ МЕТАЛЛОВ (Cu^{2+} , Cd^{2+} И Pb^{2+})

В данной главе исследуется возможность применения метода QSPR для прогнозирования потенциометрической чувствительности пластифицированных полимерных мембранных сенсоров на основе химической структуры ионофора. QSPR-модель разрабатывается на основе литературных данных о чувствительности ранее изученных структурно подобных ионофоров. Оценивается эффективность полученной модели в установлении связи между структурами ионофоров и их чувствительностью к Cu^{2+} , Cd^{2+} и Pb^{2+} . Прогнозы модели для четырёх новых ионофоров на основе дифенилфосфорил ацетамида сопоставляются с экспериментальными данными, полученными для этих же ионофоров.

Несмотря на успешное применение QSPR-моделирования для прогнозирования сенсорных свойств потенциометрических сенсоров на основе ионофоров, включая параметры чувствительности и селективности, до настоящего времени этот подход не использовался для прогнозирования чувствительности новых мембранных сенсоров на основе еще не изученных ионофоров. Данная работа направлена на восполнение этого пробела путем разработки QSPR-модели на основе литературных данных о чувствительности к Cu^{2+} , Cd^{2+} и Pb^{2+} для структурно подобных ионофоров. Выбор этих конкретных металлов был продиктован их токсичностью и, следовательно, выраженным интересом к этим элементам в экологических исследованиях.

3.1. Экспериментальная часть

3.1.1. Выбор дескрипторов

Для построения QSPR моделей для прогнозирования потенциометрической чувствительности были собраны литературные данные об откликах сенсоров с ПВХ-пластифицированными мембранами на ионы металлов Cu^{2+} , Cd^{2+} и Pb^{2+} . Выбор именно этих металлов был продиктован их токсичностью и, следовательно, выраженным интересом к этим элементам в экологических исследованиях. Большая часть набора данных (35 структур и данные об их потенциометрических откликах) была взята из [60]. Для улучшения обучающего набора по фосфорсодержащим ионофорам в него были добавлены шесть веществ из [83] (№36-41, Таблица 1, ПриложениеА).

Структуры ионофоров, их названия по ИЮПАК и литературные ссылки приведены в Таблице 1 в Приложении А.

Структуры ионофоров были описаны с помощью молекулярных дескрипторов. В данной работе, аналогично выбору дескрипторов, описанному в разделе 2.1.1, для этого использовались подструктурные молекулярные фрагменты, полученные с помощью программного комплекса ISIDA. Главная идея получения таких молекулярных дескрипторов в том, что подструктурный фрагмент представляет собой последовательность атомов химической структуры, соединенных химическими связями. Каждому фрагменту соответствует количество раз, которое этот фрагмент встречается в молекуле. Затем эти числа собираются в матрицу, где каждая строка соответствует определенному ионофору, а каждый столбец указывает на количество раз, которое определенный фрагмент встречается в ионофоре. В этом исследовании рассматривались фрагменты длиной от 2 до 9 атомов. Итоговая матрица с дескрипторами имеет размерность $41 * 1095$, где 41 – количество ионофоров в обучающем наборе, а 1095 – количество

различных найденных молекулярных дескрипторов, описывающие структуры ионофоров, составляющие обучающий набор.

3.1.2. QSPR моделирование

Для того чтобы найти математическую зависимость, связывающую химические структуры ионофоров, представленные молекулярными дескрипторами, и потенциметрическую чувствительность этих ионофоров к ионам меди, кадмия и свинца, был применена PLS регрессия.

Для нахождения оптимального числа латентных переменных модели PLS подвергались полной кросс-валидации, такая разновидность кросс-валидации была выбрана из-за сравнительно малого размера обучающего набора.

Рассчитанные QSPR модели были использованы для прогнозирования потенциметрической чувствительности к меди, кадмию и свинцу для четырех новых синтезированных ионофоров. Структуры новых веществ и их названия по номенклатуре ИЮПАК приведены в Таблице 5.

Таблица 5. Структуры новых ионофоров и их названия по номенклатуре ИЮПАК

	Химическая структура	Название по ИЮПАК
1		3,3'-((этан-1,2-диилбис(пропилазандил))бис(1-(дифенилфосфорил)пропан-2-он)

Таблица 5. (продолжение)

	Химическая структура	Название по ИЮПАК
2		3,3'-(бутан-1,4-диилбис(октилазандиил))бис(1-(дифенилфосфорил)пропан-2-он)
3		3,3'-(гексан-1,6-диилбис(октилазандиил))бис(1-(дифенилфосфорил)пропан-2-он)
4		1,1'-(пиперазин-1,4-диил)бис(2-(дифенилфосфорил)этанон)

3.2. Обсуждение результатов

На первом этапе эксперимента были построены три отдельные PLS-модели, связывающие молекулярные дескрипторы ионофоров из обучающего набора с их потенциометрической чувствительностью к ионам меди, кадмия и свинца. На Рисунке 16 показан типичный результат такого моделирования: график зависимости измеренной чувствительности от спрогнозированной для ионов кадмия. Видно, что наблюдается достаточно хорошая корреляция, а ошибка при кросс-валидации составила около 4 мВ/дек. Эти результаты позволяют предположить, что формализованное описание ионофора с помощью

подструктурных молекулярных фрагментов может быть использовано для прогнозирования чувствительности соответствующих пластифицированных полимерных сенсорных мембран.

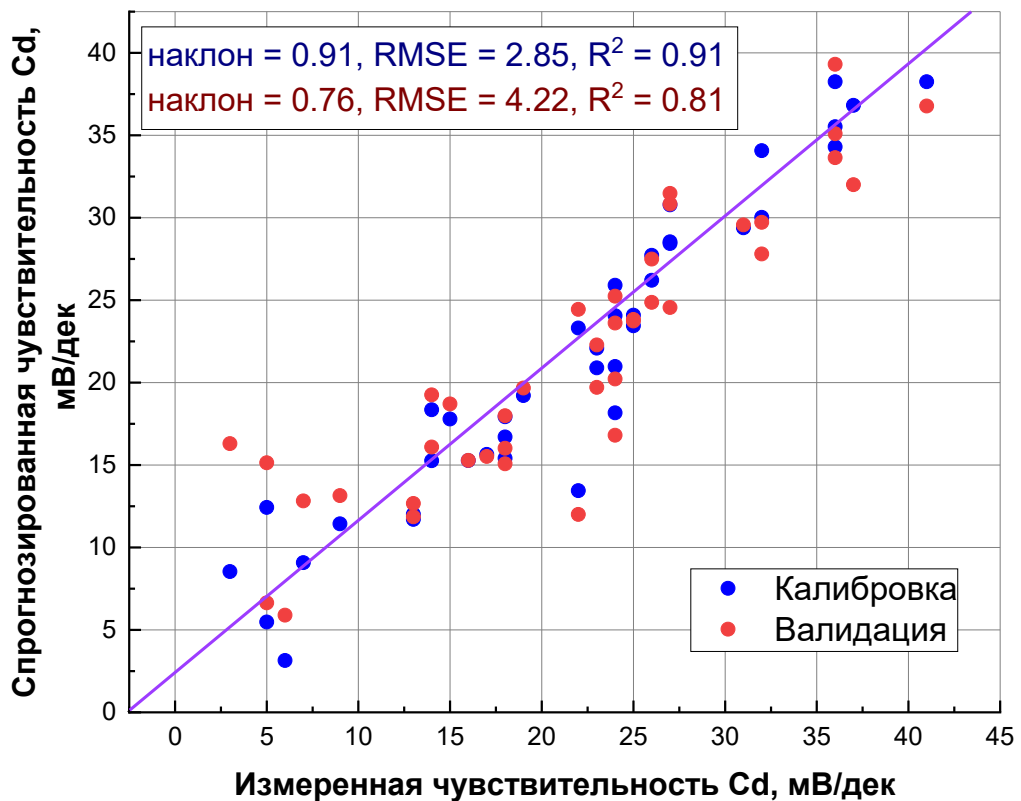


Рисунок 16. График введено-найдено модели чувствительности к ионам кадмия (четыре латентные переменные) с отмеченной линией идеальной зависимости

По значениям коэффициентов регрессии в PLS-моделировании можно судить о значимости молекулярных фрагментов для конкретного целевого свойства. График коэффициентов регрессии - удобный способ визуализации этих значений. Взвешенные коэффициенты регрессии рассчитываются из взвешенной матрицы данных с дескрипторами в ходе PLS-регрессии. На Рисунке 17 представлен график для PLS-модели чувствительности к ионам кадмию: по оси X указан номер дескриптора, а по оси Y - численное значение конкретного коэффициента регрессии для каждого дескриптора. Чем выше абсолютное значение коэффициента регрессии, тем более значимым был вклад

соответствующего дескриптора. Видно, что некоторые дескрипторы не вносили существенного вклада в модель, а значит, могли нести в себе только шум. Для того чтобы оптимизировать модель, мы исключили все переменные (дескрипторы) в пределах отмеченной области на Рисунке 17, имеющие небольшие значения коэффициента регрессии между $-0,05$ и $+0,05$. Избавление от нерелевантных переменных привело к улучшению моделей. Аналогичная процедура была проведена для PLS-моделей чувствительностей к ионам меди и свинца, и полученные статистические данные приведены в Таблице 6. Графики «введено-найденно» для чувствительности к ионам меди и свинца представлены на Рисунках 18 и 19.

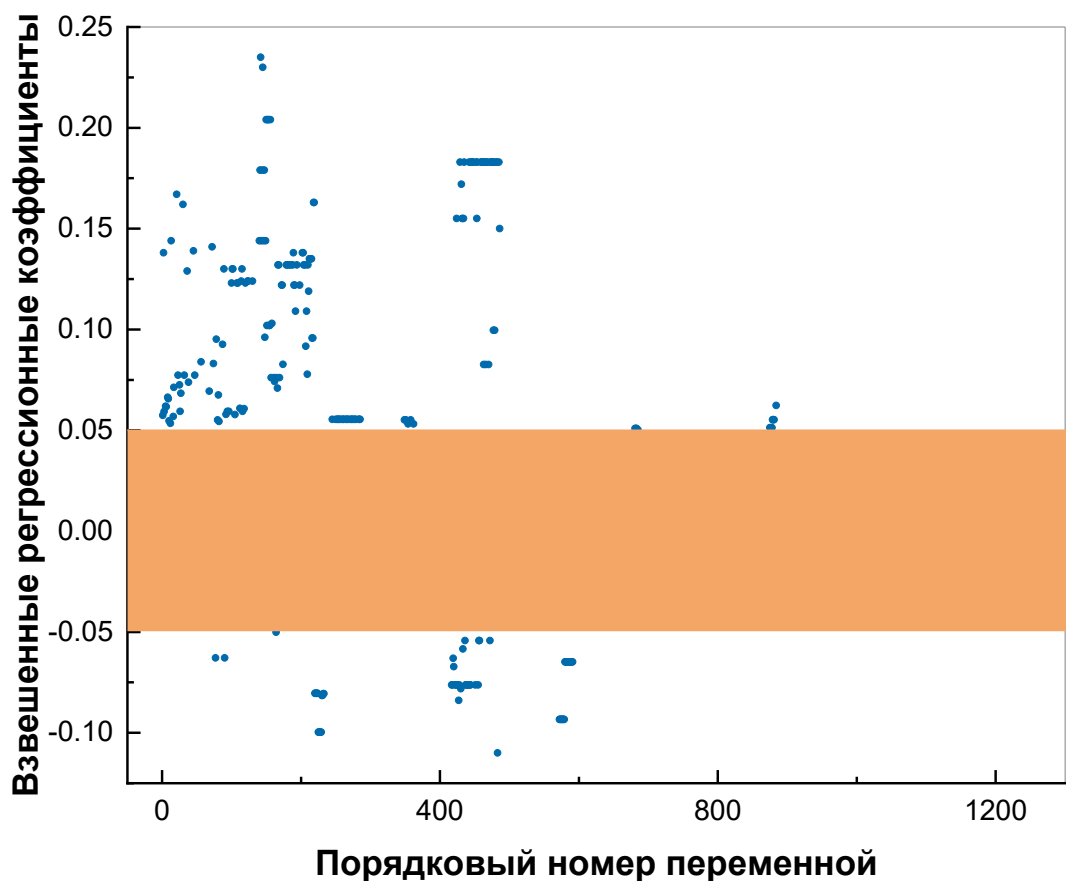


Рисунок 17. График коэффициентов регрессии для PLS-модели чувствительности к кадмию с отмеченной областью фрагментов с небольшими вкладами в модель

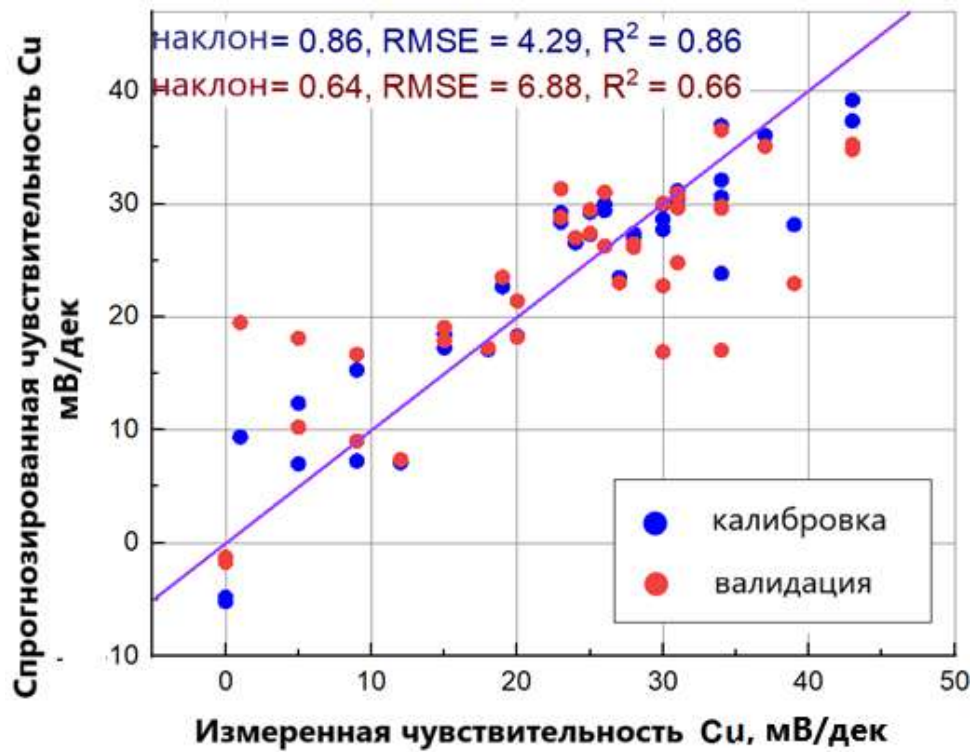


Рисунок 18. График введено-найдено модели чувствительности к ионам меди (две латентные переменные) с отмеченной линией идеальной зависимости

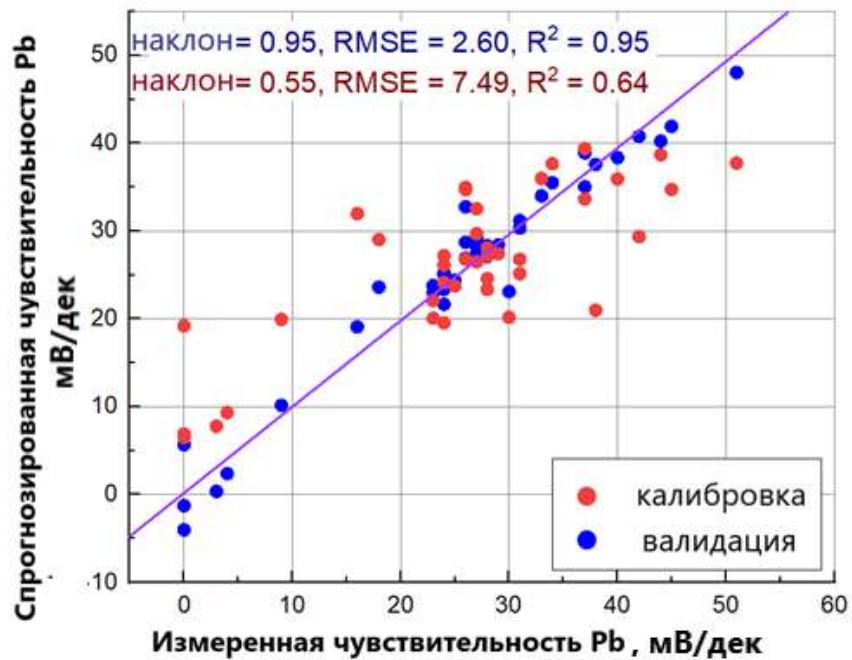


Рисунок 19. График введено-найдено модели чувствительности к ионам свинца (шесть латентных переменных) с отмеченной линией идеальной зависимости

Таблица 6. Характеристики моделей, прогнозирующих чувствительность к ионам кадмия, меди и свинца

		Наклон	RMSE	R ²
Cd	калибровка	0,91	2,85	0,91
	валидация	0,76	4,22	0,81
Cu	калибровка	0,86	4,29	0,86
	валидация	0,64	6,88	0,66
Pb	калибровка	0,95	2,60	0,95
	валидация	0,55	7,49	0,64

Видно, что наихудшие показатели модели наблюдались для модели чувствительности к ионам свинца, где RMSE при кросс-валидации составила более 7 мВ/дек. Очевидно, что такая модель подходит скорее для качественной оценки ионофоров с точки зрения их пригодности для создания сенсоров на ионы свинца, чем для численного прогнозирования. Несколько лучшие результаты были получены в случае чувствительности к ионам меди, однако они все равно носили полуколичественный характер с точки зрения точности прогнозирования.

