

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
имени М.В. ЛОМОНОСОВА  
Химический факультет

*На правах рукописи*

**Карпов Кирилл Викторович**

**De novo дизайн комплексообразователей  
трехвалентных f-элементов**

1.4.13 Радиохимия

**ДИССЕРТАЦИЯ**

на соискание ученой степени

кандидата химических наук

Научный руководитель:

кандидат химических наук

Митрофанов Артем Александрович

Москва - 2026

## Содержание

Введение .....	5
Глава 1. Обзор литературы .....	10
1.1 Обращение с отработанным ядерным топливом .....	10
1.1.1 Переработка отработавшего ядерного топлива .....	10
1.1.2 Требования при дизайне новых экстрагентов .....	11
1.2 Методы вычислительной химии .....	13
1.2.1 Методы квантовой химии .....	13
1.2.2 Методы искусственного интеллекта в химии .....	16
1.3 Применение методов вычислительной химии для моделирования комплексообразования f-элементов .....	22
1.3.1 Квантовохимические расчеты в радиохимических системах .....	22
1.3.2 Предсказательные модели в комплексообразовании f-элементов .....	26
1.3.3 Генеративные модели для дизайна лигандов .....	29
1.4 Выводы из обзора литературы .....	31
Глава 2. Расчетная часть .....	33
2.1 Квантово-химические расчеты .....	33
2.2 Построение моделей структура-свойство .....	34
2.3 Анализ вкладов молекулярных фрагментов .....	41
2.4 Методы глобальной оптимизации .....	42
2.5 Построение генеративной модели .....	44
Глава 3. Экспериментальная часть .....	45
3.1 Реагенты и материалы .....	45
3.2 Жидкостная экстракция .....	46
3.2.1 Проведение экстракционных экспериментов .....	46
3.2.2 Методы определения содержания металлов .....	46
Глава 4. Результаты и обсуждение .....	48
4.1 Моделирование свойств лигандов и комплексов .....	48
4.1.1 Расчет жесткости лигандов .....	48
4.1.2 Апробация методологии трансферного обучения на примере физико-химических и биомедицинских свойств .....	54

4.1.3 Построение трансферных моделей для предсказания констант устойчивости комплексов .....	56
4.1.4 Построение гибридных моделей для предсказания констант устойчивости комплексов .....	64
4.1.5 Алгоритм поиска равновесной геометрии малых органических молекул .....	71
4.2 Жидкостная экстракция .....	72
4.2.1 Сравнение результатов эксперимента и расчетов .....	73
4.3 Генеративная модель .....	75
4.3.1 Генерация на основе случайных выборок .....	75
Заключение .....	80
Выводы .....	81
Список литературы .....	82

## Список сокращений и условных обозначений

- ОЯТ — отработавшее ядерное топливо
- DFT — density functional theory — теория функционала плотности
- ВЗМО — высшая заполненная молекулярная орбиталь
- НСМО — низшая свободная молекулярная орбиталь
- ТОДГА — тетраоктилдигликольамид
- ТБФ — трибутилфосфат
- ВАО — высокоактивные отходы
- FNN - Feed Forward Network — нейронная сеть прямого распространения
- GNN — Graph Neural Network - графовая нейронная сеть
- MPNN — Message Passing Neural Network - алгоритм передачи сообщений
- t-SNE - t-distributed Stochastic Neighbor Embedding - Стохастическое вложение соседей с t-распределением
- PCA - Principal Component Analysis - Метод главных компонент

## Введение

Получение дешевой и экологически чистой энергии – одна из основных задач промышленности в настоящее время. Важным источником такой энергии является атомная энергетика, чья доля в мировой выработке электроэнергии растет с каждым годом [1]. Переход от углеводородной энергетике к атомной значительно сокращает выбросы парниковых газов, однако использование атомных электростанций (АЭС) предполагает накопление отработавшего ядерного топлива (ОЯТ). Для обращения с ОЯТ принято две стратегии - открытый ядерный топливный цикл (ОЯТЦ) и закрытый ядерный топливный цикл (ЗЯТЦ). В Российской Федерации принята концепция ЗЯТЦ, что ставит перед индустрией задачу переработки ОЯТ.

Переработка ОЯТ является одной из самых важных проблем атомной энергетике. На сегодняшний день, основным способом переработки является выделение урана и плутония для последующего переиспользования для выработки электроэнергии. Оставшаяся часть ОЯТ является высокоактивными отходами (ВАО), которые требуют отдельного обращения с фракционированием и трансмутацией отдельных компонентов. Наиболее сложной задачей при переработке ВАО является выделение и разделение трехвалентных актинидов и лантанидов, которая в промышленности решается методом жидкость-жидкостной экстракции с помощью органических экстрагентов.

Дизайн лигандов для экстракционного разделения f-элементов — нетривиальная задача, требующая учета большого количества факторов, которым должна соответствовать итоговая молекула. Основным способом поиска новых экстрагентов является экспериментальный скрининг, в рамках которого синтезируется ряд перспективных соединений с их последующим тестированием с целью выяснения целевых параметров. Этот процесс требует большого количества ресурсов и исследует небольшое число соединений, из-за чего экспериментальный скрининг, как правило, ограничен определенным набором классов соединений. Существующие вычислительные методы квантовой химии способны с хорошей точностью рассчитывать свойства химических соединений, однако их применение для виртуального скрининга также ограничено выбором соединения и/или

ресурсами требующимися для расчета.

### **Цель работы**

Создание подхода для дизайна новых экстрагентов трехвалентных f-элементов, способного предлагать структуры соединений, обладающих заданными свойствами. В процессе работы решались следующие задачи: (1) Сбор базы данных комплексообразователей катионов трехвалентных лантаноидов и актиноидов, (2) Разработка и проверка на литературных данных предсказательных моделей для оценки устойчивости комплексов металл-лиганд, (3) Экспериментальная проверка полученных предсказательных моделей, (4) Разработка алгоритма de novo дизайна для получения структур новых перспективных комплексообразователей f-элементов.

### **Научная новизна**

1. Впервые проведен количественный анализ того как такие свойства как жесткость, энергия предорганизации и электростатический потенциал экстрагента связаны с эффективностью и селективностью на широком круге соединений в разных экспериментальных условиях.
2. Полученная модель структура-свойство для константы устойчивости металл-лиганд, с одной стороны превосходит аналогичные представленные в литературе модели для ранее моделируемых металлов, а с другой стороны расширяет список металлов для которых может предсказывать целевое свойство.
3. Представленная генеративная модель впервые использовалась для генерации перспективных комплексообразователей f-элементов.

### **Теоретическая и практическая значимость работы**

- Собрана обширная база данных комплексообразователей трехвалентных f-элементов с экспериментальными значениями констант устойчивости комплексов 1:1.
- Установлено, что сочетание таких свойств лиганда как жесткость, энергия предорганизации и электростатический потенциал в точке координации металла вместе с экспериментальными условиями экстракции позволяет предсказывать селективность лиганда с точностью до 0.44 [lg(SF)].

- Предложена новая численная модель, основанная на графовых нейронных сетях для предсказания константы устойчивости комплексов с f-элементами со стехиометрией 1:1 с точностью до 0.4 [lg(K)].
- Создана генеративная модель, которая способна с высокой эффективностью предлагать структуры перспективных комплексообразователей трехвалентных лантанидов и актинидов.

**Методология и методы исследования** В работе использовались современные расчетные методы квантовой химии, основанные на приближении сильной связи и на теории функционала электронной плотности. Для построения статистических предсказательных моделей машинного обучения применялись современные алгоритмы и архитектуры, такие как графовые нейронные сети и бустинговые алгоритмы основанные на деревьях решений. Также для проведения экспериментальных исследований использовались такие современные инструментальные методы как альфа-спектрометрия и гамма-спектрометрия.

#### **Положения, выносимые на защиту**

- Комбинация таких свойств органической молекулы как жесткость, энергия предорганизации и электростатический потенциал могут служить для количественного прогнозирования селективности лиганда по отношению к паре Am(III)/Eu(III), при этом их индивидуальное использование не позволяет осуществлять количественных оценок.
- Разработанные архитектуры нейронных сетей позволяют успешно предсказывать константы устойчивости комплексов трехвалентных f-элементов с точностью до 0.4 [lg(K)], а также корректно ранжировать лиганды по селективности в проведенном экстракционном эксперименте.
- Применение генеративной модели, основанной на эволюционном алгоритме с предложенными моделями “структура-свойство” в качестве функции оценки, позволяет получать структуры перспективных органических комплексообразователей, подходящих для дальнейших исследований.

**Соответствие паспорту научной специальности** Диссертационная работа соответствует паспорту специальности 1.4.13 Радиохимия по следующим областям исследований: Соединения радиоактивных элементов. Свойства. Методы выделения, разделения и очистки радиоактивных элементов и изотопов.

Экстракционные процессы разделения в радиохимии. Химия ядерного топлива. Научные основы радиохимической технологии и проблемы обращения с радиоактивными отходами.

### **Степень достоверности**

Достоверность результатов обеспечивается использованием апробированных теоретических методов и известных литературных данных. Корректность построенных моделей и выполненных расчетов подтверждена сопоставлением с данными литературы и результатами собственных экспериментов, выполненных при непосредственном участии автора

### **Личный вклад**

автора состоит в критическом обзоре литературных данных, сборе экспериментальной базы данных и разработке методов для тренировки моделей структурасвойство, разработке метода поиска равновесной геометрии органических соединений, применении генеративной модели для дизайна новых комплексообразователей f-элементов, написании программного кода и проведении экстракционного эксперимента. Личный вклад автора в совместных публикациях основополагающий.

### **Апробация работы**

Результаты диссертационной работы были представлены в виде стендовых и устных докладов на следующих российских и международных научных конференциях: III-я Международная конференция "Математическое моделирование в материаловедении электронных компонентов" (Москва, Россия, 2021; стендовый доклад), Международная научная конференция студентов, аспирантов и молодых учёных «Ломоносов-2022» (Москва, Россия, 2022; стендовый доклад), X Российская конференция Радиохимия-2022 (Санкт-Петербург, Россия, 2022; стендовый доклад); Научная конференция-школа «Искусственный интеллект в химии и материаловедении» "Artificial Intelligence in Chemistry and Materials Science" (Москва, Россия, устный доклад); XXII Менделеевский съезд по общей и прикладной химии (Федеральная территория "Сириус Россия, устный доклад); XI Российская конференция Радиохимия-2025 (Красноярск, устный доклад).

### **Публикации**

Основные результаты, положения и выводы изложены в 4х статьях общим объемом 2.9 п.л. в рецензируемых научных изданиях, рекомендованных для

защиты в МГУ имени М.В. Ломоносова по специальности и отрасли науки

### **Объём и структура работы**

Диссертация изложена на 93 страницах, состоит из списка сокращений и условных обозначений, введения, обзора литературы, расчётной части, экспериментальной части, обсуждения результатов, заключения, выводов и списка цитируемой литературы (170 наименований). Работа содержит 10 таблиц и 40 рисунков.

# Глава 1. Обзор литературы

## 1.1 Обращение с отработанным ядерным топливом

### 1.1.1 Переработка отработавшего ядерного топлива

Одним из важнейших вызовов атомной промышленности в настоящее время является переработка большого количества ОЯТ, образующегося при работе АЭС. Стандартным подходом при переработке ОЯТ на сегодняшний день является ПУРЕКС процесс (PUREX - Plutonium Uranium Recovery by Extraction) [2], где ОЯТ сначала выдерживают для распада короткоживущих радионуклидов, а затем растворяют в азотной кислоте. Далее из ОЯТ выделяют оставшийся уран и плутоний с помощью жидкость-жидкостной экстракции с 30% раствором три-*n*-бутилфосфата (ТБФ) в керосине в качестве экстрагента.

Водная фаза, образующаяся после выделения урана и плутония, называется высокоактивными отходами (ВАО) и требует дополнительных процедур для безопасной утилизации. Одним из способов обращения с ВАО является его внедрение в стеклянную или керамическую матрицу с последующим захоронением в подземные хранилища. Важной проблемой данного подхода является дизайн такой матрицы, которая могла бы содержать в себе ВАО в течении сотен тысяч лет, что на данный момент не представляется возможным [3]. Альтернативой такому подходу является предварительное фракционирование ВАО — разделение входящих в него компонентов по группам в зависимости от их свойств. Такое разделение позволяет как упростить подбор матрицы под конкретную фракцию, так и дает возможность провести трансмутацию части компонентов [4] (в первую очередь минорных актинидов), что дополнительно снижает итоговую активность 1.1.

Наиболее изученным процессом при фракционировании ВАО является совместное выделение *f*-элементов, для этого применяются каскадные экстракционные системы с триалкилфосфиноксидами с добавлением ТБФ. Дальнейшее выделение и разделение *f*-элементов требует иных экстракционных систем, экстрагенты в которых имеют большие коэффициенты разделения для целевых металлов.

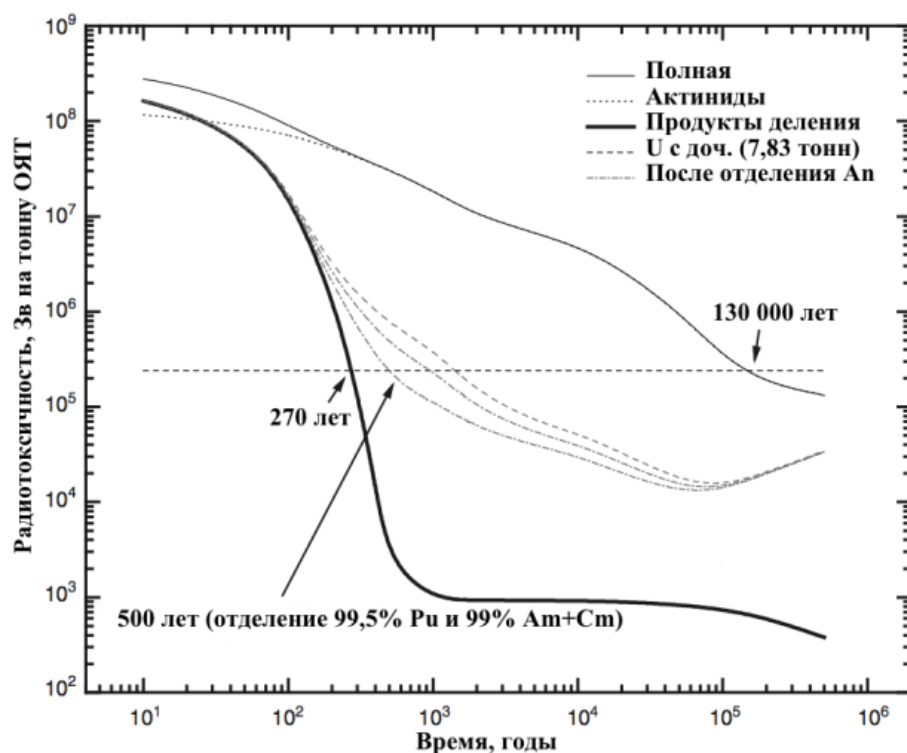


Рис. 1.1 — Зависимость уменьшения радиотоксичности ВАО от времени с учетом выделения различных компонентов [5]

С целью поиска оптимальной экстракционной схемы были созданы несколько систем, использующие различные классы экстрагентов. Так, например TRUEX (TRansUranic Extraction) [6] процесс использует карбомоилфосфиноксиды в качестве экстрагентов, а ряд других систем используют малонамиды и фосфиноксиды, которые, в свою очередь, имеют приемлемые для фракционирования лантанидов и актинидов коэффициенты разделения [4; 7]. Широко исследуемыми в данный момент экстрагентами для разделения актинидов и лантанидов являются производные дигликольамидов, такие как TODGA [8], на основе которых создаются новые экстракционные схемы. И хотя в разработке новых систем экстракции f-элементов происходят серьезные продвижения, все предложенные подходы не решают полностью проблему низкой селективности рассматриваемых лигандов.

### 1.1.2 Требования при дизайне новых экстрагентов

Несмотря на успехи в разделении актинидов и лантанидов существующими экстракционными системами, существует потребность в экстрагентах с большими коэффициентами разделения, которые были бы более селективны по отношению к определенным металлам. Также существует набор параметров, не свя-

занных напрямую с разделением, которые необходимо учитывать при дизайне нового экстрагента.

Основным промышленным методом фракционирования ВАО является противоточная жидкостная экстракция, следовательно, экстрагент (и его комплекс с металлом) должен иметь высокую растворимость целевом растворителе. Низкая растворимость лиганда не позволит эффективно использовать его в качестве комплексообразователя. Современная промышленная технология накладывает ряд ограничений на химическую структуру экстрагента, в первую очередь это касается его радиолитической и химической стабильности в условиях проведения экстракции. Так, например, дозовая нагрузка на экстрагент, участвующий в процессе фракционирования ВАО может составлять порядка 10 кГр за один экстракционный цикл [9], учитывая также необходимость многократного использования органической фазы, её экстракционные свойства не должны ухудшаться при значениях поглощённой дозы в диапазоне 100-1000 кГр [10]. Поскольку основным методом переработки ОЯТ является PUREX процесс, который предполагает растворение ОЯТ в азотной кислоте, органические экстрагенты для фракционирования ВАО должны быть химически устойчивы при ее концентрации в пределах 3-4 моль/л.

Следующей важной характеристикой лиганда является его элементный состав. Существует целый ряд экстракционных систем, имеющих подходящие параметры селективности [4; 7; 11], но при этом содержащие в составе экстрагентов атомы фосфора и серы. Наличие этих атомов приводит к образованию твердых остатков фосфатов и сульфатов при озолении комплексов после экстракции, что в свою очередь негативно сказывается на стабильности матриц при остекловывании ВАО. Таким образом при дизайне нового экстрагента важно придерживаться CHON-принципа (сокращение от Carbon, Hydrogen, Oxygen, Nitrogen – углерод, водород, кислород, азот), то есть использовать в структуре экстрагента только атомы водорода, углерода, кислорода и азота.

Учитывая необходимость в переработке тысяч тонн ОЯТ, предложенные экстрагенты должны быть настолько просты и доступны в синтезе, насколько это возможно. Данная характеристика не связана напрямую с технологической стороной переработки ВАО, но сложный и дорогой в синтезе лиганд приведет к трудностям при масштабировании процесса. Перечисленный выше список свойств, которым должны соответствовать перспективные экстрагенты не яв-

ляется исчерпывающим, однако без соответствия этим параметрам, экстрагент не сможет стать частью промышленной экстракционной системы. Важно отметить, что большинство из перечисленных параметров (а иногда и все из них) проверяются уже после синтеза исследуемых экстрагентов. Это приводит к тому, что ресурсы, потраченные на разработку и синтез перспективного соединения зачастую не приводят к ожидаемым результатам, и даже удачный с точки зрения селективности лиганд может не подойти по другим параметрам. Таким образом, исследователи в этой области сталкиваются с многопараметрической задачей, решение которой с помощью экспериментального скрининга не всегда является оптимальным. Большинство из перечисленных параметров можно предварительно оценить с помощью все более развивающихся в последнее время вычислительных методов и таким образом уменьшить количество необходимых на одно соединение ресурсов, что в свою очередь позволяет охватить больше перспективных лигандов.