Показателем достоверности модели QSPR было соответствие важных молекулярных фрагментов установленным химическим закономерностям. На Рисунке 20 показаны фрагменты, которые были важны для характеристики чувствительности к ионам кадмия. Они были отобраны при условии, что встречаются не менее чем в пяти структурах, чтобы избежать случайных корреляций. При построении диаграмм мы опускали фрагменты с углеводородными цепями на концах, если вложенные фрагменты имели схожий вклад. В качестве примера таких фрагментов можно рассмотреть пары C-C-P-C=C-C и C-C-P-C=C. На Рисунках 21 и 22 показаны эти диаграммы для чувствительности к ионам меди и свинца соответственно.



Рисунок 20. Фрагменты с высоким вкладом в значение чувствительности к кадмию

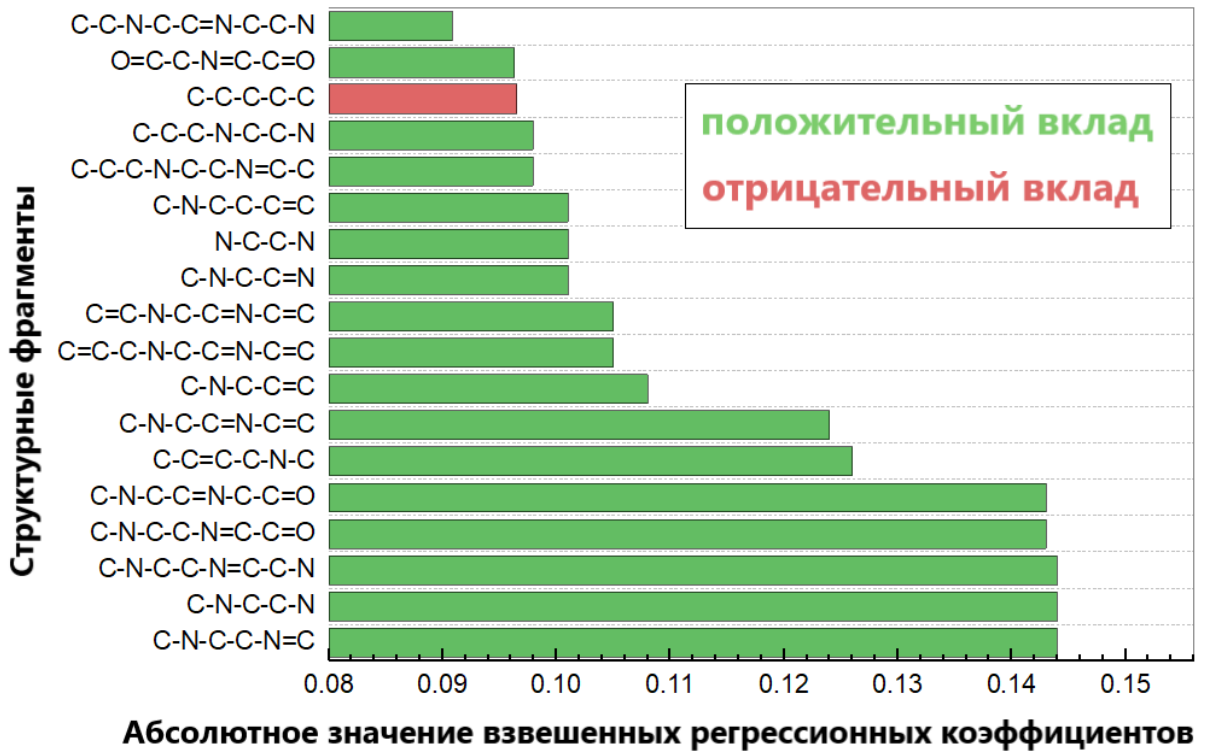


Рисунок 21. Фрагменты с высоким вкладом в значение чувствительности к меди

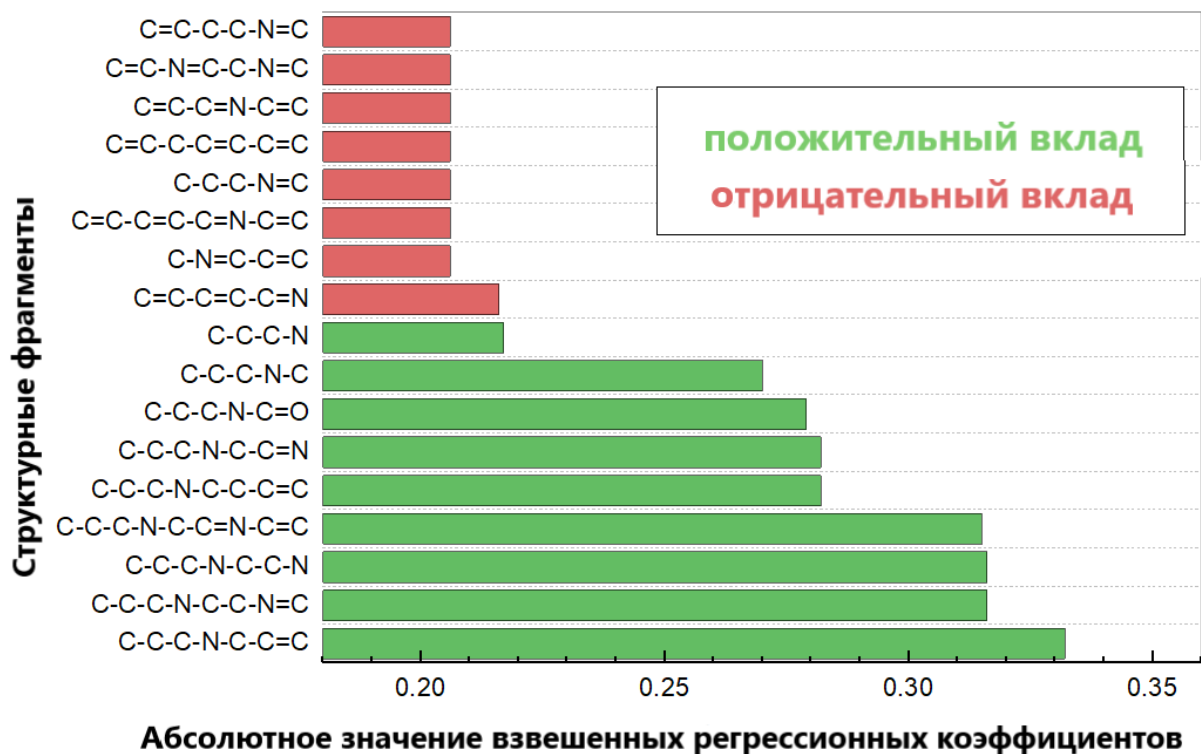


Рисунок 22. Фрагменты с высоким вкладом в значение чувствительности к свинцу

Анализ диаграммы на Рисунке 20 показывает, что фрагменты с наибольшим вкладом в чувствительность к кадмию включают атом фосфора или N,N-диметилпиколинамидный фрагмент (рисунок 23а). В отличие от чувствительности к меди (Рисунок 21), фрагменты структуры (C=C-C=C-C=N, C-N=C) в основном имеют отрицательный вклад. Наибольшей чувствительностью к кадмию в обучающем наборе данных (41 мВ/дек) обладал 4,4'-дибром-N6,N6'-диэтил-N6,N6'-бис(4-гексилфенил)-[2,2'-бипиридин]-6,6'-дикарбоксаимид. Химическая структура этого соединения приведена на Рисунке 23б. Соединение содержало фрагменты как с положительным (выделены зелёным), так и с отрицательным (выделены красным) вкладом. В целом, выявленные важные структурные фрагменты следовали химической логике комплексообразования: ионофоры с пиколинамидным фрагментом (Рисунок 23а) показали выраженную чувствительность к кадмию при

использовании в пластифицированных полимерных мембранах потенциометрических сенсоров (см. также Таблицу 6). То же самое наблюдалось и для зелёного фрагмента на Рисунке 23б, который представлял собой часть пиколинамида. Еще одно несоответствие экспериментальным наблюдениям - красный фрагмент на Рисунке 23б - в целом наличие пиридиновой структурной единицы наблюдалось для ионофоров как с высокой, так и со средней чувствительностью к Cd^{2+} . Мы считаем, что эти расхождения можно объяснить довольно небольшим размером использованного набора данных (всего 41 ионофор), в то время как традиционные исследования QSPR обычно включают сотни или тысячи записей. Следует отметить, что в случае потенциометрических экспериментов создание такого большого набора данных вряд ли возможно из-за различий в экспериментальных протоколах, используемых исследователями для изучения потенциометрических сенсоров (составы мембран, расчёты чувствительности, pH образцов).

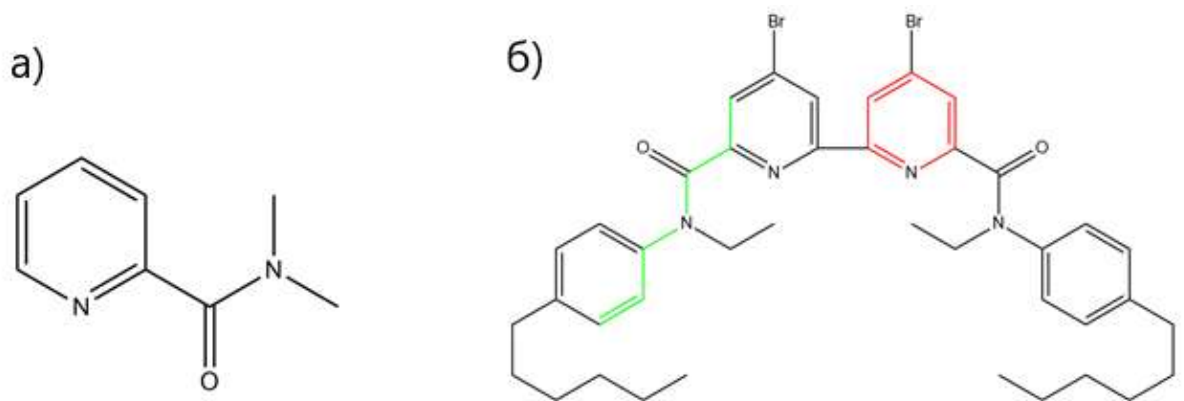


Рисунок 23. *N,N*-диметилпиколинамидный фрагмент (а), 4,4'-дибром-*N6,N6'*-диэтил-*N6,N6'*-бис(4-гексилфенил)-[2,2'-бипиридин]-6,6'-дикарбоксамид с выделенными C-C-C-N-C-C=C (зелёный) и C=C-C=C-C=N (красный) фрагментами (б)

Полученные PLS-модели были применены для прогнозирования потенциометрической чувствительности четырех новых ионофоров (1-4), структуры которых приведены в Таблице 5. Химические структуры этих соединений были проанализированы с помощью ранее оптимизированного

набора молекулярных фрагментов. Затем эти четыре ряда молекулярных дескрипторов послужили исходными данными для трех PLS-моделей, чтобы спрогнозировать чувствительность к каждому из катионов металлов. Результаты представлены в Таблице 7.

Соответствие прогнозов модели и экспериментальных данных было наиболее высоким для чувствительности к кадмию. Прогнозируемые значения чувствительности к меди и свинцу продемонстрировали некоторое несоответствие с экспериментальными данными. В частности, для соединений 2 и 3 прогнозируемые значения чувствительности к меди отличались от экспериментальных на 2 и 6 мВ/дек соответственно. Это объясняется отсутствием в структуре этих соединений фрагментов, вносящих наибольший вклад в чувствительность к меди (C-N-C-C-N=C, C-N-C-C-N и другие). Для свинца наибольшее отклонение прогноза от эксперимента наблюдалось для соединения 4, что, вероятно, связано с наличием фосфора в его составе. Несмотря на присутствие соединений с фосфором в обучающем наборе, среди фрагментов с наибольшей чувствительностью к свинцу отсутствуют фосфорсодержащие.

Таблица 7. Экспериментальная и спрогнозированная чувствительность. Значения стандартного отклонения для экспериментальных данных, полученных для трёх идентичных сенсоров в трёх повторных измерениях, не превышали 1,5 мВ/дек

	Cd ²⁺		Cu ²⁺		Pb ²⁺	
	Эксперимент	Прогноз	Эксперимент	Прогноз	Эксперимент	Прогноз
1	24,6	24,7	20,3	20,9	32,9	29,7
2	22,1	22,5	18,3	16,4	31,7	29,2
3	24,7	22,5	21,1	15,3	34,6	29,1
4	23,0	23,1	23,4	19,4	34,9	27,2

Увеличение длины алкильной цепи между атомами азота в лиганде (вещества 1, 2 и 3) коррелировало с повышением чувствительности сенсоров к кобальту и никелю. В то же время, для остальных металлов эта тенденция не наблюдалась. Вещество 4 с пиперазиновым мостиком демонстрировало отличное от веществ с алкильными связями поведение. Можно предположить, что механизм комплексообразования между лигандом и ионом-мишенью, основанный на взаимодействии кислорода оксида фосфина и кето-группы с определенным влиянием атомов азота, является общим для всех исследованных веществ. Однако, для более детального понимания механизма взаимодействия необходимы дополнительные исследования.

Сравнение прогнозов QSPR-модели и результатов потенциометрического эксперимента, представленное в Таблице 7, демонстрирует высокую степень соответствия в большинстве случаев. Особенно точное совпадение наблюдалось для кадмия, в то время как для свинца прогнозируемые значения были систематически ниже экспериментальных. Тем не менее, относительный порядок увеличения чувствительности от меди к свинцу хорошо согласовывался с прогнозами модели.

Заключение к Главе 3

В данной главе продемонстрирована возможность успешного применения QSPR-модели, обученной на литературных данных по потенциометрической чувствительности пластифицированных полимерных мембранных сенсоров, для *in silico* прогнозирования потенциометрического поведения новых ионофоров без необходимости проведения их синтеза и экспериментальной характеристики. Достоверность прогнозов модели для чувствительности к Cu^{2+} , Cd^{2+} и Pb^{2+} была подтверждена экспериментально для четырех новых дифенилфосфорилацетамидных ионофоров. Высокая степень соответствия между

прогнозами модели и экспериментальными значениями чувствительности свидетельствует о значительном потенциале применения QSPR-подхода для разработки новых пластифицированных мембранных сенсоров с заданными характеристиками.

Глава 4. ПРИМЕНЕНИЕ QSPR-МОДЕЛИРОВАНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ ФИЗИКО-ХИМИЧЕСКИХ СВОЙСТВ ГЛУБОКИХ ЭВТЕКТИЧЕСКИХ РАСТВОРИТЕЛЕЙ НА ОСНОВЕ ХОЛИН ХЛОРИДА

В данной главе исследуется возможность применения метода QSPR для прогнозирования физико-химических свойств глубоких эвтектических растворителей на основе холин хлорида с органическими кислотами в качестве доноров водородной связи. Цель данного исследования – разработать эффективные QSPR-модели, способные предсказывать свойства ГЭР на основе их молекулярной структуры и состава, что позволит оптимизировать выбор компонентов и соотношений для получения ГЭР с заданными свойствами. Это особенно актуально в связи с растущим интересом к ГЭР как к экологически безопасным и эффективным растворителям для различных применений.

Для построения QSPR-моделей используется набор данных, состоящий из 36 образцов ГЭР с различными составами, в которых варьировались органические кислоты и содержание воды. Такой подход позволяет учесть влияние как природы кислоты, так и содержания воды на свойства ГЭР. В качестве независимых переменных выбраны подструктурные молекулярные фрагменты, характеризующие химическую структуру кислот, полуэмпирические параметры, полученные с помощью квантово-химического расчета PM3, отражающие электронные свойства молекул, а также процентное содержание воды в каждом растворе. Зависимыми переменными являются вязкость, коэффициент преломления, электропроводность и плотность, определённые экспериментально. Выбор этих свойств обусловлен их важностью для практического применения ГЭР в различных областях.

Прогностическая способность разработанных моделей дополнительно проверяется на двух ГЭР, не входивших в обучающий набор. Это позволяет оценить эффективность моделей для прогнозирования свойств новых, ранее не изученных ГЭР.

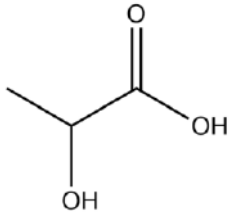
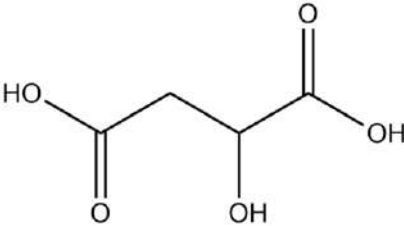
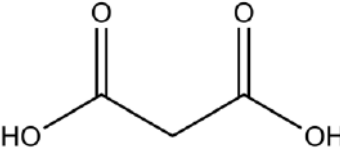
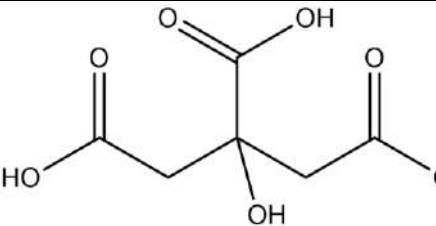
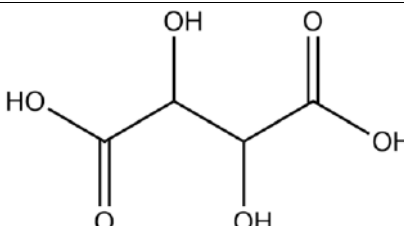
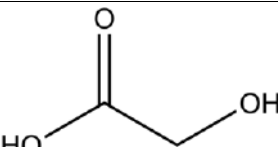
4.1. Экспериментальная часть

4.1.1. Выбор дескрипторов

При разработке моделей QSPR для ГЭР на основе карбоновых кислот целью являлось установление взаимосвязи между дескрипторами и зависимыми переменными, представленными вязкостью, плотностью, электропроводностью и коэффициентом преломления. Для моделирования с помощью подхода QSPR был собран набор данных, содержащий 36 образцов ГЭР, в составах которых изменяются органические кислоты и содержание воды. Из них два образца были исключены из расчётов, поскольку кислоты, содержащиеся в этих ГЭР, в концентрированном виде (содержание воды 0%) являются кристаллами. В отличие от предыдущих исследований, описанных в Главах 2 и 3, набор независимых переменных включал в себя не только подструктурные молекулярные фрагменты, но и полуэмпирические параметры, полученные с помощью квантово-химического расчёта PM3, а также процентное содержание воды в каждом растворе. Зависимые переменные – вязкость, коэффициент преломления, электропроводность и плотность – были получены экспериментально и предоставлены коллегами с кафедры аналитической химии, Татьяной Бочко и Андреем Шишовым. В Таблице 2 в Приложении А представлены измеренные характеристики ГЭР, а в Таблице 8 приведены тривиальные и IUPAC названия и структуры исследуемых органических кислот.

Структурные дескрипторы были рассчитаны с помощью программного пакета ISIDA аналогично экспериментам, описанным в Главах 2 и 3. Поскольку химические соединения в обучающем наборе этого эксперимента значительно меньше, чем в описанных ранее экспериментах, то размер подструктурных молекулярных фрагментов был выбран от двух до четырёх атомов.

Таблица 8. Органические кислоты, составляющие обучающий набор, их структуры и названия по номенклатуре ИЮПАК

№	Тривиальное название	Структура	Название ИЮПАК
1	Молочная кислота		2-гидроксипропановая кислота
2	Яблочная кислота		2-гидроксипропан-1,2,3-трикарбоновая кислота
3	Малоновая кислота		пропандиовая кислота
4	Лимонная кислота		2-гидроксипропан-1,2,3-трикарбоновая кислота
5	Винная кислота		2,3-дигидроксипропан-1,2,3-трикарбоновая кислота
6	Гликолевая кислота		2-гидроксиэтановая кислота

Квантово-химические дескрипторы определяли с помощью программного обеспечения HyperChem компании Hypercube, Inc. (<http://www.hypercubeusa.com/>) с использованием параметрического метода PM3.

PM3 позволяет проводить квантовые расчёты электронной структуры молекул в вычислительной химии. С помощью HyperChem были определены несколько дескрипторов, таких как энергии высшей занятой молекулярной орбитали (HOMO) и низшей незанятой молекулярной орбитали (LUMO), дипольный момент и теплоемкость (рассчитанная для 300K). Эти расчёты добавили дополнительные данные к обучающему набору и позволили подробнее анализировать природу полученных зависимостей.

Комплексное использование структурных дескрипторов вместе с полуэмпирическими параметрами обеспечивает надежное представление независимых переменных в рамках моделирования QSPR, повышая предсказательную способность моделей.

4.1.2. QSPR моделирование

Для установления математической связи между набором дескрипторов для исследуемых ГЭР и их вязкостью, показателем преломления, плотностью и проводимостью мы использовали PLS регрессию в сочетании с кросс-валидацией для определения оптимального количества латентных переменных, обеспечивающих наибольшую предсказательную способность. Также был произведён отбор дескрипторов по их вкладу в модели, и для уменьшения шума из набора данных были исключены те дескрипторы, которые оказывали незначительное влияние на модель.

В связи с ограниченным размером обучающего набора данных, было необходимо провести тщательную проверку качества модели и её способности к обобщению данных. Для проверки статистической значимости моделей была применена вложенная кросс-валидация, также известная как двойная кросс-валидация. Такой вид перекрёстной проверки является с точки зрения вычислений более затратным, чем K-кратная или полная кросс-валидация, но при этом позволяет избежать переобучения модели. Вложенная кросс-валидация включает

в себя два этапа, для наглядности представленные на Рисунке 24: внутренний этап для выбора модели или настройки гиперпараметров, в первую очередь, количество латентных переменных, и внешний этап, полная кросс-валидация, для оценки эффективности модели. Вложенная кросс-валидация имеет две итерации разделения данных на несколько обучающих и проверочных наборов, тем самым обеспечивает более достоверную оценку эффективности и снижает риск переобучения модели.

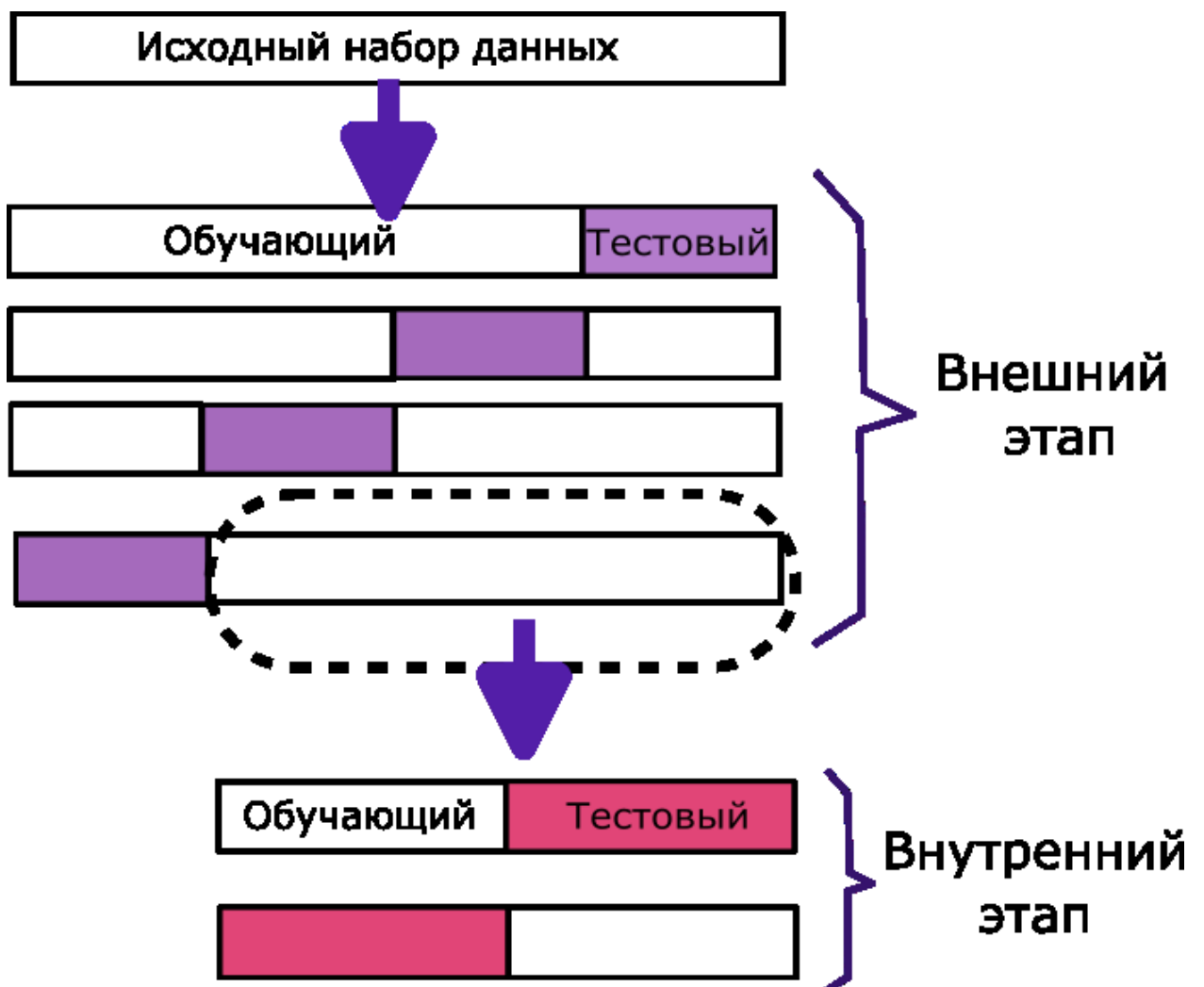


Рисунок 24. Схема вложенной кросс-валидации

В качестве дополнительного этапа проверки модели было проведено перестановочное тестирование. В перестановочных тестах из значений зависимой

переменной создаётся некоторое количество новых наборов значений зависимой переменной. Для этого значения исходной выборки меняются местами, перемешиваются, и для каждого полученного набора рассчитываются показатели эффективности модели, этот процесс проиллюстрирован на Рисунке 25. Таким образом, предполагается, что все модели, кроме той, где данные не перемешаны, не обладают предсказательной силой. Значит, сравнивая показатели реальной модели с этим распределением, можно оценить значимость предсказательной способности модели и убедиться, что она не является результатом случайной корреляции.

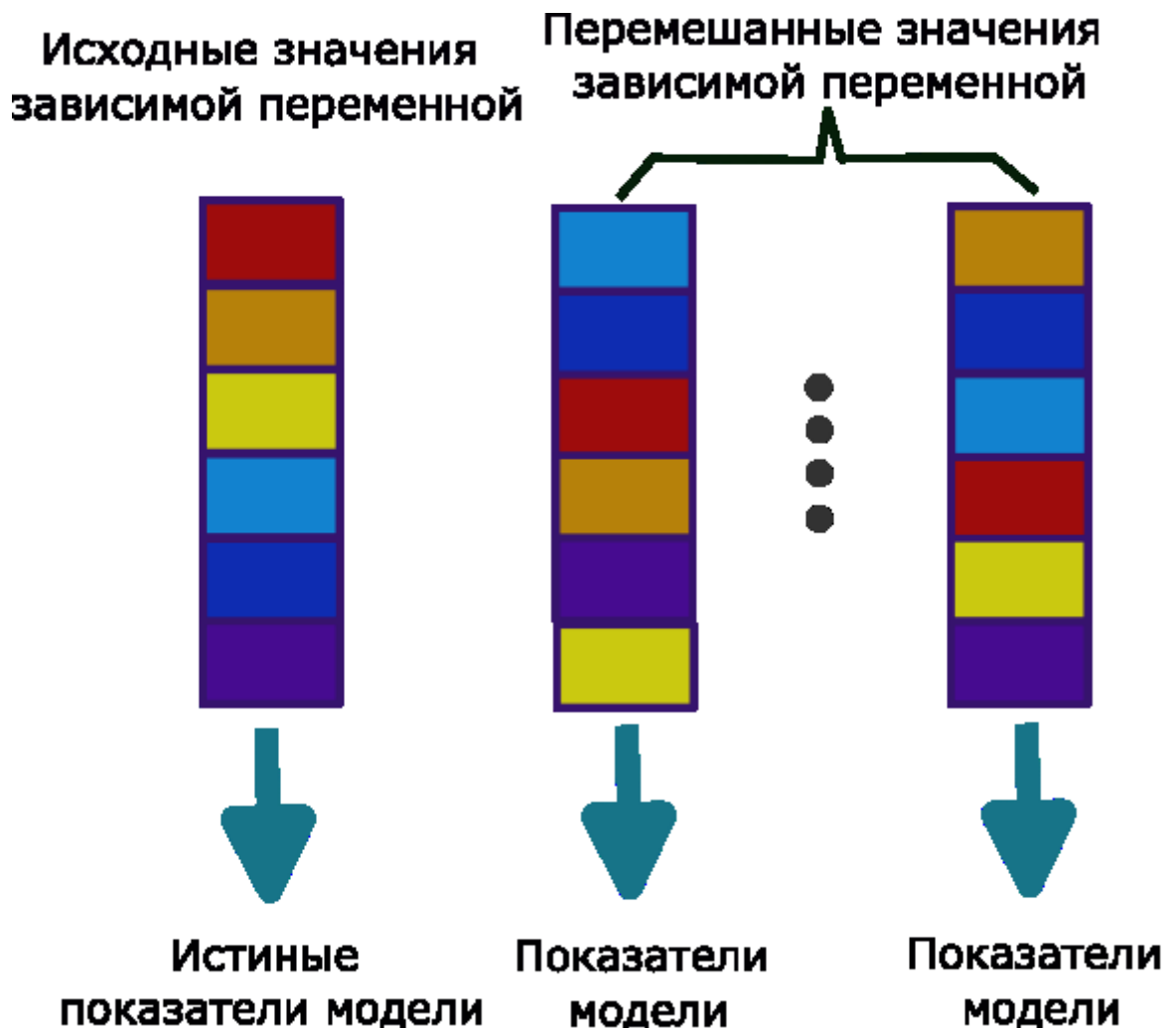


Рисунок 25. Схема перестановочного тестирования

Регрессионное моделирование проводилось в среде программирования R (R версии 4.1.3) с использованием пакета `caret` для PLS и пакета `nestedcv` для вложенной кросс-валидации.