## 1.2 Методы вычислительной химии

### 1.2.1 Методы квантовой химии

#### Методы волновой функции

Исторически первым методом квантовой химии можно считать метод Хартри-Фока (ХФ). Этот метод был разработан Дугласом Хартри и Владимиром Фоком как приближенный метод расчета свойств атомов и молекул [12–15]. Необходимость появления такого метода происходит из того, что аналитическое решение уравнения Шредингера (УШ) возможно получить только для одноэлектронных систем. Из записи УШ для атома гелия в приближении Борна-Оппенгеймера (1.1) видно, что 3 член, отвечающий за межэлектронное взаимодействие не позволяет решить это уравнение аналитически для системы с более чем одним электроном.

$$-\frac{\hbar^2}{2m_e}(\nabla_1^2 + \nabla_2^2)\psi(r_1, r_2) - \frac{2e^2}{4\pi\epsilon_0}\left(\frac{1}{r_1} + \frac{1}{r_2}\right)\psi(r_1, r_2) + \frac{e^2}{4\pi\epsilon_0|r_1 - r_2|}\psi(r_1, r_2) = E\psi(r_1, r_2) \quad (1.1)$$

Метод ХФ обходит эту проблему таким образом, что в нем не учитывается

электронная корреляция системы. В нем электрон рассматривается на одноэлектронной орбитали в усредненном поле остальных электронов системы. В настоящее время метод ХФ не применяется для расчетов молекулярных систем ввиду недостаточной точности [16].

С целью повышения точности расчетов были разработаны так называемые «пост Хартри-Фоковские методы», которые позволяют получить приближенные к эксперименту результаты путем введения дополнительных поправок на электронную корреляцию.

Первым из пост-ХФ методов является метод конфигурационного взаимодействия (CI — configuration interaction) [17]. Этот метод является расширением метода ХФ с учетом дополнительных конфигураций — способов размещения электронов системы по всем орбиталям, реальным и виртуальным. Эти методы подразделяются по количеству учитываемых возбужденных состояний — однократно возбужденные (CIS), двойные возбужденные (CISD) и т.д. до полного конфигурационного взаимодействия (FCI). Так CISD позволяет учесть до 95% корреляционной энергии, но сложность расчета растет с учетом каждого дополнительного взаимодействия и FCI, например, может реалистично использоваться только для двухатомных систем.

Следующим способом учета корреляции является применение методов, основанных на теории возмущений [18; 19]. В рамках этой теории, система, точное решение для которой получить нельзя, разделяется на две части — упрощенную, точно решаемую часть и вторую, которая рассматривается как малое возмущение к первой части. Наибольшее распространение получила теория Меллера-Плессе, основанная на теории возмущения многих тел [20]. Как известно, полная энергия получаемая из решения УШ методом ХФ составляет более 99% точного нерелятивистского значения энергии, соответственно оставшаяся часть может рассматриваться как малое возмущение с возможностью применения для этого соответствующего математического аппарата. Так, полный гамильтониан системы складывается из ХФ гамильтониана, для которого есть точное решение и из его возмущения (Уравнение 1.2).

$$\hat{H}_\lambda = \hat{H}_0 + \lambda \hat{V} \quad (1.2)$$

Далее, возмущенная волновая функция представляется в виде степенного ряда по безразмерному параметру, отсюда возмущенная волновая функция пред-

ставляется как линейная комбинация возбужденных детерминантов. Эти методы подразделяются по тому на каком члене обрывается ряд разложения полной энергии системы, так метод, в котором учитываются порядок возбуждения вплоть до  $k$  называется MP $k$ . Наиболее часто используемым методом из этой группы является MP2, поскольку он предлагает хороший баланс между точностью и вычислительными затратами, методы более точные чем MP4 используются редко, поскольку требуют значительных вычислительных ресурсов.

Группа методов, которая обеспечивает наиболее полный учет электронной корреляции называется методами связанных кластеров [21; 22]. В рамках этой теории, коэффициенты членов конфигурационного взаимодействия не рассматриваются как независимые, а также выражаются через кластерные амплитуды с помощью экспоненциального анзаца, что позволяет учитывать не только независимые возбуждения, но и их связные комбинации. Такой подход позволяет с большой точностью учитывать энергию электронной корреляции, но требует больших вычислительных затрат и по этой причине применим только для относительно небольших систем.

Наиболее совершенные методы из группы основанных на волновой функции способны с большой точностью рассчитывать характеристики различных химических систем, но при это они все требуют значительных вычислительных затрат, что ограничивает область их применимости.

### **Теория функционала плотности**

Теория функционала плотности (DFT) заложена серией работ проведенных под руководством Вальтера Кона [23; 24] за что им, совместно с Джоном Поплом, впоследствии была получена Нобелевская премия по химии [25]. В них было доказано, что свойства многоэлектронной системы определяются только электронной плотностью, зависящей от трех координат, а также то, что энергия электронной подсистемы, записанная как функционал электронной плотности, имеет минимум, равный энергии основного состояния. Применение этой теории с формализмом Кона-Шэма позволило разработать квантово-химический метод DFT, который в настоящее время находит широкое применение в вычислительной химии.

Основным вопросом при использовании метода DFT является выбор функционала обменно-корреляционной энергии, который не может быть вычислен

аналитически и подбирается в зависимости от системы. Для описания этого функционала используются различные приближения, которые различаются в точности и вычислительной сложности. К наиболее известным относятся: приближение локальной плотности [26] (LDA - Local Density Approximation), в котором используется приближение однородного электронного газа для неизвестной части функционала, градиентно-исправленные методы [27] (GGA - Generalized Gradient Approximation), в которых энергия зависит также и от градиента электронной плотности, что дополнительно увеличивает точность метода, а также гибридные и двойные-гибридные функционалы [28], в которых в величину обменно-корреляционного члена добавляется часть ХФ или пост-ХФ обмена. При этом наибольшую популярность при изучении молекулярных систем получили именно гибридные функционалы, дающие разумный компромисс между точностью и вычислительной сложностью метода.

### 1.2.2 Методы искусственного интеллекта в химии

С появлением новых высокопроизводительных компьютеров и улучшением алгоритмов [29], методы глубокого машинного обучения, до этого не пользовавшиеся большой популярностью, получили широкое развитие, как в человеческой деятельности в целом, так и в химии в частности.

Наиболее простой и при этом наиболее широко применяемой разновидностью нейронной сети является нейронная сеть прямого распространения (Feed-forward neural network (FNN)) (Рис. 1.2) которая представляет собой набор связанных между собой слоев, состоящих из элементарных искусственных нейронов (перцептронов), каждый из которых принимает на вход набор значений, преобразует их в зависимости от своих внутренних параметров (которые меняются в процессе обучения) и передаёт полученное в результате преобразования значение на последующие перцептроны. Принцип подбора внутренних параметров перцептрона в процессе тренировки называется методом обратного распространения ошибки [30] и является основным методом тренировки нейронных сетей. Данная архитектура нашла широкое применение в химии, в основном в области разработки лекарственных средств (drug-design) [31] и разработке новых материалов (material science) [32].

Одной из основных сложностей при использовании FNN является необходимость подбора способа представления объекта моделирования — перевода его

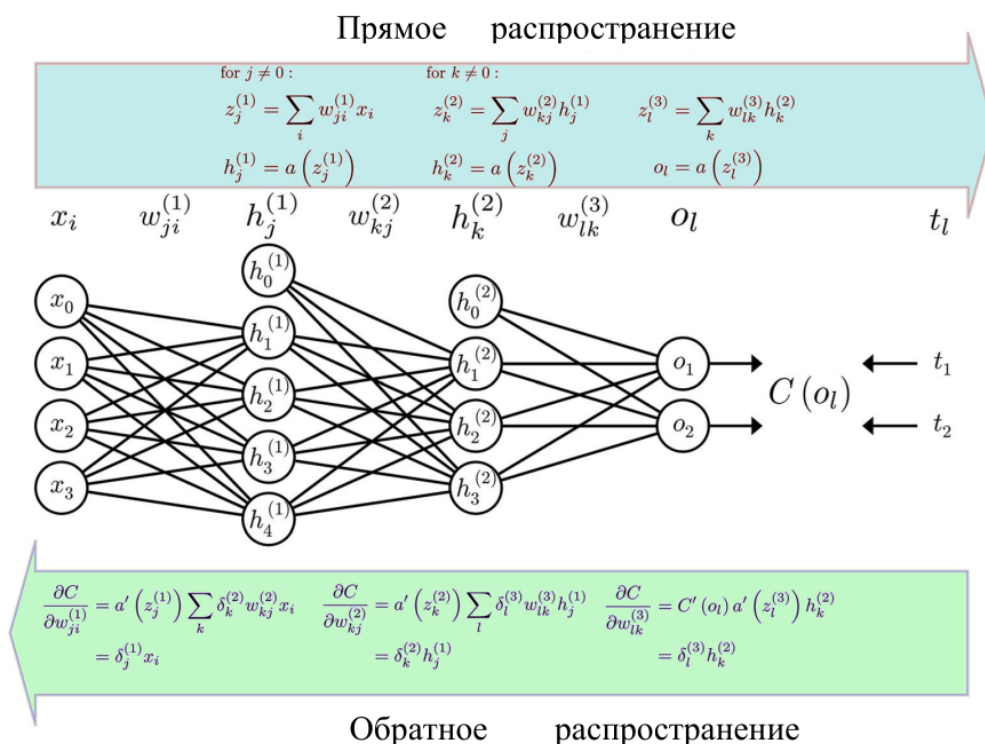
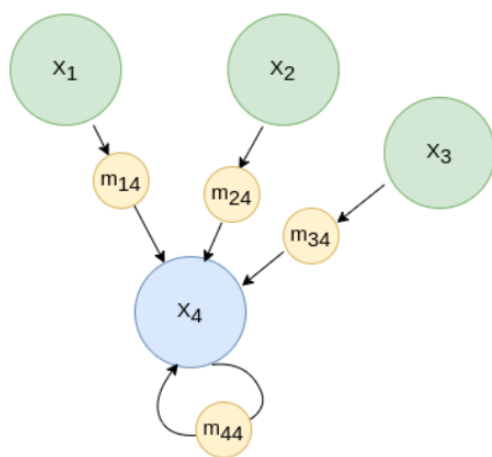


Рис. 1.2 — Схема полносвязной нейронной сети (FNN) с входным, двумя скрытыми и выходным слоями. Синяя стрелка указывает на прямой проход данных, а зеленая — на обратное распространение ошибки (back-propagation). В ходе прямого прохода выходы нейронов вычисляются через взвешенную сумму и функцию активации. На выходе предсказания сети сравниваются с эталонными значениями, а полученная ошибка используется для расчета градиентов и корректировки весов.