4.2. Обсуждение результатов

4.2.1. Плотность

В качестве первого шага в анализе данных мы построили график зависимости плотности и показателя преломления от содержания воды, как показано на Рисунках 26-27. Каждому составу ГЭР соответствует одна точка на графике, при этом содержание воды в составе ГЭР отображено на оси X. Эти графики выявили четко выраженную линейную зависимость между этими переменными и процентным содержанием воды.

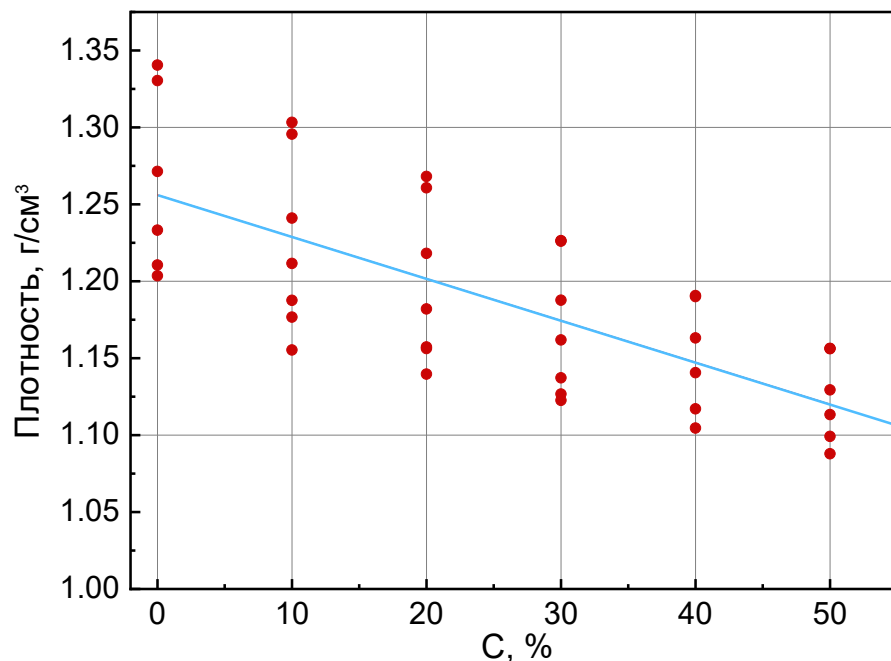


Рисунок 26. Точечная диаграмма, показывающая зависимость между плотностью и содержанием воды. Голубая линия отображает зависимость, полученную с помощью простой линейной регрессии

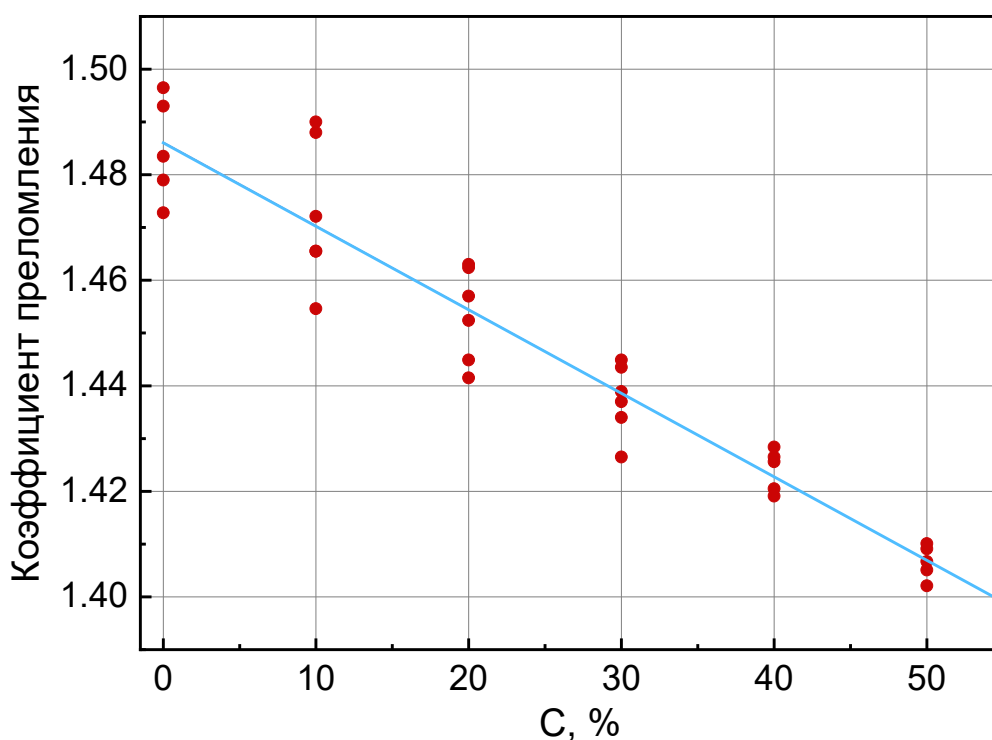


Рисунок 27. Точечная диаграмма, показывающая зависимость между коэффициентом преломления и содержанием воды. Голубая линия отображает зависимость, полученную с помощью простой линейной регрессии

Для более глубокого понимания, от каких параметров зависят значения переменных и как эти зависимости можно объяснить, к этим данным было применено PLS-моделирование, график введено-найдено для результатов моделирования представлен на Рисунке 28. Эта корреляция может быть более подробно изучена с помощью коэффициентов регрессии PLS. Абсолютное значение каждого коэффициента пропорционально вкладу соответствующей переменной в итоговую модель, при этом знак коэффициента указывает на его отрицательный или положительный вклад. Соответствующий график коэффициентов регрессии для PLS-модели, описывающей ГЭР с одинаковым содержанием воды, приведен на Рисунке 29. Анализ этого графика позволяет предположить, что основное влияние на плотность оказывает водородная связь, так как молекулярные фрагменты, содержащие C-OH и C=O фрагменты,

демонстрируют положительную корреляцию из-за их склонности к водородной связи. Увеличение количества водородных связей приводит к более плотному расположению молекул, что ведет к увеличению общей плотности [84]. Эти взаимодействия также отражаются на энергии гидратации: большее количество связей соответствует меньшей энергии гидратации и, следовательно, большей плотности. Кроме того, усиление водородных связей в целом снижает липофильность кислот. Это связано с полярными свойствами водородных связей, которые повышают растворимость в воде, в результате чего кислоты становятся менее растворимыми в неполярных растворителях, что отражается в отрицательной корреляции с $\log p$ (\log коэффициента распределения вода-октанол).

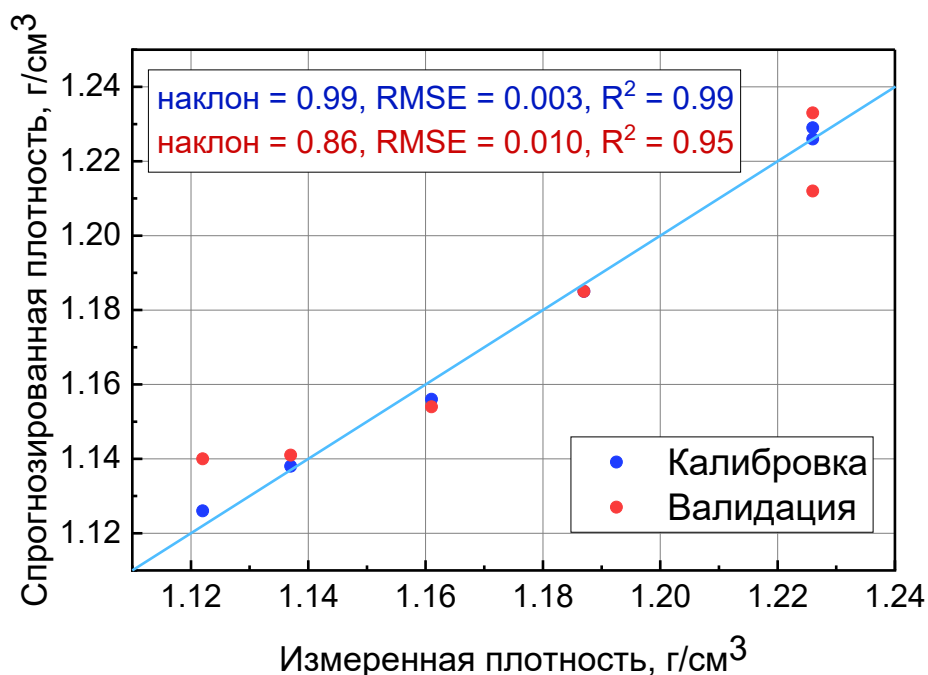


Рисунок 28. График введено-найдено для модели плотности PLS на основе образцов с содержанием воды = 30% (2 латентные переменные, полная кросс-валидация)

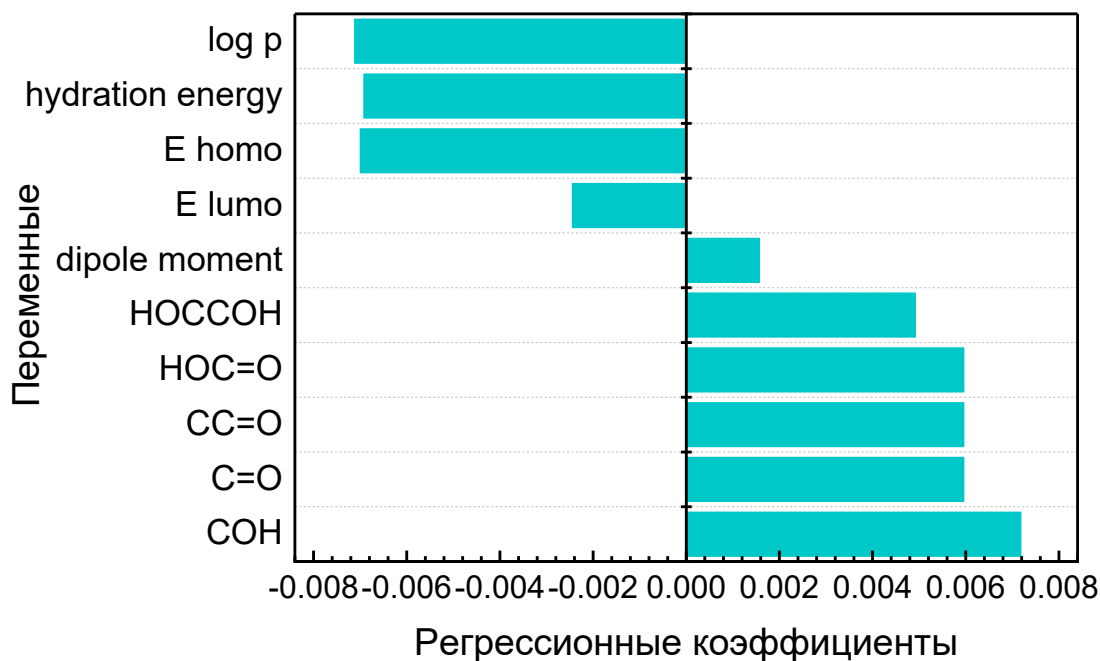


Рисунок 29. Регрессионные коэффициенты переменных с наибольшими вкладами в модель плотности PLS

4.2.2. Проводимость

По той же методике было проведено PLS-моделирование для прогнозирования проводимости ГЭР - было исследовано подмножество образцов с постоянным содержанием воды 30 мас.%, и в модель были включены только переменные с высокими абсолютными значениями коэффициентов регрессии. График зависимости измеренных и предсказанных значений для соответствующей PLS-модели приведен на Рисунке 30. Модель продемонстрировала достаточно высокую предсказательную способность.

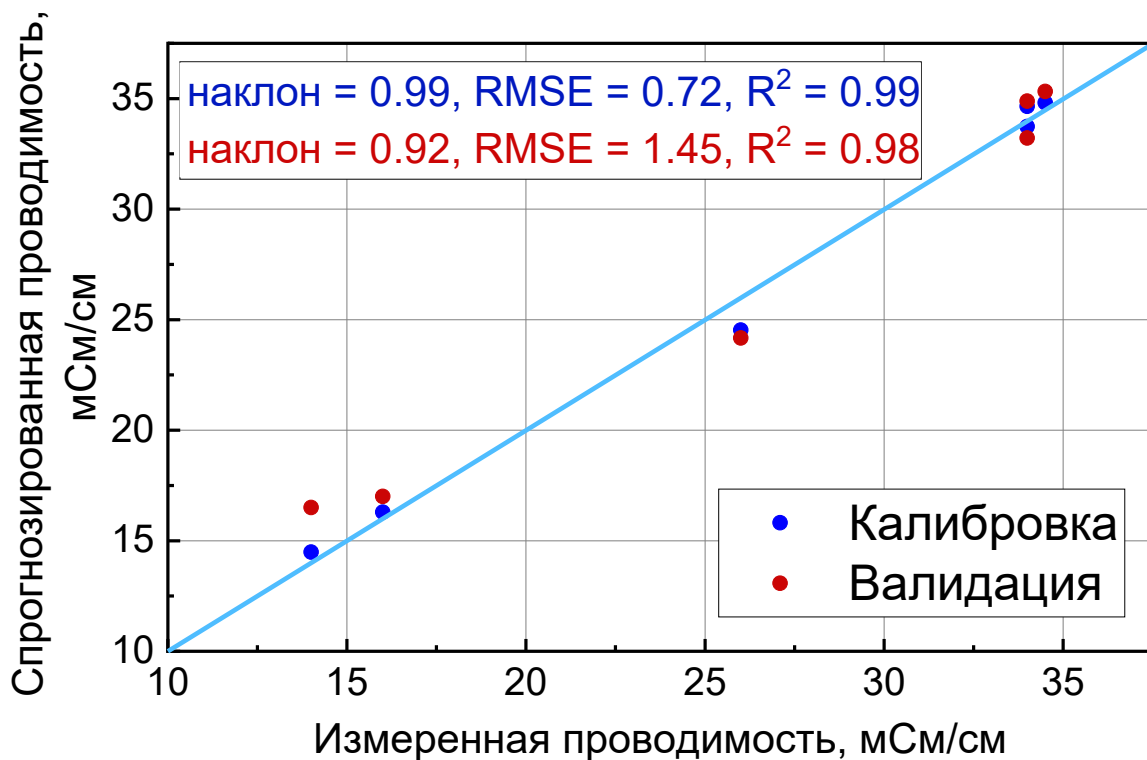


Рисунок 30. График введено-найдено для модели электропроводности на основе 6 образцов с содержанием воды = 30% (2 латентные переменные)

Анализ коэффициентов регрессии (Рисунок 31) показывает, что на проводимость существенно влияют несколько факторов: положительно – $\log P$ (липофильность), общая и свободная энергия, отрицательно - полярные СОН-группы и энтропия. Увеличение $\log P$ повышает проводимость, вероятно, из-за влияния на подвижность ионов [85].

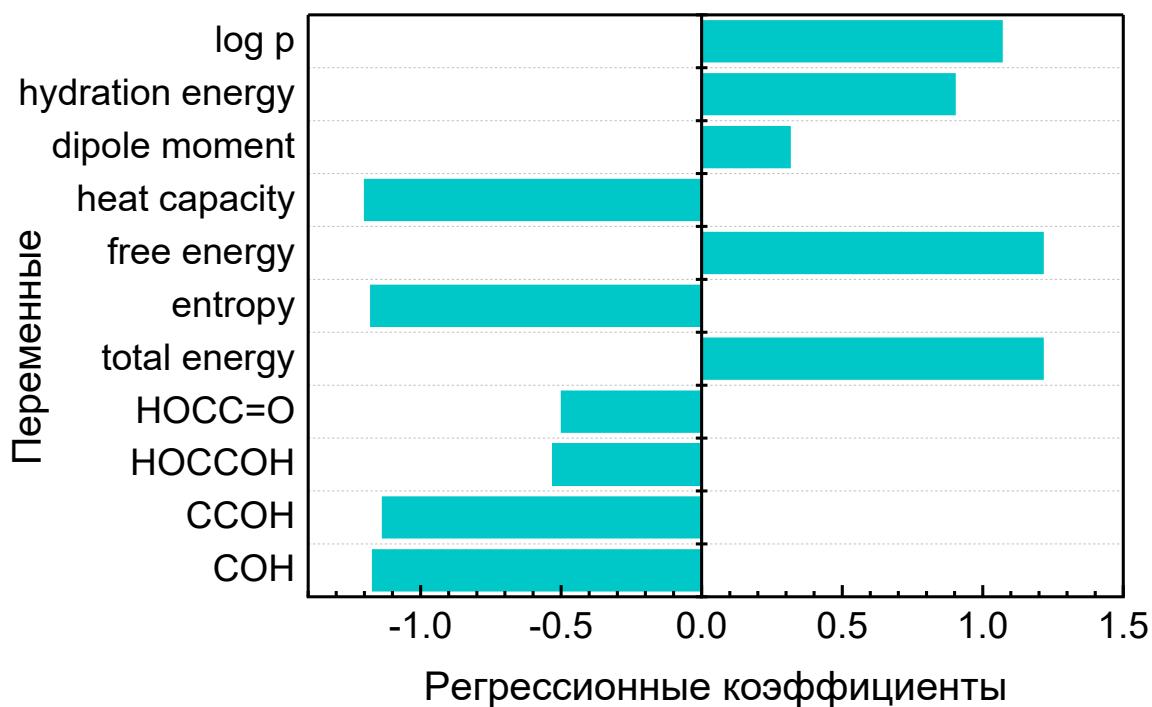


Рисунок 31. Регрессионные коэффициенты для PLS-модели проводимости на основе образцов с содержанием воды = 30%

4.2.3. Вязкость

Первые опыты по прогнозированию вязкости ГЭР с помощью PLS-моделирования показали, что модели имеют явно нелинейную структуру из-за очень широкого диапазона изменений экспериментально определенных значений вязкости. Чтобы обойти эту проблему, мы провели регрессионное моделирование относительно логарифма значений вязкости. На Рисунке 32 показан соответствующий график зависимости измеренных значений от спрогнозированных после исключения дескрипторов с незначительными коэффициентами регрессии.

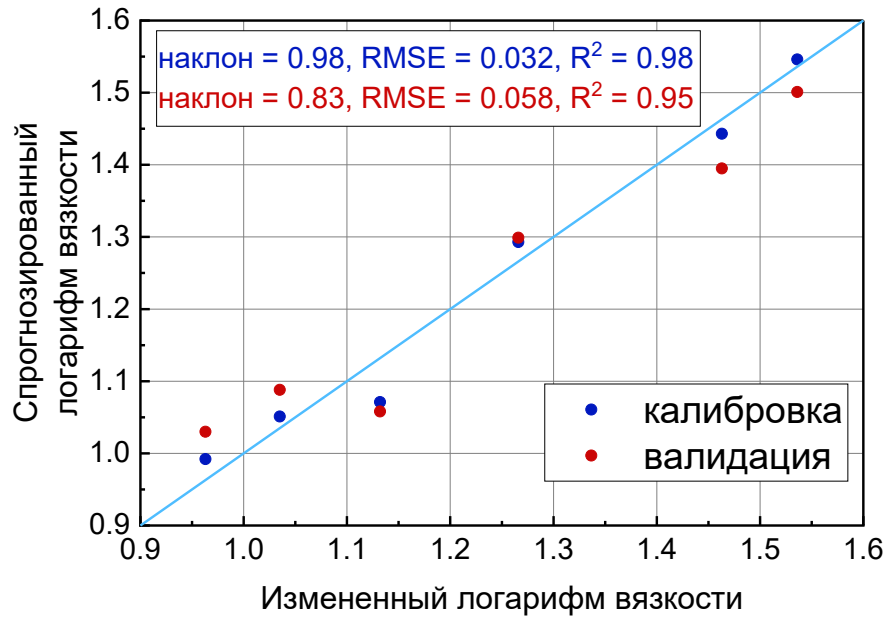


Рисунок 32. График введено-найдено для модели логарифма вязкости на основе 6 образцов с содержанием воды = 30% (1 латентная переменная)

Коэффициенты регрессии для остальных переменных представлены на Рисунке 33.

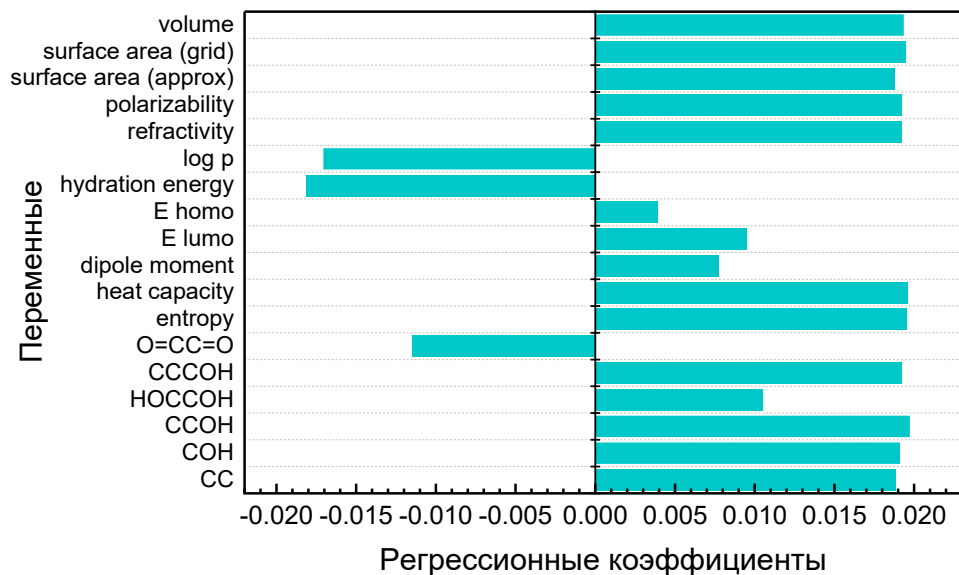


Рисунок 33. Регрессионные коэффициенты для модели вязкости PLS на основе образцов с содержанием воды = 30% для оптимизированного набора переменных

Более половины представленных переменных оказывают значительное влияние на логарифм вязкости. В частности, $\log R$ и энергия гидратации оказывают отрицательное влияние на логарифм вязкости, то есть чем выше $\log R$ или энергия гидратации органической кислоты, тем ниже её вязкость. Напротив, десять переменных, включая SS , COH , $SSOH$, $SSCOH$, энтальпию, теплоемкость, поляризуемость, преломляемость, площадь поверхности и объем молекул кислоты, оказывают положительное влияние. Наблюдаемая корреляция между липофильностью и вязкостью, вероятно, может быть объяснена взаимодействием компонентов ГЭР с водой; более высокая липофильность предполагает меньшее взаимодействие с водой, что приводит к снижению вязкости. Влияние площади поверхности и объема предполагает, что более крупные молекулы способствуют увеличению вязкости, что согласуется с общефизическими соображениями и другими экспериментами [86]. Авторы работы [86] показали, что вязкость в ряду растворов 1-алканолов коррелирует с числом атомов C в молекуле (что аналогично дескриптору SS в нашем исследовании, коррелирующему с общим числом атомов C). Наличие гидроксильных ($-OH$) групп в молекулярных фрагментах указывает на возможность образования водородных связей, которые, возможно, увеличивают сопротивление течению. Кроме того, поляризуемость, теплоёмкость и рефракция могут оказывать значительное влияние на вязкость в результате усиления межмолекулярных взаимодействий или сил, которые вносят вклад в общую вязкость.

4.2.4. Коэффициент преломления

Следующим изученным свойством был показатель преломления ГЭР. В соответствии с аналогичными рассуждениями, приведенными в разделе 4.2.1 с анализом плотности, моделирование было проведено для набора образцов с

одинаковым содержанием воды (30 мас.%). График введено-найдено для соответствующей PLS-модели представлен на Рисунке 34.

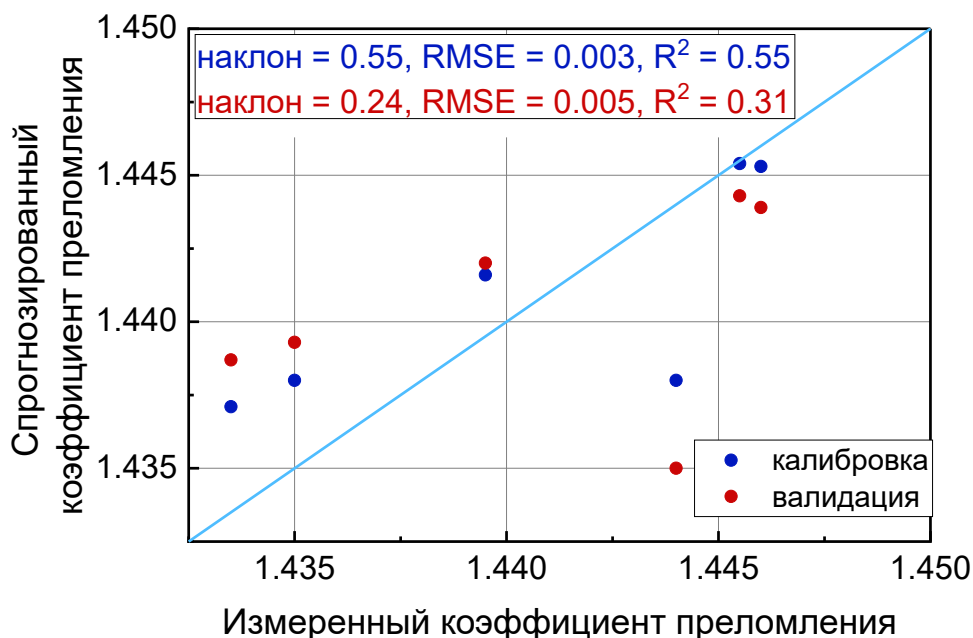


Рисунок 34. График введено-найдено для модели коэффициента преломления для образцов с содержанием воды = 30% (1 латентная переменная)

Видно, что корреляция между изученным набором дескрипторов и показателем преломления была довольно слабой (R^2 при калибровке 0,55, R^2 при валидации 0,31), при этом образец гликолевой кислоты был явным выбросом. Эти результаты означают, что выбранные дескрипторы недостаточны для предсказания оптических свойств ГЭР, полученного набора не хватает для объяснения зависимости структуры от коэффициента преломления. В этом случае для получения зависимости потребуются дальнейшее расширенное исследование.

Одним из перспективных подходов к решению этой проблемы представляется включение в модель дескрипторов, отражающих электродинамическое поведение молекул. Поскольку природа показателя преломления связана с взаимодействием электромагнитного излучения с

веществом, на него влияет способность молекулярной структуры поляризоваться в ответ на воздействие электромагнитного поля. Эта поляризация, в свою очередь, зависит от распределения и динамики электронных зарядов и молекулярного импульса. Примером таких дескрипторов могут служить топологические индексы заряда [87].

4.2.5. Статистическая проверка моделей

Конечной целью моделирования QSPR является оценка применимости этих моделей для прогнозирования свойств новых ГЭР. Учитывая ограниченное количество образцов, модели были тщательно проверены с помощью вложенной кросс-валидации и перестановочного тестирования, как указано в разделе 4.1.2. Такой подход позволяет предварительно оценить прогностическую способность регрессионных моделей в условиях ограниченного количества выборок. На следующем этапе прогностическая способность модели оценивалась для двух ГЭР на основе кислот, которые не использовались в обучении.

Графики соотношения измеренных и спрогнозированных значений для вложенной кросс-валидации PLS-моделей приведены на Рисунке 35. Показатели RMSE и R^2 свидетельствуют о возможности прогнозирования всех свойств с достаточно высокой точностью, кроме показателя преломления. Например, прогнозирование плотности в диапазоне 1,12 - 1,23 г/см³ достигается с RMSE 0,012. В случае показателя преломления модель не продемонстрировала удовлетворительных результатов при вложенной кросс-валидации, что указывает

на необходимость использования дополнительных дескрипторов или альтернативных подходов к моделированию.

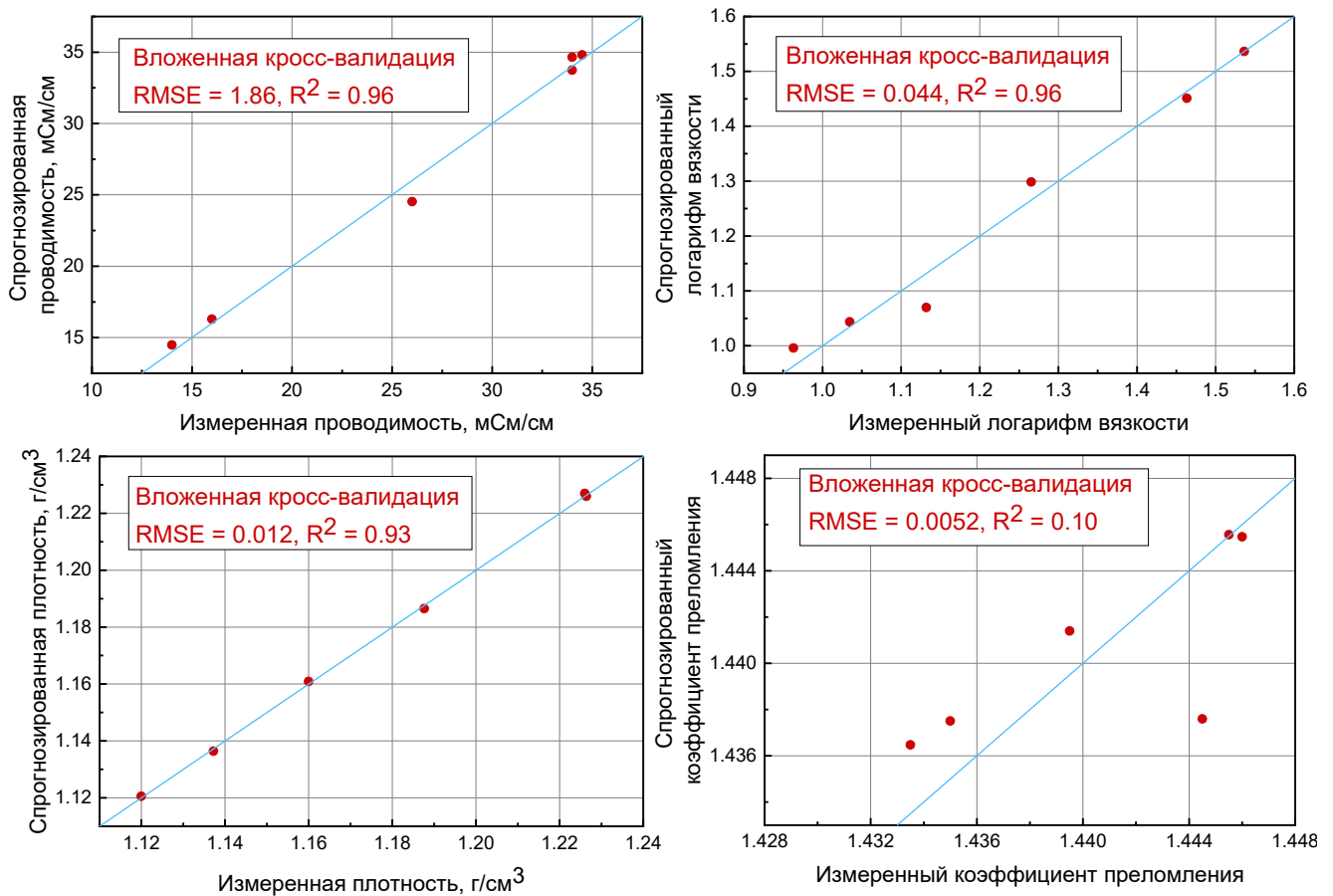


Рисунок 35. Графики введено-найдено для PLS-моделей с вложенной кросс-валидацией. Синяя линия показывает гипотетическую идеальную зависимость

Аналогичные результаты были получены при перестановочном тестировании (Рисунок 36). Из графиков следует, что набор использованных дескрипторов не позволяет построить надежную модель для прогнозирования показателя преломления, в то время как плотность, электропроводность и вязкость хорошо коррелируют с оптимизированными дескрипторами.