в вектор конечной длины, применительно к химии — рассчитать дескриптор химического соединения [33]. Выбор такого представления представляет из себя нетривиальную задачу и оказывает сильное влияние на качество итоговой модели. Существуют альтернативные архитектуры нейронных сетей, которые не нуждаются в специальном представлении объекта, вместо этого обработка объекта происходит во время тренировки модели. Наиболее известными видами таких нейронных сетей являются сверточные нейронные сети для обработки изображений [34] и трансформеры для обработки текста [35]. Еще одной важной архитектурой нейронной сети является графовая нейронная сеть (graph neural network - GNN), которая позволяет обрабатывать объекты представимые в виде графа [36].

Графовые нейронные сети играют особую роль в применении методов искусственного интеллекта в химии, поскольку их архитектура позволяет работать непосредственно с химическими объектами. Особенности строения химических

объектов (малых органических молекул, твердых тел, белков и т.д.) позволяет естественно представлять их в виде графа, в котором атомы или более сложные функциональные группы являются вершинами графа, а химические связи — ребрами [37]. Строение и тренировка графовых нейронных сетей не особо отличается от FNN, за исключением слоев, непосредственно занимающихся обработкой графа. Подробную классификацию графовых нейронных сетей можно найти в работе Джоу и соавторов [36], при этом широкое распространение для химических задач получили графовые сверточные нейронные сети, основанные на передаче сообщений между вершинами (Message Passing Neural Networks - MPNN) [38]. Основой обработки графа в MPNN является обновление значений каждой вершины путем пересылки сообщений между вершинами графа через соединяющие их ребра (Рис. 1.3).



1.  $h_v^{(0)} = x_v \quad \forall v \in V$
2.  $m_v^{(t)} = f_{\text{Agg}}^{(t)}(h_v^{(t-1)}, \{h_u^{(t-1)} : u \in \mathcal{N}(v)\}) \quad 1 \leq t < T$
3.  $h_v^{(t)} = f_{\text{Update}}^{(t)}(h_v^{(t-1)}, m_v^{(t)})$

Рис. 1.3 — Схема работы алгоритма пересылки сообщений в графе. Первым шагом происходит инициализация значений каждой вершины начальным вектором признаков, затем каждая вершина графа обменивается сообщениями с своими соседями и, исходя из значений этих сообщений, происходит обновление списка признаков для каждой вершины.

Важным фактором при этом остается выбор списка признаков вершин и ребер графа. Зачастую в вопросах химии и материаловедения ребра графа не имеют отдельного списка признаков, а определяют лишь связь между вершинами, также существует ряд работ в которых порядки связи в химических структурах использовались в качестве признаков ребер графа [39; 40]. И хотя список свойств ребер графа может влиять на качество модели, наибольший эффект оказывает то, как именно представляются вершины графа — атомы. Как правило сам химический элемент записывается с помощью быстрого кодирования (one-hot-encoding) номера этого элемента в таблице Менделеева, список при-

знаков также дополняется информацией о валентности, заряде, числе соседей и некоторых других свойствах конкретной вершины [41; 42]. В работе Фана [43] они дополнили список признаков физическими свойствами конкретного элемента, а также воспользовались трехмерной структурой молекулы. В настоящее время GNN и, в частности, MPNN показывают одни из лучших результатов при моделировании свойств малых молекул [44] и кристаллических структур [45].

### **De novo дизайн новых химических соединений**

Модели машинного обучения способны за доли секунды получать значение целевого свойства молекулы, что делает их очень удобными при проведении виртуального скрининга, что в свою очередь экономит ресурсы экспериментаторов, давая им возможность исследовать только наиболее перспективные молекулы. Одним из недостатком такого подхода является необходимость ручной генерации базы данных для проведения виртуального скрининга, это приводит к сужению области поиска и трате дополнительных ресурсов на формирование такой выборки.

Возможной альтернативой виртуальному скринингу при поиске новых перспективных химических соединений являются генеративные модели [46; 47]. Генеративные модели представляют из себя алгоритмы, которые предлагают химическое соединение, опираясь на значения оценочной функции, тем самым автоматизируя поиск новых химических соединений. Существует множество типов генеративных моделей, их можно разделить в зависимости от архитектуры, принципа работы, способа представления объектов и т.д., однако их всех объединяет конечная цель — исследование неизученных ранее областей химического пространства в поисках молекул и материалов, которые можно было бы успешно применить под конкретную задачу. Генеративные модели уже достаточно давно используются в области поиска новых лекарственных средств и даже являются одним из промышленных этапов разработки новых соединений, хотя и не могли похвастаться полным экспериментальным подтверждением [48; 49]. Последние работы, например исследования Коршуновой [50] и Жаворонкова [51], показывают значительное улучшение в способности генеративных моделей не просто предлагать выбранные из химического пространства соединения, а непосредственно помогать в решении сложных химических проблем.

Успехи в развитии методов искусственного интеллекта дали начало множеству разнообразных подходов к построению генеративных моделей. Популярными типами генеративных моделей являются генеративно-состязательные сети (Generative adversarial network — GAN) [52–54], генеративные модели на основе вариационных автокодировщиков (VAE) [55], а также модели на основе нормализованного потока [56; 57]. Отдельно от них стоят генеративные модели, основанные на генетическом алгоритме (ГА) [58]. Сам алгоритм состоит из нескольких шагов — формирование начальной популяции, расчет целевой функции для каждого из индивидов в популяции и операции кроссовера и мутации над наиболее подходящими индивидами (Рис. 1.4).

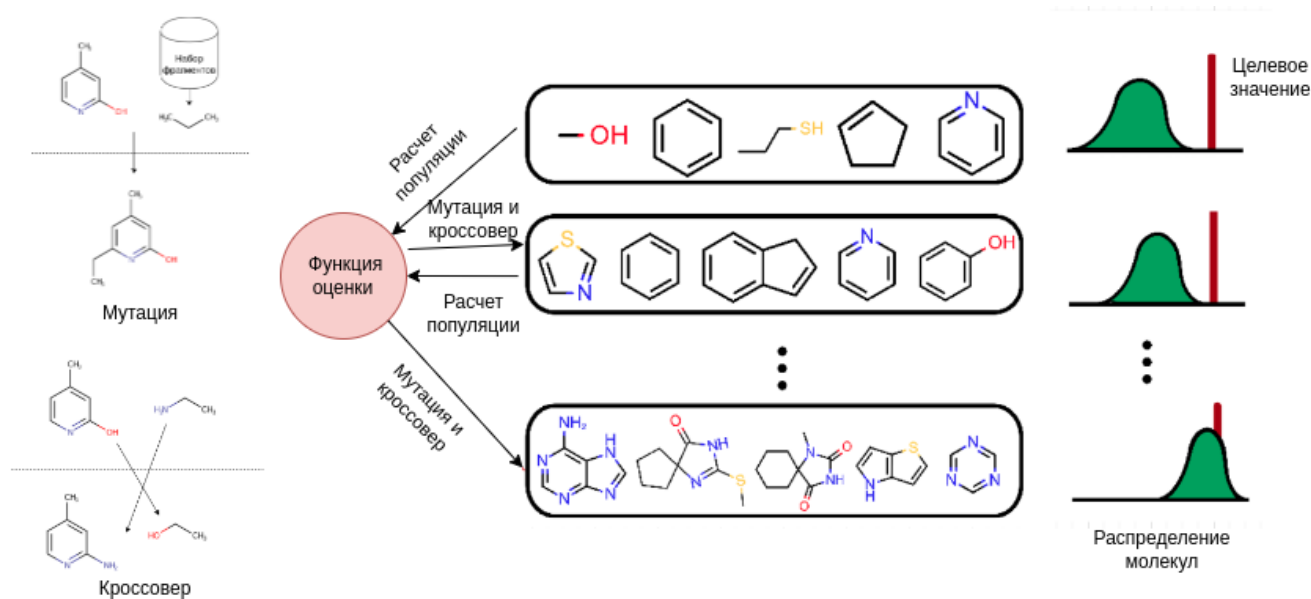


Рис. 1.4 — Схема работы генетического алгоритма для поиска новых химических соединений

ГА как метод глобальной оптимизации находит применение в решении сложных, многопараметрических задач, где традиционные градиентные методы часто неприменимы. Популярность ГА обусловлена их способностью эффективно исследовать обширные пространства поиска, не требуя вычисления производных целевой функции или её непрерывности. Эти алгоритмы особенно эффективны для задач с дискретными переменными благодаря популяционному подходу и операторам кроссовера. Они применяются в инженерном проектировании, машинном обучении, биоинформатике и других областях.

В химии ГА утвердились как стандартный инструмент для решения таких фундаментальных задач, как: оптимизация геометрии и поиск глобальных ми-

нимумов на потенциальных энергетических поверхностях для кластеров, наночастиц и биомолекул; планирование экспериментов и подбор условий синтеза, направленный на максимизацию выхода целевого продукта; дизайн новых молекул и материалов с заданными свойствами, в том числе для поиска новых катализаторов. В этих областях ГА часто выступают основным оптимизационным алгоритмом в специализированных программных пакетах (таких как OGOLEM, GMIN, BCGA), которые интегрированы с квантово-химическими расчётами для прямой оценки энергии системы.

Генетические алгоритмы одними из первых начали применяться для дизайна новых химических соединений [59—61], но при этом все еще активно развиваются, работы по теме выходят достаточно часто и предложенные модели показывают хорошие результаты. Так, Дженсен [62] продемонстрировал генетическую генеративную модель на графах, показывающую высокое качество и малое время работы при относительно небольшой начальной популяции. Еще одна модель, которую предложил Ан [63] в которой объединились генетический алгоритм и нейронная сеть для отбора наиболее подходящих молекул. Квон и Ли предложили MolFinder [64] в котором оптимизация идет на основе метрики сходства между предлагаемыми молекулами, которая уменьшается с каждым новым поколением. Важный толчок применению генетического алгоритма в *de novo* дизайне стал новый метод текстового представления молекул — SELFIES (Self-referencing embedded strings) [65]. В отличие от уже ставшей стандартом нотации SMILES, SELFIES разработаны таким образом, что любая комбинация символов из его алфавита даст на выходе химически правильную молекулу (Рис. 1.5). Эта особенность привела к появлению серии работ посвящённых генеративным моделям, которые используют посимвольную модификацию SELFIES строки как способ поиска новых соединений. Одной из таких работ является STONED алгоритм [66]. В основе этого алгоритма лежат три возможные модификации SELFIES строки — добавление случайного символа из алфавита в случайное место исходной строки, удаление случайного символа из изначальной строки и замена случайного символа на любой другой. В отличие от многих других генеративных моделей, целью которых является поиск одной или нескольких молекул наилучшим образом подходящих под целевое свойство, авторы этой работы сосредоточились на создании локального химического подпространства, при этом учитывая сходство с изначальным со-

единением. Подход STONED оказался удачным и на его основе был создан ряд генеративных моделей с использованием генетического алгоритма. Так, в работе Стейнмана и Дженсена [67] подход к модификации STONED был совмещен с генетическим алгоритмом и целевая функция — показатель докинга, была дополнена фильтром, основанным на наличии определенных функциональных группах и показателем синтетической доступности.

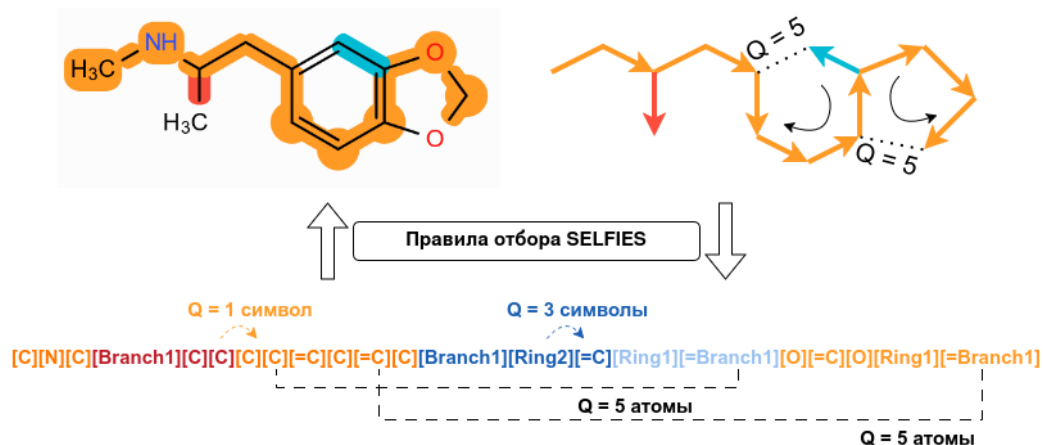


Рис. 1.5 — Пример построения SELFIES нотации

Нигам и соавторы в своей работе [66] усовершенствовали процесс отбора наилучших молекул используя две параллельно изменяющиеся выборки, периодически обменивающиеся между собой соединениями. Одна из выборок отвечает за глобальный поиск по химическому пространству, а другая за локальную оптимизацию наиболее удачных соединений. Полученная модель показала наилучшие результаты в общепринятых тестах генеративных моделей при этом обгоняя методы с похожими метриками качества в скорости.

## 1.3 Применение методов вычислительной химии для моделирования комплексообразования f-элементов

### 1.3.1 Квантовохимические расчеты в радиохимических системах

**Особенности применения квантовохимических методов для систем с тяжелыми элементами**

Среди всех прочих вычислительных методов, методы квантовой химии нашли наиболее широкое применение в области радиохимии. Ранее, ввиду сравнительной простоты и недостаточной развитости вычислительных возможностей, они

позволяли решать сравнительно несложные задачи, например, оценивать вероятности реакций протекающих при получении меченых соединений [68]. По мере совершенствования вычислительных методов, стало возможным перейти к оценке свойств радиоактивных элементов в химических молекулах [69; 70], включая новые сложные классы неорганических соединений [71—73] или материалах матриц, используемых при циклотронном получении изотопов [74]. При этом важно отметить, что расчет геометрии и свойств соединений содержащих тяжелые элементы является достаточно сложной задачей. Помимо того, что для этих многоэлектронных систем существует большое количество близких по энергии состояний, для получения точного результата важное значение принимает учет релятивистских эффектов. Подробнее ознакомиться с особенностями моделирования таких систем можно в монографии Михаэля Долга [75]. Выбор оптимального подхода для расчета структуры и свойств соединений с тяжелыми элементами рассматривался в множестве работ. Так, в работах [76; 77] были проведены сравнения популярных методов квантовой химии в расчетах систем с тяжелыми элементами. Были отмечены необходимость учета эффектов электронной корреляции и релятивистских эффектов. Ожидаемо, наилучшую точность показали расчет методом связанных кластеров, также хорошие результаты были получены при использовании DFT. При этом было показано, что способ учета релятивизма не сильно влиял на точность. Важно отметить, что несмотря на высокую точность расчетов методом связанных кластеров, значительные вычислительные затраты ограничивают этот метод в размере системы, таким образом для большинства реальных приложений метод функционала плотности оказывается предпочтительным вариантом.

При использовании DFT, важным вопросом остается выбор функционала и базисного набора. Подбору оптимального функционала было посвящено несколько работ. Так, в работе [78] были протестированы популярные функционалы (Рис. 1.6) и из результатов можно сделать вывод, что наилучшую точность дают относительно простые функционалы. Это может быть связано с тем, что в работе исследовались относительно простые соединения, с другой стороны, похожие результаты были получены в [79].

К другому выводу пришли в работах [80—82], в них наиболее точные результаты показали гибридные и двойные гибридные функционалы. В последнее время начали появляться работы, в которых предлагаются функционалы непо-

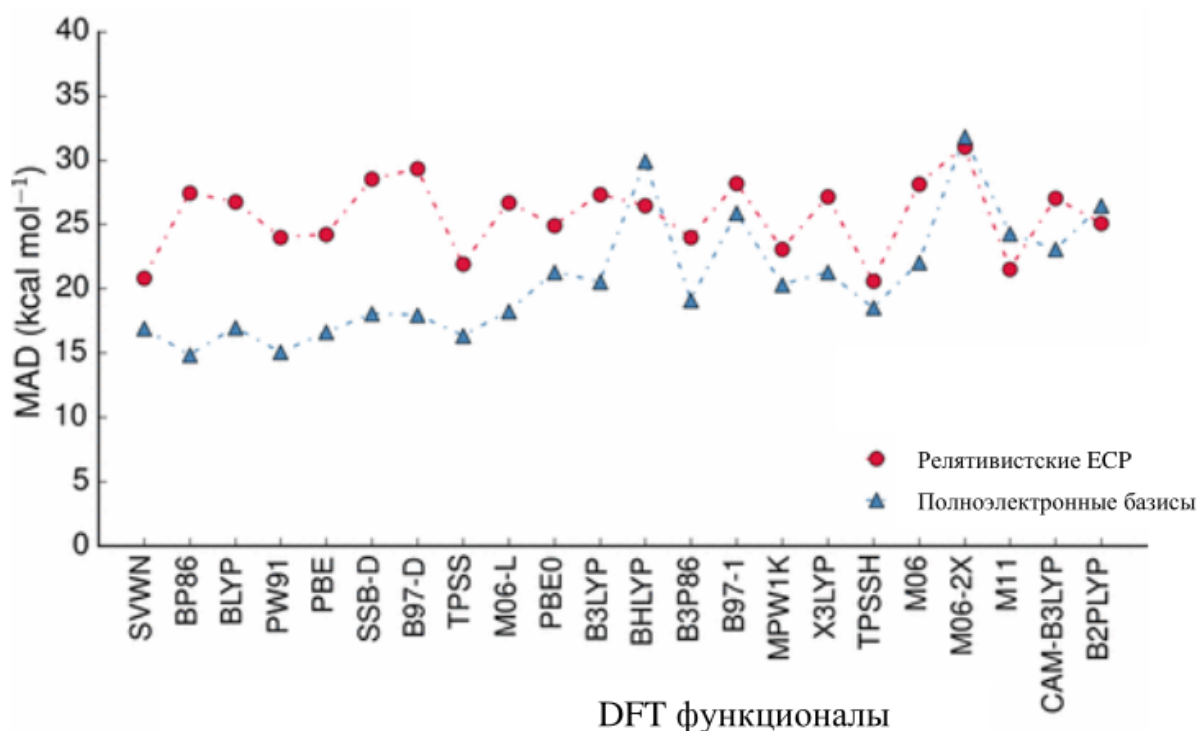


Рис. 1.6 — Погрешности расчета энтропии, полученные при использовании различных функционалов с использованием релятивистских псевдопотенциалов (красные) или полноэлектронных базисных наборов.