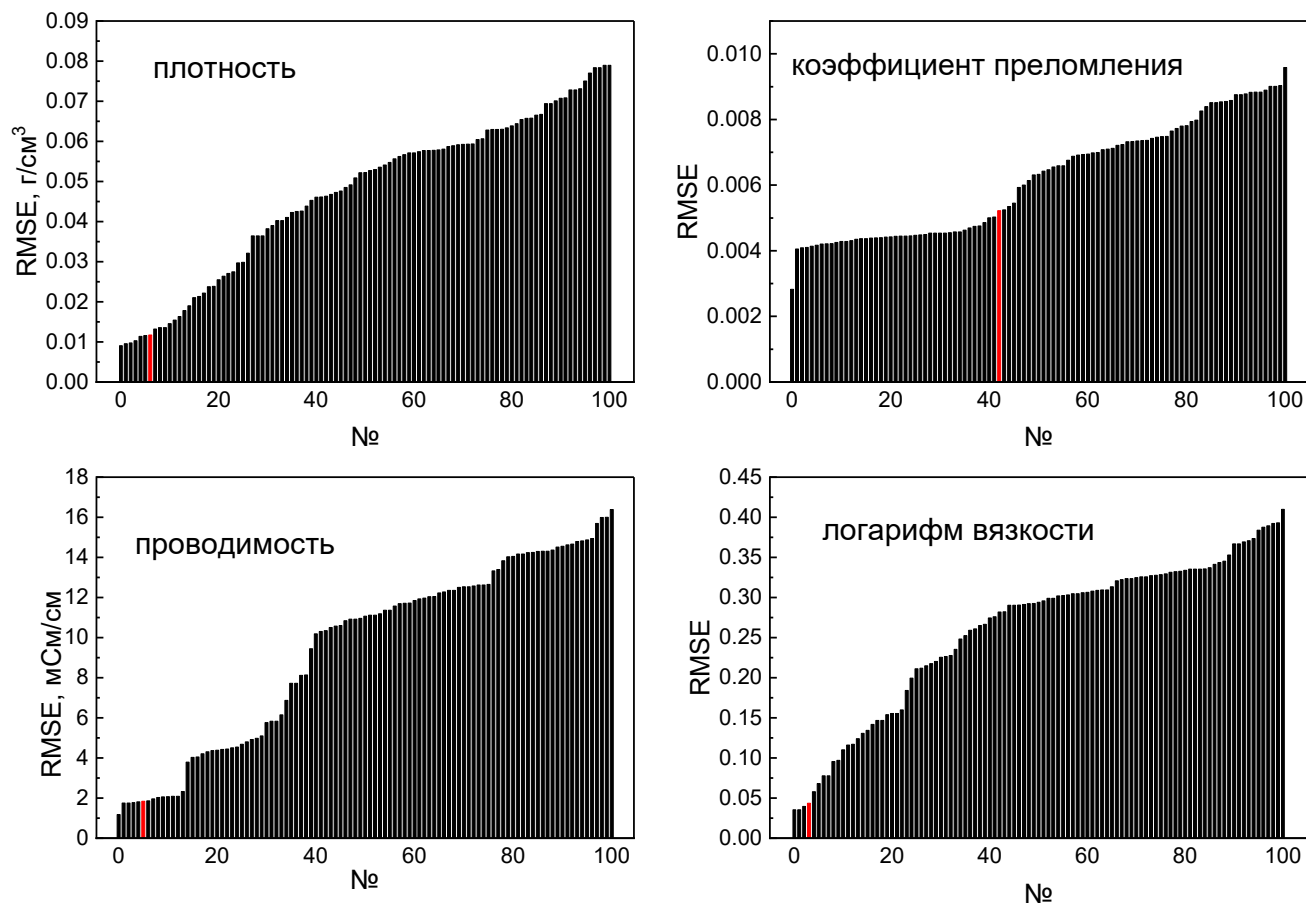


Рисунок 36. Результаты теста на перестановку для моделирования свойств ГЭР. RMSE исходной модели без проведённых перестановок обозначен красной линией

Небольшое количество образцов в наборе данных (6) и наличие близких по значениям пар привели к тому, что в некоторых перестановках получались модели с обучающими наборами, незначительно отличающимися от исходного. Это объясняет, почему в отдельных случаях показатели моделей при перестановках оказывались немного лучше, чем для исходной модели.

4.2.6. Оценка прогностической эффективности моделей

Для оценки прогностической способности разработанных QSPR-моделей были рассчитаны дескрипторы для малеиновой и уксусной кислот, не входивших в

обучающий набор. На основе этих дескрипторов были спрогнозированы плотность, вязкость, электропроводность и показатель преломления двух соответствующих ГЭР, содержащих 30 мас.% воды. Полученные предсказания были сопоставлены с экспериментальными данными, полученными для синтезированных ГЭР. Результаты сравнения представлены в Таблице 9.

Таблица 9. Свойства ГЭР на основе малеиновой и уксусной кислот с 30 мас.% воды, полученные в эксперименте и спрогнозированные с помощью моделей QSPR. Экспериментальные значения приведены в колонках «Эксп.», спрогнозированные – в колонках «Прогн.»

Химическая структура органической кислоты	Коэффициент преломления		Плотность, г/см ³		Вязкость, мПа·с		Проводимость, мСм/см	
	Прогн.	Эксп.	Прогн.	Эксп.	Прогн.	Эксп.	Прогн.	Эксп.
	1,438	1,444	1,210	1,145	17,4	13,8	29,0	28,5
	1,435	1,429	1,102	1,085	7,1	7,4	42,5	37,9

Как видно из полученных результатов, наблюдается удовлетворительное согласие между спрогнозированными и измеренными значениями новых ГЭР. Наибольшее расхождение наблюдается для вязкости в случае малеиновой кислоты - это, вероятно, связано с наличием двойной связи C=C в этой молекуле - особенность, которая не присутствовала в молекулах обучающего набора.

Некоторое отклонение прогнозируемого значения от экспериментального также наблюдается для проводимости ГЭР с уксусной кислотой в составе. Это, вероятно, связано с особенностями структуры уксусной кислоты (например, с наименьшим размером молекулы среди входящих в обучающий и тестовый наборы данных), которые не полностью учтены в модели.

Несмотря на некоторые расхождения, результаты исследования подтверждают перспективность использования метода QSRP для моделирования различных физических свойств эвтектических растворителей.

Заключение к Главе 4

Было показано, что математическое моделирование с использованием подхода QSPR может быть использовано для прогнозирования свойств ГЭР на основе молекулярных дескрипторов, охарактеризованных подструктурными фрагментами органических кислот. Подструктурные фрагменты и результаты полуэмпирических квантовых расчетов достаточны для количественного предсказания значений плотности, электропроводности и вязкости ГЭР. Продемонстрированная возможность применения QSPR в данной области исследований указывает на актуальность дальнейших специальных исследований для расширения спектра моделируемых свойств и описываемых глубоких эвтектических растворителей. Такое моделирование может существенно облегчить разработку систем ГЭР для конкретных практических применений

ЗАКЛЮЧЕНИЕ

1. Для выборки из 40 лигандов, описанных в литературе по анионселективным электродам на основе подструктурных молекулярных фрагментов и проводимости пластификатора, созданы QSPR модели, которые позволяют прогнозировать потенциометрическую селективность полимерных пластифицированных мембранных сенсоров на основе таких лигандов к карбонат-аниону. Модель позволяет получать прогноз селективности для новых соединений с погрешностью в среднем не более $1,5 \log K^{\text{сел}}(\text{HCO}_3^- / \text{Cl}^-)$.

2. На основе подструктурных молекулярных фрагментов для 41 соединения, содержащего амидные и фосфорильные функциональные группы, разработаны QSPR модели, которые позволяют прогнозировать потенциометрическую чувствительность полимерных пластифицированных мембранных сенсоров на основе таких лигандов к катионам тяжёлых металлов (Cu^{2+} , Cd^{2+} и Pb^{2+}). Модель позволяет получать прогноз чувствительности для новых соединений с погрешностью не более 8 мВ/дек.

3. Для выборки глубоких эвтектических растворителей на основе холин хлорида с 6 различными органическими кислотами в качестве доноров водородной связи, были получены молекулярные дескрипторы в виде подструктурных молекулярных фрагментов и квантово-химических параметров молекул кислот. На основе этих данных разработаны QSPR модели, позволяющие для новых структурно схожих ГЭР прогнозировать плотность с погрешностью не более 0,065 г/см, проводимость с погрешностью не более 4,6 мСм/см, вязкость с погрешностью не более 3,6 мПа·с.

4. Такие подходы, как вложенная кросс-валидация и проверка перестановочными тестами, позволяют проводить проверку и обеспечивать надежность QSPR-моделей, построенных на основе ограниченных выборок обучающих данных.

5. Анализ регрессионных коэффициентов для PLS регрессии в QSPR-моделях позволяет выявить ключевые структурные дескрипторы, определяющие необходимые свойства новых материалов.

СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ

ВЭЖХ - высокоэффективная жидкостная хроматография

ГЭР - глубокие эвтектические растворители

ИСЭ - ионоселективные электроды

CPSA - Charged partial surface area

DFT - density functional theory

ECFP - Extended-Connectivity Fingerprints

J3D - 3D Индекс Балабана

LV - latent variable

LOO CV - leave-one-out cross-validation

MACCS - Molecular ACCess System

MLR - Multiple Linear Regression

PLS - Partial Least Squares

PM3 - Parametric Method 3

SMF - substructure molecular fragments

SMILES - Simplified Molecular-Input Line-Entry System

SVM - support vector machine

SVR - support vector regression

QSAR - Quantitative Structure-Activity Relationships

QSPR - Quantitative Structure-Property Relationships

W3D - 3D индекс Винера

СПИСОК ЛИТЕРАТУРЫ

1. Vladimirova N. et al. Prediction of Carbonate Selectivity of PVC-Plasticized Sensor Membranes with Newly Synthesized Ionophores through QSPR Modeling // Chemosensors. 2022. Vol. 10, № 2. P. 43.
2. Vladimirova N. et al. Predicting the Potentiometric Sensitivity of Membrane Sensors Based on Modified Diphenylphosphoryl Acetamide Ionophores with QSPR Modeling // Membranes (Basel). 2022. Vol. 12, № 10. P. 953.
3. Vladimirova N. et al. Predicting the properties of deep eutectic solvents based on choline chloride and carboxylic acids and their mixtures with water using QSPR approach // Colloids Surf A Physicochem Eng Asp. 2024. Vol. 692. P. 133961.
4. Brown A.C., Fraser T.R. On the Connection between Chemical Constitution and Physiological Action; with special reference to the Physiological Action of the Salts of the Ammonium Bases derived from Strychnia, Brucia, Thebaia, Codeia, Morphia, and Nicotia. // J Anat Physiol. 1868. Vol. 2, № 2. P. 224–242.
5. Hansch Corwin., Anderson S.M. The Structure-Activity Relationship in Barbiturates and Its Similarity to That in Other Narcotics // J Med Chem. 1967. Vol. 10, № 5. P. 745–753.
6. Hansch Corwin., Fujita Toshio. ρ - σ - π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure // J Am Chem Soc. 1964. Vol. 86, № 8. P. 1616–1626.
7. HANSCH C. et al. Correlation of Biological Activity of Phenoxyacetic Acids with Hammett Substituent Constants and Partition Coefficients // Nature. 1962. Vol. 194, № 4824. P. 178–180.
8. Silva de Jesus Passaes A.C. et al. Quinoxalines against *Leishmania amazonensis*: SAR study, proposition of a new derivative, QSAR prediction, synthesis, and biological evaluation // Sci Rep. 2023. Vol. 13, № 1. P. 18136.

9. Esfahani S.N. et al. Synthesis of some novel coumarin isoxazol sulfonamide hybrid compounds, 3D-QSAR studies, and antibacterial evaluation // *Sci Rep.* 2021. Vol. 11, № 1. P. 20088.
10. Qin D. et al. 5D-QSAR studies of 1H-pyrazole derivatives as EGFR inhibitors // *J Mol Model.* 2022. Vol. 28, № 12. P. 379.
11. Acharya K. et al. A quantitative structure-biodegradation relationship (QSBR) approach to predict biodegradation rates of aromatic chemicals // *Water Res.* 2019. Vol. 157. P. 181–190.
12. Najafi M. et al. QSPR model for estimation of photodegradation average rate of the porphyrin-TiO₂ complexes and prediction of their biodegradation activity and toxicity: Engineering of two annihilators for water/waste contaminants // *J Mol Struct.* 2022. Vol. 1249. P. 131463.
13. Ai H. et al. QSAR modelling study of the bioconcentration factor and toxicity of organic compounds to aquatic organisms using machine learning and ensemble methods // *Ecotoxicol Environ Saf.* 2019. Vol. 179. P. 71–78.
14. Wang C. et al. Assessment of bromide-based ionic liquid toxicity toward aquatic organisms and QSAR analysis // *Ecotoxicol Environ Saf.* 2015. Vol. 115. P. 112–118.
15. Cai J. et al. Mechanistic studies of congener-specific adsorption and bioaccumulation of polycyclic aromatic hydrocarbons and phthalates in soil by novel QSARs // *Environ Res.* 2019. Vol. 179. P. 108838.
16. Arnot J.A., Gobas F.A.P.C. A Generic QSAR for Assessing the Bioaccumulation Potential of Organic Chemicals in Aquatic Food Webs // *QSAR Comb Sci.* 2003. Vol. 22, № 3. P. 337–345.
17. Puzyn T. et al. QSPR-based estimation of the atmospheric persistence for chloronaphthalene congeners // *Atmos Environ.* 2008. Vol. 42, № 27. P. 6627–6636.

18. Villaverde J.J. et al. QSAR/QSPR models based on quantum chemistry for risk assessment of pesticides according to current European legislation // SAR QSAR Environ Res. 2020. Vol. 31, № 1. P. 49–72.
19. Yang L. et al. QSAR modeling the toxicity of pesticides against *Americamysis bahia* // Chemosphere. 2020. Vol. 258. P. 127217.
20. Rodrigues N.E. et al. QSAR-Guided Proposition of N-(4-methanesulfonyl)Benzoyl-N'-(Pyrimidin-2-yl)Thioureas as Effective and Safer Herbicides // Bull Environ Contam Toxicol. 2022. Vol. 108, № 6. P. 1019–1025.
21. Tukur S., Shallangwa G.A., Ibrahim A. Theoretical QSAR modelling and molecular docking studies of some 4-hydroxyphenylpyruvate dioxygenase (HPPD) enzyme inhibitors potentially used as herbicides // Heliyon. 2019. Vol. 5, № 11. P. e02859.
22. Isyaku Y., Uzairu A., Uba S. QSAR STUDY OF 2-SUBSTITUTED PHENYL-2-OXO-, 2-HYDROXYL- AND 2-ACYLLOXYETHYLSULFONAMIDES AS FUNGICIDES // The Journal of Engineering and Exact Sciences. 2019. Vol. 5, № 3. P. 0283–0290.
23. Kaushik P., Shakil N.A., Rana V.S. Synthesis, Biological Evaluation, and QSAR Studies of 3-Iodochromone Derivatives as Potential Fungicides // Front Chem. 2021. Vol. 9. P. 636882.
24. Knizhnik A.A. et al. New Polymers In Silico Generation and Properties Prediction // Nanomanufacturing. 2023. Vol. 4, № 1. P. 1–26.
25. Cravero F. et al. Computational modelling of mechanical properties for new polymeric materials with high molecular weight // Chemometrics and Intelligent Laboratory Systems. 2019. Vol. 193. P. 103851.
26. Buglak A.A., Zherdev A. V., Dzantiev B.B. Nano-(Q)SAR for Cytotoxicity Prediction of Engineered Nanomaterials // Molecules. 2019. Vol. 24, № 24. P. 4537.
27. Toropov A.A., Kjeldsen F., Toropova A.P. Use of quasi-SMILES to build models based on quantitative results from experiments with nanomaterials // Chemosphere. 2022. Vol. 303. P. 135086.

28. Kar S. et al. Power Conversion Efficiency of Arylamine Organic Dyes for Dye-Sensitized Solar Cells (DSSCs) Explicit to Cobalt Electrolyte: Understanding the Structural Attributes Using a Direct QSPR Approach // *Computation*. 2016. Vol. 5, № 4. P. 2.
29. Ratanasak M., Hasegawa J., Parasuk V. Design and prediction of high potent *ansa*-zirconocene catalyst for olefin polymerizations: combined DFT calculations and QSPR approach // *New Journal of Chemistry*. 2021. Vol. 45, № 18. P. 8248–8257.
30. Nkulikiyinka P. et al. Prediction of Combined Sorbent and Catalyst Materials for SE-SMR, Using QSPR and Multitask Learning // *Ind Eng Chem Res*. 2022. Vol. 61, № 26. P. 9218–9233.
31. Ojha P.K., Roy K. Development of a robust and validated 2D-QSPR model for sweetness potency of diverse functional organic molecules // *Food and Chemical Toxicology*. 2018. Vol. 112. P. 551–562.
32. Chen K. et al. Multiple quantitative structure–pungency correlations of capsaicinoids // *Food Chem*. 2019. Vol. 283. P. 611–620.
33. Belhassan A. et al. QSPR Study of the Retention/Release Property of Odorant Molecules in Water, Dairy and Pectin gels // *Mater Today Proc*. 2019. Vol. 13. P. 621–629.
34. Buckley H.L. et al. Design and Testing of Safer, More Effective Preservatives for Consumer Products // *ACS Sustain Chem Eng*. 2017. Vol. 5, № 5. P. 4320–4331.
35. Rezaei S. et al. Exploring 3D-QSPR models of human skin permeability for a diverse dataset of chemical compounds // *Journal of Receptors and Signal Transduction*. 2019. Vol. 39, № 5–6. P. 442–450.
36. Oztan Akturk S., Tugcu G., Sipahi H. Development of a QSAR model to predict comedogenic potential of some cosmetic ingredients // *Computational Toxicology*. 2022. Vol. 21. P. 100207.
37. Kim J.Y., Kim K.-B., Lee B.-M. Validation of Quantitative Structure-Activity Relationship (QSAR) and Quantitative Structure-Property Relationship (QSPR)

- approaches as alternatives to skin sensitization risk assessment // *J Toxicol Environ Health A*. 2021. Vol. 84, № 23. P. 945–959.
38. Liu R. et al. QSPR models for sublimation enthalpy of energetic compounds // *Chemical Engineering Journal*. 2023. Vol. 474. P. 145725.
 39. Wang B. et al. Prediction of Minimum Ignition Energy from Molecular Structure Using Quantitative Structure–Property Relationship (QSPR) Models // *Ind Eng Chem Res*. 2017. Vol. 56, № 1. P. 47–51.
 40. Zhang S. et al. Evaluating the properties of ionic liquid at variable temperatures and pressures by quantitative structure–property relationship (QSPR) // *Chem Eng Sci*. 2021. Vol. 231. P. 116326.
 41. Yan F. et al. QSPR models for the properties of ionic liquids at variable temperatures based on norm descriptors // *Chem Eng Sci*. 2020. Vol. 217. P. 115540.
 42. Roubehie Fissa M. et al. Development of QSPR-ANN models for the estimation of critical properties of pure hydrocarbons // *J Mol Graph Model*. 2023. Vol. 121. P. 108450.
 43. Li R. et al. Machine learning-quantitative structure property relationship (ML-QSPR) method for fuel physicochemical properties prediction of multiple fuel types // *Fuel*. 2021. Vol. 304. P. 121437.
 44. Jiao L. et al. QSPR Studies on the Octane Number of Toluene Primary Reference Fuel Based on the Electrotopological State Index // *ACS Omega*. 2020. Vol. 5, № 8. P. 3878–3888.
 45. Randić M. Molecular bonding profiles // *J Math Chem*. 1996. Vol. 19, № 3. P. 375–392.
 46. Rogers D., Hahn M. Extended-Connectivity Fingerprints // *J Chem Inf Model*. 2010. Vol. 50, № 5. P. 742–754.
 47. Bahia M.S. et al. A comparison between 2D and 3D descriptors in QSAR modeling based on bio-active conformations // *Mol Inform*. 2023. Vol. 42, № 4. P. e2200186.

48. Verloop A., Hoogenstraaten W., Tipker J. Development and Application of New Steric Substituent Parameters in Drug Design // Drug Design. Elsevier, 1976. P. 165–207.
49. Bogdanov B., Nikolić S., Trinajstić N. On the three-dimensional wiener number // J Math Chem. 1989. Vol. 3, № 3. P. 299–309.
50. Mihalic Z., Nikolic S., Trinajstic N. Comparative study of molecular descriptors derived from the distance matrix // J Chem Inf Comput Sci. 1992. Vol. 32, № 1. P. 28–37.
51. Hollas B. THE AUTOCORRELATION OF A TOPOLOGICAL STRUCTURE: A NEW MOLECULAR DESCRIPTOR // J Math Chem. 2003. Vol. 33, № 2. P. 91–101.
52. Gramatica P. WHIM Descriptors of Shape // QSAR Comb Sci. 2006. Vol. 25, № 4. P. 327–332.
53. Mulliken R.S. Electronic Population Analysis on LCAO–MO Molecular Wave Functions. I // J Chem Phys. 1955. Vol. 23, № 10. P. 1833–1840.
54. Stanton D.T. et al. Charged partial surface area (CPSA) descriptors QSAR applications // SAR QSAR Environ Res. 2002. Vol. 13, № 2. P. 341–351.
55. Wold S., Sjöström M., Eriksson L. PLS-regression: a basic tool of chemometrics // Chemometrics and Intelligent Laboratory Systems. 2001. Vol. 58, № 2. P. 109–130.
56. Athanasios Kondylis. Smooth PLS Regression for Spectral Data // REVSTAT-Statistical Journal. 2022. Vol. 20, № 4. P. 463–479.
57. Gopala Krishna J., Roy K. QSPR modeling of absorption maxima of dyes used in dye sensitized solar cells (DSSCs) // Spectrochim Acta A Mol Biomol Spectrosc. 2022. Vol. 265. P. 120387.
58. Eichenlaub J., Rakowska P.W., Kloskowski A. User-assisted methodology targeted for building structure interpretable QSPR models for boosting CO2 capture with ionic liquids // J Mol Liq. 2022. Vol. 350. P. 118511.
59. Xu H., Caramanis C., Mannor S. Robustness and Regularization of Support Vector Machines. 2008.

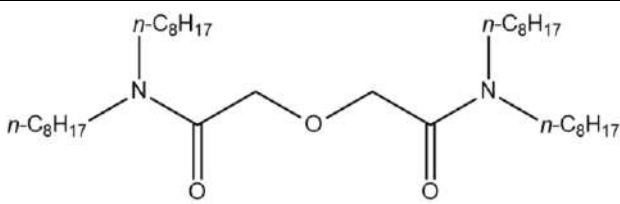
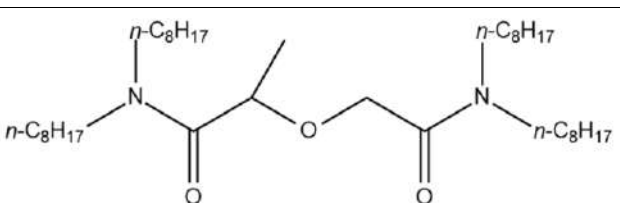
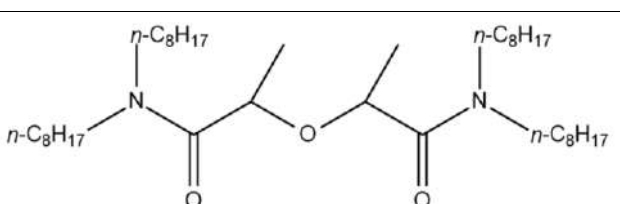
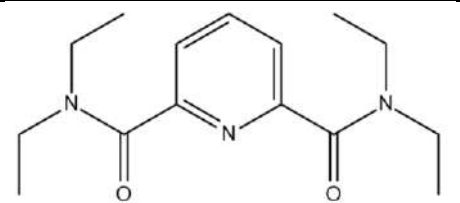
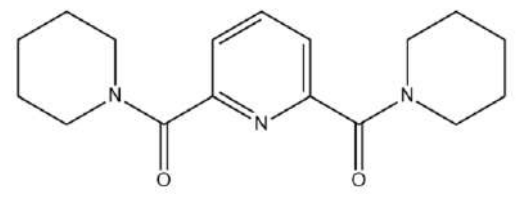
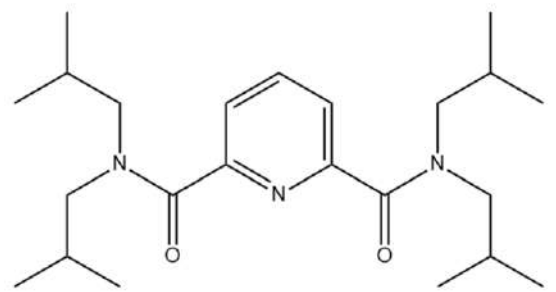
60. Soloviev V. et al. QSPR modeling of potentiometric sensitivity towards heavy metal ions for polymeric membrane sensors // *Sens Actuators B Chem.* 2019. Vol. 301. P. 126941.
61. Martynko E. et al. QSPR Modeling of Potentiometric Mg^{2+}/Ca^{2+} Selectivity for PVC-plasticized Sensor Membranes // *Electroanalysis.* 2020. Vol. 32, № 4. P. 792–798.
62. Eslam Pourbasheer. High Accurate Prediction of Carbonate Selectivity of PVC-Plasticized Membranes Sensors by Genetic Algorithm-Support Vector Machine // *Analytical and Bioanalytical Electrochemistry.* 2023. Vol. 15, № 2. P. 150–165.
63. Roya Kiani-anboui, Zeinab Mozafari. A Novel QSPR Approach in Modeling Selectivity Coefficients of the Lanthanum-Selective Electrode // *Analytical and Bioanalytical Electrochemistry.* 2023. Vol. 15, № 12. P. 1031–1045.
64. Diem-Tran P.T. et al. Stability Constant and Potentiometric Sensitivity of Heavy Metal–Organic Fluorescent Compound Complexes: QSPR Models for Prediction and Design of Novel Coumarin-like Ligands // *Toxics.* 2023. Vol. 11, № 7. P. 595.
65. Abbott A.P. et al. Preparation of novel, moisture-stable, Lewis-acidic ionic liquids containing quaternary ammonium salts with functional side chains // *Chemical Communications.* 2001. № 19. P. 2010–2011.
66. Lemaoui T. et al. Quantitative structure properties relationship for deep eutectic solvents using σ -profile as molecular descriptors // *J Mol Liq.* 2020. Vol. 309. P. 113165.
67. Boubliia A. et al. Molecular-based artificial neural network for predicting the electrical conductivity of deep eutectic solvents // *J Mol Liq.* 2022. Vol. 366. P. 120225.
68. Salahshoori I., Baghban A., Yazdanbakhsh A. Novel hybrid QSPR-GPR approach for modeling of carbon dioxide capture using deep eutectic solvents // *RSC Adv.* 2023. Vol. 13, № 43. P. 30071–30085.

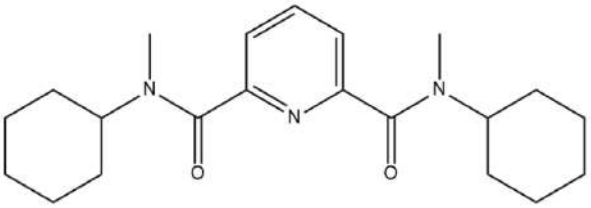
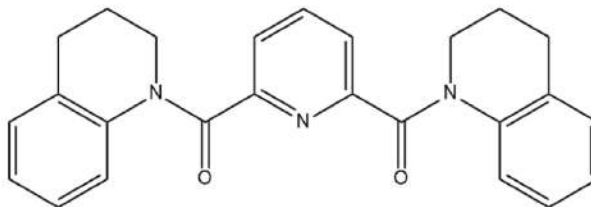
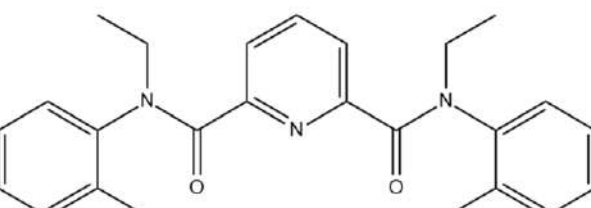
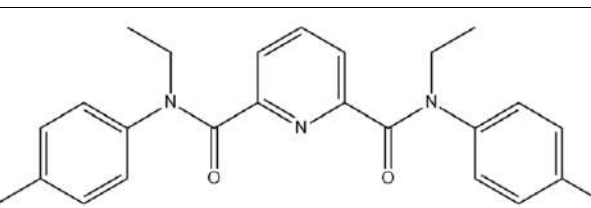
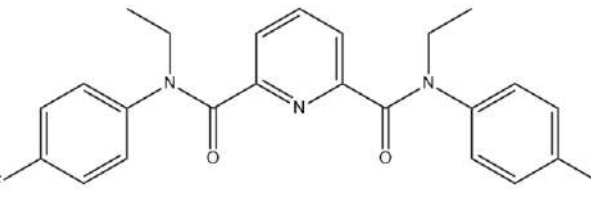
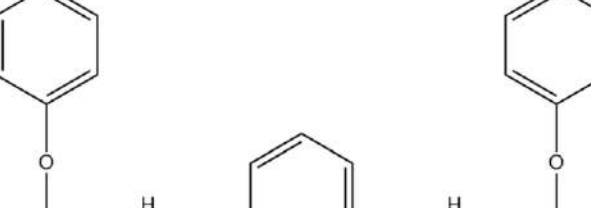
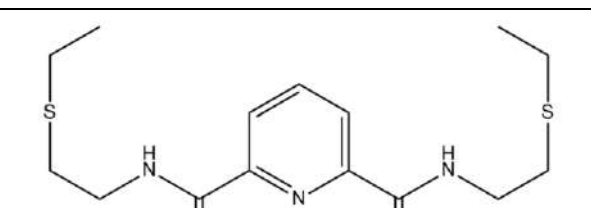
69. Hu J. et al. COSMO-SAC and QSPR combined models: A flexible and reliable strategy for screening the extraction efficiency of deep eutectic solvents // *Sep Purif Technol.* 2023. Vol. 315. P. 123699.
70. Fedorova E.S. et al. Deep learning for retention time prediction in reversed-phase liquid chromatography // *J Chromatogr A.* 2022. Vol. 1664. P. 462792.
71. Obradović D. et al. A comparative study of the predictive performance of different descriptor calculation tools: Molecular-based elution order modeling and interpretation of retention mechanism for isomeric compounds from METLIN database // *J Chromatogr A.* 2024. Vol. 1719. P. 464731.
72. Turanov A.N., Karandashev V.K., Yarkevich A.N. Extraction of Rare Earth Elements(III) from Perchlorate Solutions with Modified Diphenylphosphorylacetamides // *Russian Journal of Inorganic Chemistry.* 2021. Vol. 66, № 4. P. 572–577.
73. Sokalski T. et al. Observations on the behaviour of some trifluoroacetophenone derivatives as neutral carriers for carbonate ion-selective electrodes // *Analyst.* 1996. Vol. 121, № 2. P. 133–138.
74. Behringer C. et al. Anion selectivities of trifluoroacetophenone derivatives as neutral ionophores in solvent-polymeric membranes // *Anal Chim Acta.* 1990. Vol. 233. P. 41–47.
75. Maj-Zurawska M. Carbonate ion selective electrodes with trifluoroacetophenone derivatives in potentiometric clinical analyser // *Talanta.* 1997. Vol. 44, № 9. P. 1641–1647.
76. Hong Y.K. et al. Effect of varying quaternary ammonium salt concentration on the potentiometric properties of some trifluoroacetophenone derivative-based solvent-polymeric membranes // *Electroanalysis.* 1997. Vol. 9, № 11. P. 865–868.
77. Shim J. Ion-selective electrodes based on molecular tweezer-type neutral carriers // *Talanta.* 2004. Vol. 63, № 1. P. 61–71.