средственно для моделирования f-элементов [83; 84], точность которых для этих задач оказывается сравнимой с методом связанных кластеров при значительно меньших вычислительных затратах.

Влияние релятивистские эффекты возрастает пропорционально квадрату атомного номера элемента. Их влияние приводит к стабилизации и сжатию атомных s- и p- орбиталей, при этом d- и f- орбитали, напротив, становятся более диффузными и их энергия повышается. В частности, релятивистскими эффектами объясняется эффект лантаноидного сжатия [85]. По этой причине, при проведении расчетов для систем с тяжелыми элементами, учет эффектов релятивизма необходим для получения точных результатов. Наиболее распространенным способом учета является учет скалярных релятивистских эффектов в рамках численных двухкомпонентных (“approximate two-component”, “A2C”) приближений, таких как приближение нулевого порядка (Zeroth-Order Regular Approximation - ZORA) [86; 87] или приближение Дугласа-Кролла-Гесса (“DKH”) [88; 89]. Было проведено несколько работ, в которых эти методы сравнивались между собой и с *ab initio* мультikonфигурационным подходом [79; 90]. В них было показано, что оба подхода хорошо воспроизводят эксперимен-

тальные данные, при этом не сильно отличаясь по точности от более сложных методов.

### **Дизайн перспективных экстрагентов для разделения с помощью методов квантовой химии**

Одним из основных вопросов при исследовании комплексообразования f-элементов является подбор дескриптора, который бы хорошо описывал этот процесс. Так в работе [91] исследовались полидентатные N-донорные лиганды на роль комплексообразователей Am(III) и Eu(III). Авторы использовали энтальпию реакции, рассчитанную методом DFT как дескриптор процесса комплексообразования, результаты показали корреляцию между рассчитанными величинами энтальпии и экспериментальными данными о разделении. Похожие системы, а также некоторые другие (Рис. 1.7) были изучены в работе [92], авторы работы также использовали DFT в качестве расчётного метода, а также применили релятивистский потенциал для металлов. Анализ комплексообразования описывали сравнивая прочность и длину связи между металлом и координационными центрами лиганда, в результате чего пришли к выводу, что связь Am-N прочнее Eu-N для подобных соединений, что и обеспечивает экстракцию америция. Исследователи в работе [93] пришли к выводу, что селективность обеспечивается разностью в ковалентности связи между координационными атомами азота и металлов, к аналогичному выводу также приходили авторы работ [94; 95].

Одним из дескрипторов, с помощью которого пытаются описывать комплексообразование f-элементов является характеристика жесткости и мягкости лигандов, которая выражается как полуразность энергий ВЗМО и НСМО. Так, в работе [96] с помощью DFT был изучен набор полидентатных лигандов как экстрагентов для разделения пар Am/Eu и Am/Cm. На примере лиганда триазинил бипиридин и модификации на основе триазинила фенантролина, предложенной в статье, авторы показали (Рис. 1.8), что большая селективность в паре Am/Eu нового лиганда обеспечивается меньшей жесткостью фенантролинового лиганда по отношению к бипиридиновому. В работе [97] авторы, вместе с поиском наиболее подходящего способа учета релятивизма при анализе связей с актинидами, также проанализировали взаимосвязь между энергиями граничных орбиталей и прочностью связи между f-элементом и координационными атомами лиганда. На примере комплексов An-(DOPO)<sub>3</sub> (An = Pu, Am, Cf and

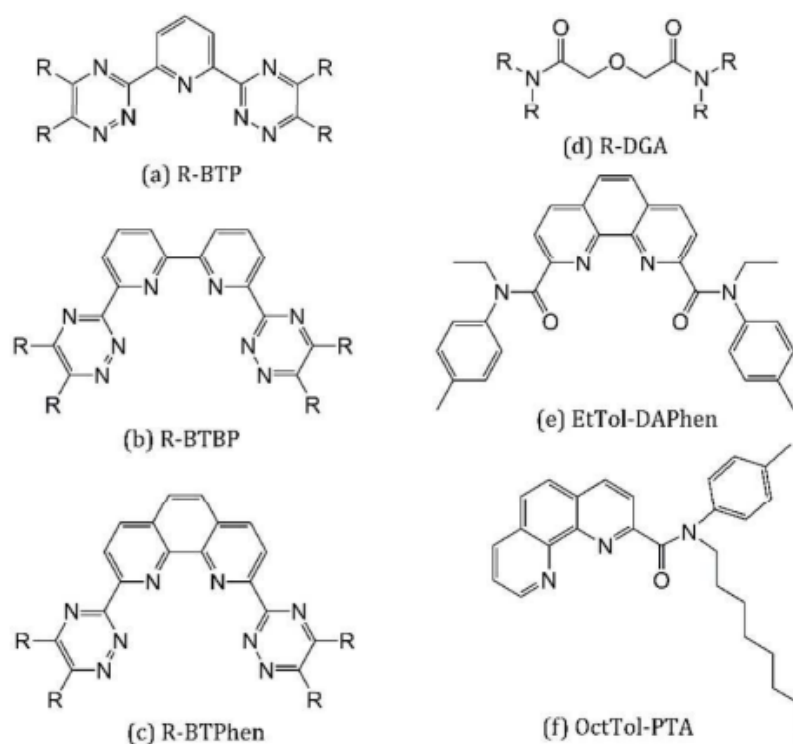


Рис. 1.7 — Примеры лигандов, пригодных для разделения Am и Eu, исследованные в работе [92].

Вк) авторы показали вклады  $5f$ - орбитали в ВЗМО и НСМО, а также в формирование связи металл-лиганд.

В указанных выше работах показано, что такие величины как константа устойчивости комплекса, значения энергий граничных орбиталей являются хорошими дескрипторами для описания комплексообразования в комплексах с  $f$ -элементами. Однако высокая вычислительная сложность квантовохимических методов позволяет провести расчет этих величин только для относительно небольшого числа соединений.

### 1.3.2 Предсказательные модели в комплексообразовании $f$ -элементов

Стремительное развитие методов искусственного интеллекта вообще и в химии в частности открыло перед исследователями новые возможности для дизайна новых молекул и материалов. Хотя наибольшее распространение эти методы нашли в области дизайна новых лекарственных средств, использование моделей «структура-свойство» в области комплексообразования металлов также развивается. Основным параметром для моделирования методами машинного обучения в комплексообразовании металлов является константа устойчивости комплекса. С одной стороны, константа устойчивости непосредственно характеризует

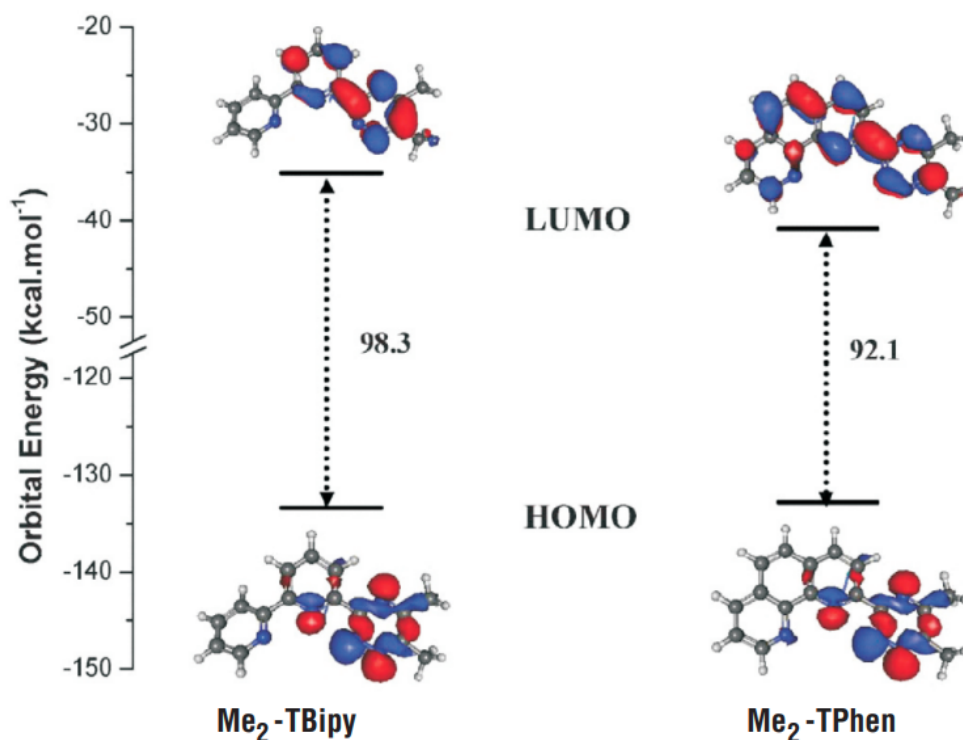


Рис. 1.8 — Сравнение величин энергий граничных орбиталей бипиридинового (Me<sub>2</sub>-TBipy) и фенантролинового (Me<sub>2</sub>-Tphen) лигандов [96].

насколько хорошо лиганд связывается с конкретным металлом, а с другой, это удобный параметр, который можно определить как с помощью вычислительных методов так и экспериментально.

Первые работы по моделированию константы устойчивости металл-лиганд появились еще до взрыва интереса к методам искусственного интеллекта. В работе Соловьева и соавторов [98] были описаны подструктурные молекулярные фрагменты в качестве способа описания лиганда, модели на которых показали высокий результат. В работе [99] Тетко и соавторы моделировали константы устойчивости Ag<sup>+</sup> и Eu<sup>3+</sup> используя различные методы машинного обучения и способы представления молекул. Из работы был сделан вывод, что выбор оптимального сочетания метода машинного обучения и способа представления молекулы позволят получать модели высокого качества. Важным этапом в моделировании константы устойчивости была разработка программы ISIDA [100], в которой реализовали метод множественной линейной регрессии (MLR). Применение ISIDA позволило получить модели машинного обучения для большого количества металлов [101; 102] (в том числе f-элементов), при этом демонстрируя хорошее качество. В этой программе также был реализован алгоритм поиска фрагментов, вносящих наибольший вклад в целевое свойство.

В более поздних работах уже применялись современные методы машинного обучения. Так, в работе [103] было проведено моделирование констант устойчивости используя MLR и FNN, авторы пришли к выводу, что модель основанная на FNN в сравнении показывает лучший результат. В работе [104] Кайто и соавторы провели моделирование общей константы устойчивости для 57 металлов (в том числе f-элементов) и 2706 лигандов используя метод гауссовой регрессии. Авторы работы провели скрупулезный отбор признаков (итоговый вектор длиной 118 был составлен из смеси физико-химических и топологических дескрипторов) для моделирования и в итоге смогли выделить наиболее важные из них (Рис. 1.9).

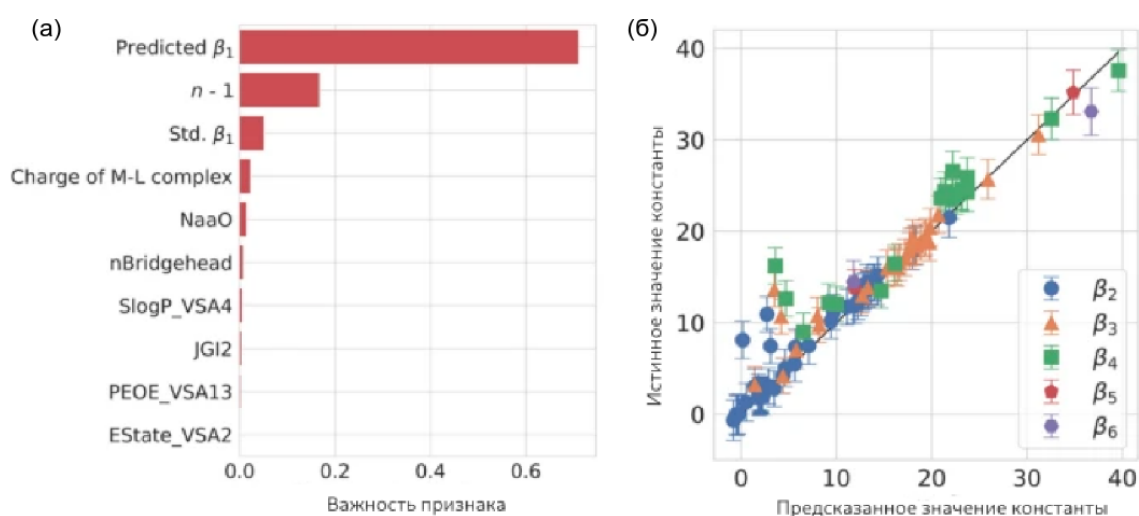


Рис. 1.9 — Наиболее важные признаки (а) и качество полученной модели (б) в работе [104]

Еще одной работой, в которой современные модели машинного обучения применялись для моделирования комплексообразователей f-элементов была работа Митрофанова и соавторов [105]. В ней были построены модели машинного обучения на основе FNN, лиганды описывались топологическими молекулярными фрагментами. Применённые методы позволили получить модели высокого качества, а проанализировав вклады дескрипторов в константу устойчивости, авторы определили фрагменты которые вносят наибольший вклад в комплексообразование (Рис. 1.10).

Из рассмотренных работ видно, что применение методов машинного обучения для моделирования константы устойчивости металл-лиганд позволяет получить точность предсказаний, достаточную для качественной, а иногда и количественной оценки целевого свойства, при этом вычислительная сложность

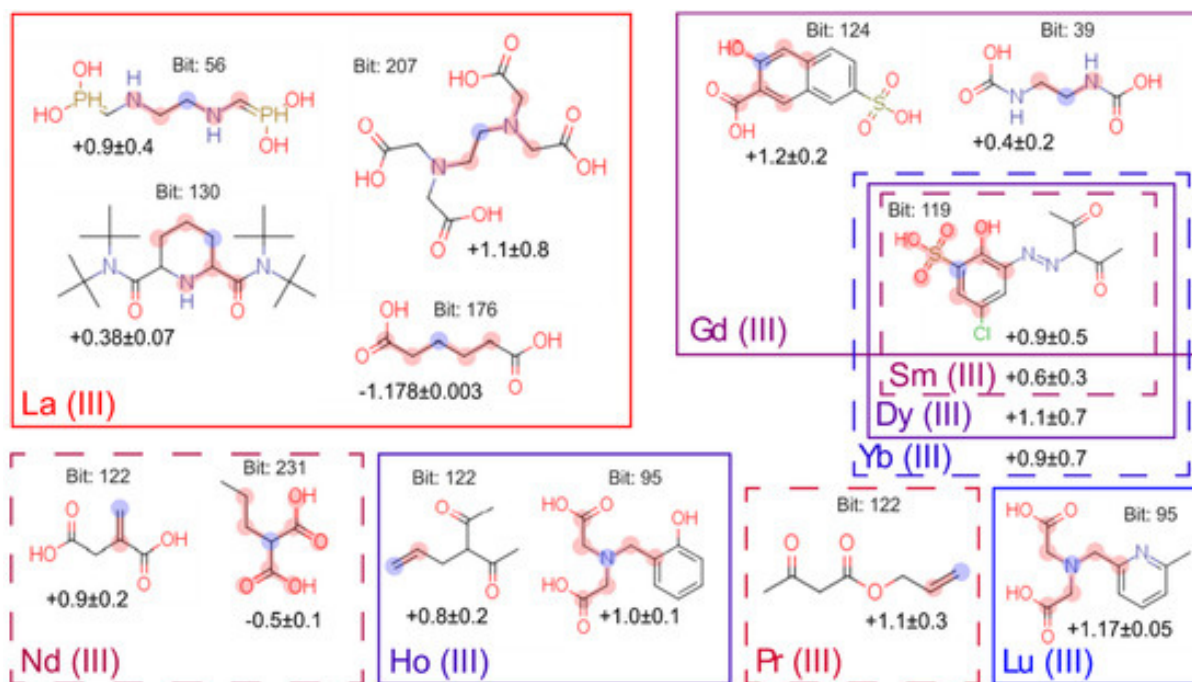


Рис. 1.10 — Анализ наиболее важных признаков, которые влияют на комплексообразование металл-лиганд. Фрагменты выделены красным цветом (синим — центральный атом в соответствии с алгоритмом Моргана); значения аддитивного вклада в константу устойчивости приведены под каждым металлом. [105].

таких методов несравнимо ниже других методов вычислительной химии. Нерешенной проблемой применения методов искусственного интеллекта в этой области остаётся недостаток данных. Для большинства современных методов машинного обучения необходимо несколько сотен, а лучше тысяч точек в обучающем наборе для получения качественной модели. Количество экспериментальных данных для минорных актининов при этом не превышает десятков измерений, что делает невозможным прямое построение моделей для них. Малые обучающие выборки также негативно влияют на область применимости получаемых моделей, из за чего их использование в реальных приложениях также сильно ограничено.

### 1.3.3 Генеративные модели для дизайна лигандов

Возможность быстрой оценки свойств характеризующих комплексообразование позволяет значительно расширять область поиска перспективных лигандов с помощью виртуального скрининга по известным базам данных. Несмотря на это, суммарное количество малых органических молекул в существующих на сегодняшний день базах данных составляет лишь малую долю химическо-

го пространства [106; 107]. И хотя специфика области комплексообразования ОЯТ немного уменьшает потенциальную область поиска, исследовать ее всю методом прямого перебора, даже имея быстрый способ первичной оценки целевого свойства, не представляется возможным. Применение методов de novo дизайна в этой области пока недостаточно развито, однако способы решения аналогичных задач в литературе присутствуют.

Так, существует серия недавних работ по применению генеративных моделей в дизайне комплексообразователей переходных металлов. Так, в работе Шилтера и соавторов [108] решалась задача поиска комплексообразователей для задачи катализа. В качестве генеративной модели использовался вариационный автоенкодер на SMILES строках. В результате работы авторы смогли сгенерировать более 8000 новых и химически правильных молекул с подходящими свойствами. В еще одной работе [109] Ли и соавторы применили вариационный автоенкодер на деревьях сочленений в качестве генеративной модели, рассматривая силовое поле лиганда в качестве целевого параметра. В статье был также применен способ аугментации данных с целью увеличения обучающей выборки. Модель генерировала новые соединения основываясь одновременно на структурном сходстве и на основе величины параметра (Рис. 1.11).