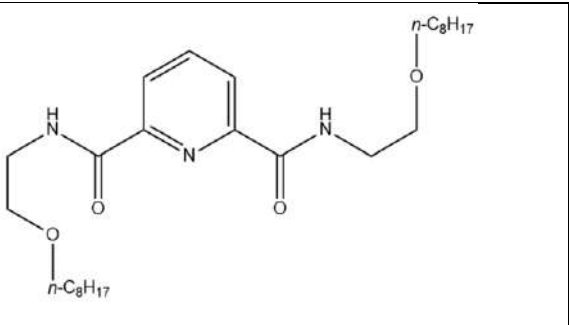
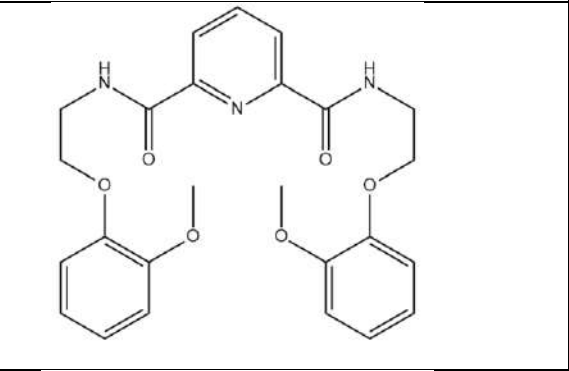
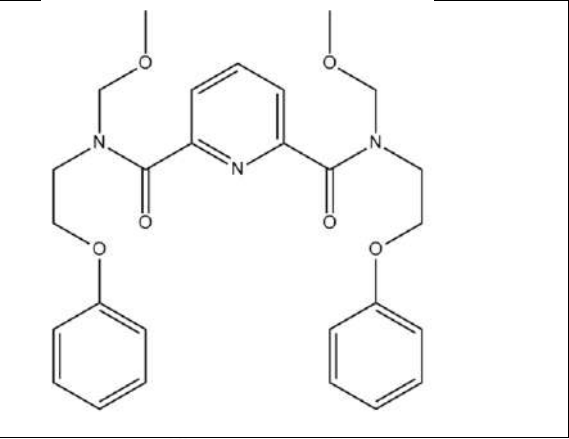
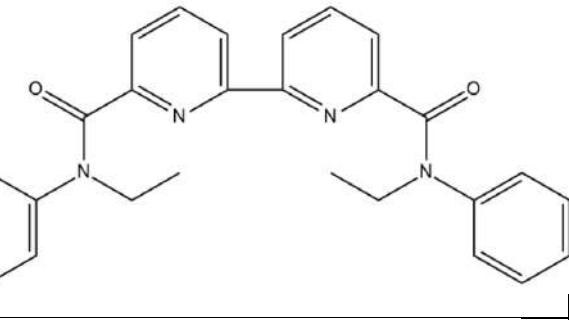
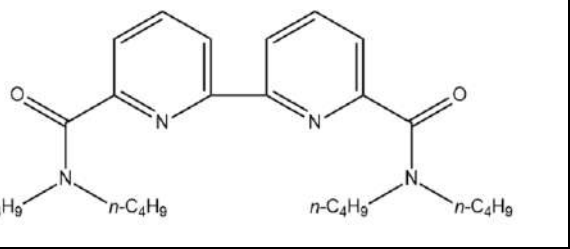
78. KAMATA S., NOMURA S., OHASHI K. Anion-selective membrane electrode based on bis(diphenylphosphino)alkane-copper(II) complexes. // BUNSEKI KAGAKU. 1990. Vol. 39, № 11. P. 677–681.
79. Rothmaier M. et al. Response mechanism of anion-selective electrodes based on mercury organic compounds as ionophores // Anal Chim Acta. 1996. Vol. 327, № 1. P. 17–28.
80. Umezawa Y. et al. Potentiometric Selectivity Coefficients of Ion-Selective Electrodes. Part I. Inorganic Cations (Technical Report) // Pure and Applied Chemistry. 2000. Vol. 72, № 10. P. 1851–2082.
81. Solov'ev Vitaly, Varnek Alexandre. Qspr Models on Fragment Descriptors [Electronic resource] // <http://vpsolovev.ru/wp-content/uploads/sites/9/2017/05/isida-qspr-help-2017.pdf>.
82. Meyerhoff M.E. et al. Role of trifluoroacetophenone solvents and quaternary ammonium salts in carbonate-selective liquid membrane electrodes // Anal Chem. 1987. Vol. 59, № 1. P. 144–150.
83. Kirsanov D. et al. Potentiometric Sensor Array for Analysis of Complex Rare Earth Mixtures // Electroanalysis. 2012. Vol. 24, № 1. P. 121–130.
84. Zhang J. et al. One Step Closer to an Ideal Insensitive Energetic Molecule: 3,5-Diamino-6-hydroxy-2-oxide-4-nitropyrimidone and its Derivatives // J Am Chem Soc. 2021. Vol. 143, № 32. P. 12665–12674.
85. Every H. et al. Ion diffusion in molten salt mixtures // Electrochim Acta. 2000. Vol. 45, № 8–9. P. 1279–1284.
86. Saleh M.A. et al. Density and Viscosity of 1-Alkanols // Phys Chem Liquids. 2004. Vol. 42, № 6. P. 615–623.
87. Galvez J. et al. Charge Indexes. New Topological Descriptors // J Chem Inf Comput Sci. 1994. Vol. 34, № 3. P. 520–525.

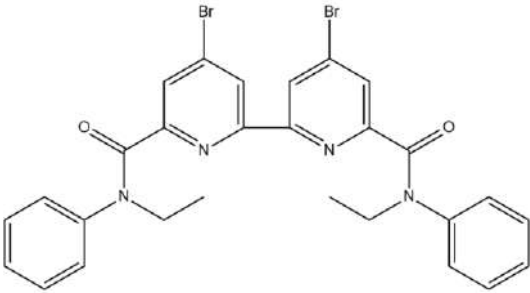
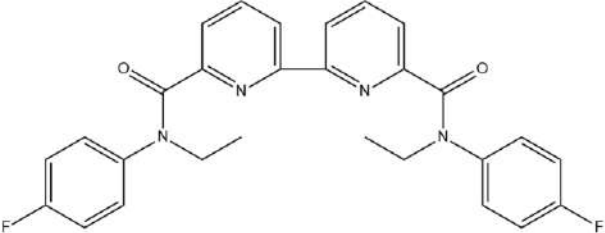
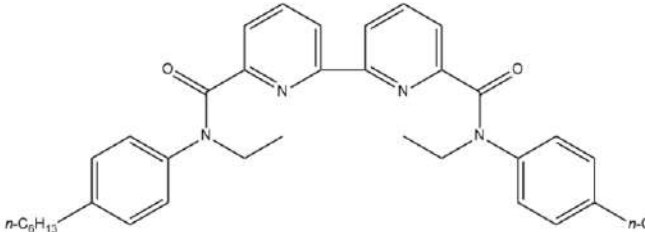
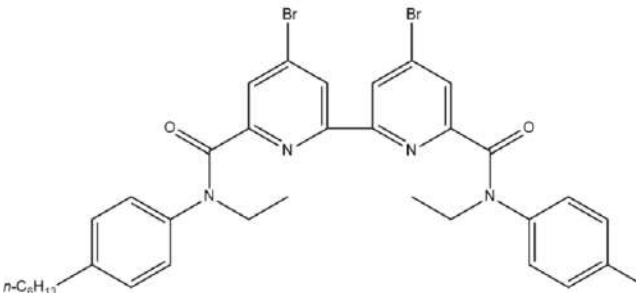
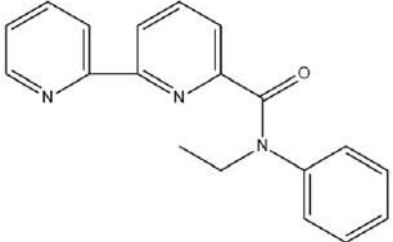
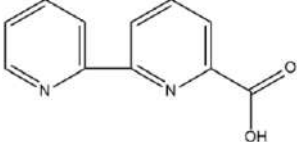
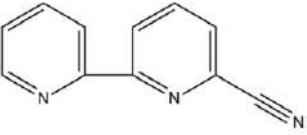
ПРИЛОЖЕНИЕ А

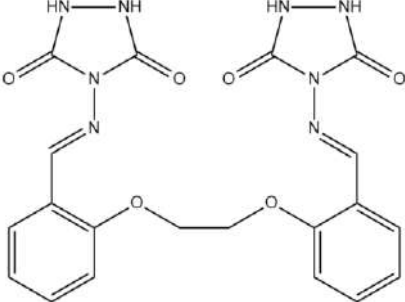
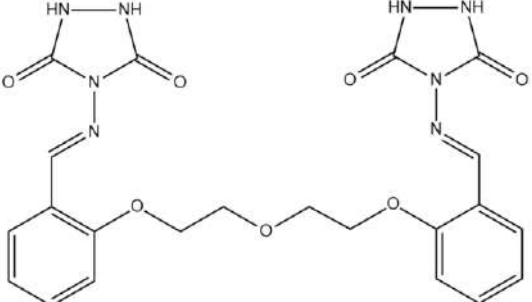
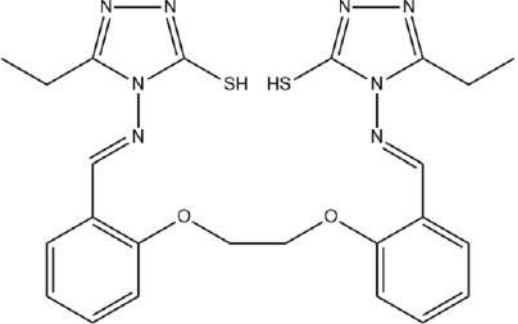
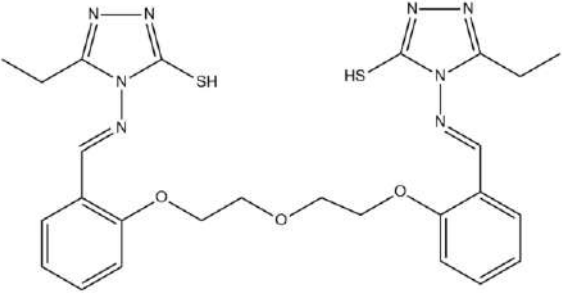
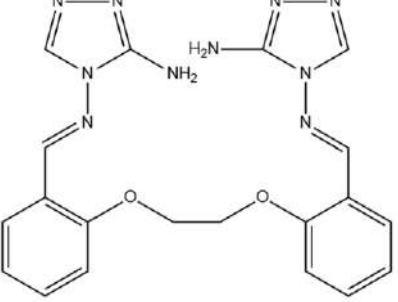
Таблица 1. Структуры и чувствительности к тяжёлым металлам ионофоров

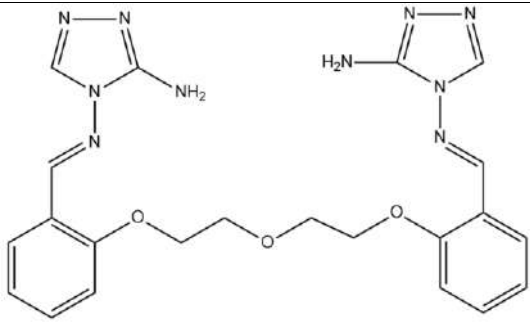
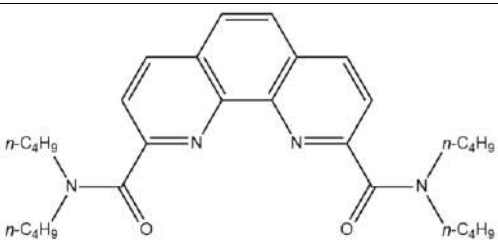
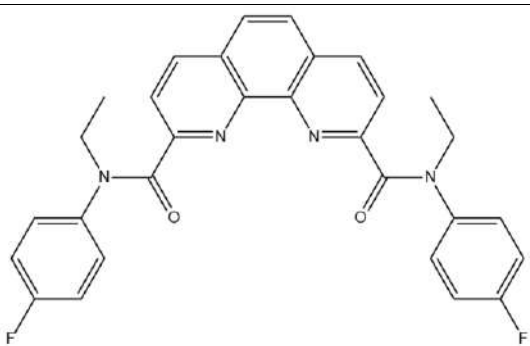
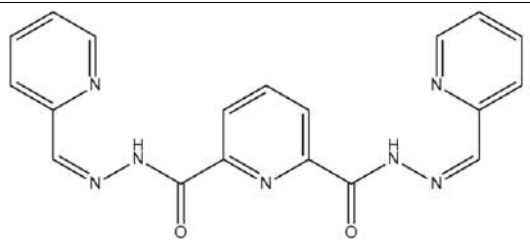
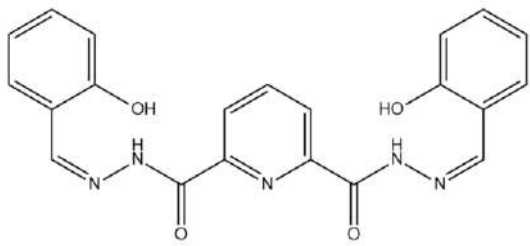
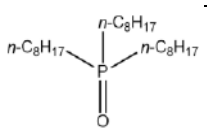
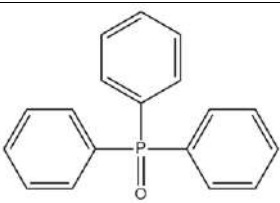
№	Структура ионофора	Чувствительность, мВ/дек		
		Cd ²⁺	Cu ²⁺	Pb ²⁺
1		9	5	30
2		13	12	24
3		13	9	18
4		14	25	27
5		14	23	26
6		27	34	51

7		21	31	37
8		26	34	34
9		32	43	44
10		32	43	45
11		19	37	37
12		23	28	28
13		22	27	28

14		24	31	31
15		25	28	29
16		22	25	42
17		36	31	24
18		37	39	26

19		36	26	28
20		31	30	23
21		36	34	23
22		41	31	27
23		3	-12	0
24		6	1	0
25		5	-10	3

26		16	15	24
27		15	15	24
28		17	19	25
29		24	34	24
30		18	28	27

31		18	20	28
32		26	24	31
33		27	23	26
34		7	0	0
35		5	0	4
36		18	9	9
37		-10	5	-20

38		23	30	38
39		25	30	33
40		27	26	40
41		24	20	16

Таблица 2. Свойства ГЭР на основе кислот и их смесей с водой. Звёздочкой (*) обозначены расчетные значения

Кислота в ГЭР	Содержание воды, масс. %					
	0	10	20	30	40	50
	Коэффициент преломления					
Молочная	1.4811 (*)	1.4660	1.4505	1.4350	1.4170	1.4035
Винная	1.4910	1.4910	1.4635	1.4460	1.4290	1.4115

Яблочная	1.4875	1.4870	1.4575	1.4395	1.4245	1.4060
Лимонная	1.4955	1.4960	1.4645	1.4455	1.4295	1.4130
Малоновая	1.4830	1.4665	1.4500	1.4335	1.4200	1.4020
Гликолевая	1.4856	1.4675	1.4565	1.4445	1.4360	1.4290
Плотность, г/см ³						
Молочная	1.172 (*)	1.155	1.139	1.122	1.104	1.087
Винная	1.330	1.297	1.260	1.226	1.190	1.156
Яблочная	1.271	1.241	1.218	1.187	1.163	1.129
Лимонная	1.340	1.306	1.268	1.226	1.189	1.156
Малоновая	1.232	1.211	1.181	1.161	1.140	1.113
Гликолевая	1.203	1.176	1.157	1.137	1.117	1.099
Вязкость, мПа·с						
Молочная	-	153.6	36.32	13.55	6.752	4.297
Винная	4504994	2435.7	157.5	29.07	10.09	5.099
Яблочная	19098.6	400.8	69.2	18.43	8.499	4.345
Лимонная	10099000	8327	342.7	34.39	11.27	5.598
Малоновая	2208.41	147	21.8	10.83	5.944	3.454
Гликолевая	797.6	73.1	20.08	9.185	4.996	3.373
Проводимость, мСм/см						
Молочная	-	5	16	34	55	75

Винная	0.01	0.015	5	16	34	53
Яблочная	0.077	0.13	9	26	43	63
Лимонная	-	0.002	3	14	29	51
Малоновая	0.38	4.47	16.15	34.5	53	77
Гликолевая	1	8	18	34	56	79