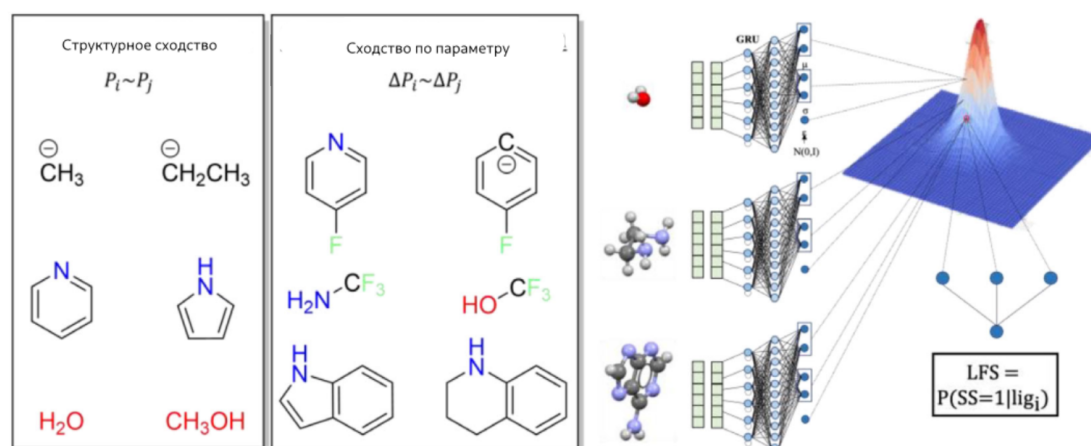


Рис. 1.11 — Типы сходства для аугментации данных (слева) и структура генеративной модели (справа).[109]

Иной подход к генерации новых соединений применили Дуан с соавторами [110]. В этой работе они использовали алгоритм глобальной оптимизации в искусственно расширенном наборе соединений для поиска наиболее подходящих кандидатов для последующего расчета их параметров и последующего применения в активном обучении модели. По словам автором, их подход по выбору

наиболее подходящего лиганда превосходит по скорости базовый случайный выбор в 1000 раз. Наиболее подходящие из найденных молекулы были оценены с помощью метода DFT и большинство из них подтвердили свои свойства. Также для задач катализа был адаптирован метод основанный на SELFIES. Лаплаза и соавторы [111] предложили свой метод NaviCatGA, который использует генетический алгоритм в сочетании с несколькими способами представления (XYZ, SMILES, SELFIES). По словам авторов предложенный подход позволяет с большей эффективностью подбирать катализатор под определенную реакцию.

Несмотря на то, что применение генеративных моделей в вопросах комплексообразования металлов развито не так сильно как, например, в дизайне лекарственных средств, необходимо отметить неоспоримые преимущества, которые приносит использование этой группы методов. Из них наиболее гибкими оказались методы основанные на генетическом алгоритме, позволяющие с высокой эффективностью исследовать химическое пространство, при этом оптимизируя целевое свойство.

## 1.4 Выводы из обзора литературы

Из обзора существующих методов дизайна перспективных комплексообразователей для переработки ОЯТ можно сделать вывод, что вычислительные методы получают все большее распространение среди исследователей в этой области. Возможность оценить такие важные для комплексообразования свойства как константа устойчивости комплекса, энергии граничных орбиталей без предварительного синтеза и проведения экспериментального тестирования открывает новые возможности для дизайна. Современные квантовохимические методы позволяют с точностью, близкой к экспериментальной, рассчитать необходимые параметры системы и сделать выводы о целесообразности дальнейших тестов. Результаты квантовохимического моделирования также позволяют больше узнать о самом процессе комплексообразования и экстракции, выделить важные функциональные группы и определить природу происходящих процессов. При этом высокие вычислительные затраты на такие расчеты не позволяют провести оценку свойств большого количества соединений. Для наиболее точных методов также существуют ограничения в размерах исследуемых систем. Сочетание высоких вычислительных затрат с огромными размерами химиче-

ского пространства приводит к тому, что поиск новых комплексообразователей ведется в относительно узкой области классов соединений. Современные методы машинного обучения способны частично решить эту проблему, на оценку свойства отдельно взятой молекулы с помощью предсказательной модели «структура-свойство» уходят доли секунды, что значительно упрощает виртуальный скрининг по большим базам данных. Однако даже самые большие базы химических данных, доступные на сегодняшний день, являются лишь малой долей полного химического пространства и прямой его перебор, пусть и значительно более быстрым методом, все равно не является возможным. Потенциальное использование генеративных моделей, пусть и не применявшихся ранее в этой области, способно помочь в решении этой задачи. Важным нерешенным вопросом является отсутствие достаточного количества данных для построения предсказательных моделей для наиболее интересных исследователям объектов. Возможность обойти данную проблему и получить предсказательную модель высокого качества для дескрипторов, описывающих комплексообразование металл-лиганд, в сочетании с генеративной моделью, которая бы предлагала перспективные химические соединения под заданное свойство, позволило бы создать инструмент для эффективного дизайна комплексообразователей для переработки ОЯТ.

## Глава 2. Расчетная часть

### 2.1 Квантово-химические расчеты

Одной из важных частей этой работы является определение величин свойств лиганда, от которых зависит стабильность комплексов "металл-лиганд" и анализ их взаимосвязи. Для определения этих свойств мы использовали различные квантово-химические методы и силовые поля. Одна из первых задач, которую необходимо решить при расчете электронных свойств соединения это определение его конформации. В случае если в литературе нет экспериментальных или расчетных данных о структуре выбранной системы, ее необходимо определить самостоятельно. Для этого, после построения двумерного молекулярного графа выбранной системы [112; 113], определяется ее трехмерная геометрия, используя методы основанные на силовых полях. В этой работе мы применяли GFN-FF в сочетании с алгоритмом ANCopt [114; 115], с помощью которых генерировались как конформационные глобальные минимумы, так и ансамбли конформаций рассматриваемых систем. Геометрии всех лигандов оптимизировались в приближении газовой фазы.

Из литературы известно, что силовые поля обладают ограниченной точностью, особенно в отношении систем с тяжелыми элементами, такими как лантаниды и актиниды, что осложняет поиск равновесной геометрии. Для получения более точных результатов при оптимизации геометрии мы применяли полуэмпирические квантово-химические методы, в частности для уточнения геометрии органических лигандов мы использовали метод GFN2-xTB [115], а для комплексов с лантанидами метод PM7 из программного пакета MOPAC в сочетании со Sparkle моделью для описания лантанидов [116; 117].

Для определения свойств лигандов и комплексов мы использовали теорию функционала электронной плотности. При расчете систем без тяжелых элементов применялся гибридный функционал B3LYP [118; 119] и Попловский базисный набор 6-31G\* [120]. В случае систем с лантанидами или актинидами мы применяли гибридный функционал relPBE [83] в сочетании с базисным набором def2-TZVP [121]. Все квантово-химические расчеты проводились в програм-

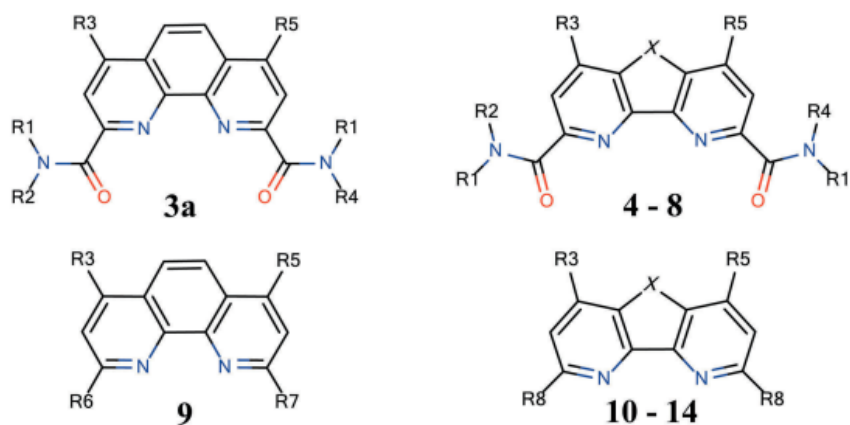
ме Orca [122–128].

Моделируемая жесткость лигандов вычислялась как полуразность энергий ВЗМО и НСМО в соответствии с теорией Пирсона [129; 130]. Энергия предорганизации лиганда вычислялась как разница полных энергий равновесной геометрии и предорганизованного состояния, которое определялась оптимизируя геометрию комплекса и последующего удаления металла и противоионов. Для определения конформационной мобильности лиганда мы составляли набор конформаций с помощью алгоритма ConfGen библиотеки RDkit [131; 132] и вычисляли дисперсию энергий сгенерированных конформаций. Для определения электростатического потенциала лиганда мы рассчитывали электростатический потенциал предорганизованного лиганда в точке координации с помощью программы MultiWFN [133; 134].

## 2.2 Построение моделей структура-свойство

В рамках этой работы были построены модели структура-свойство для предсказания жесткости лиганда и константы устойчивости комплекса — дескрипторов комплексообразования металл-лиганд. Для обоих свойств, процесс сводился к нескольким этапам. В первую очередь собиралась база данных соединений и соответствующих им целевых свойств. Следующим шагом эти данные обрабатывались и подготавливались для их последующего использования в качестве обучающего набора модели. Далее для собранных данных проводился поиск оптимального способа представления. После подготовки данных подбирались оптимальная архитектура модели машинного обучения, на основе которой уже тренировалась модель структура-свойство.

Для предсказания жёсткости N-донорных гетероциклов мы использовали программный пакет ISIDA [100; 135], в котором реализован метод множественной линейной регрессии. Для представления молекул был применен подход подструктурных молекулярных фрагментов, включенный в пакет ISIDA. Данные по жесткости лигандов, как и сами структуры лигандов были получены нами самостоятельно. База структур насчитывающая 2000 соединений была собрана из молекул, принадлежащих к перспективному классу N-донорных гетероциклов (Рис. 2.1), а их жесткость была рассчитана с помощью методов квантовой химии.



X =  $-\text{CH}_2\text{CH}_2-$  (**4**, **10**),  $>\text{C}=\text{O}$  (**5**, **11**),  $>\text{NMe}$  (**6**, **12**),  $>\text{CMe}_2$  (**7**, **13**),  $-\text{O}-$  (**8**, **14**)

R1, R2, R4 = H, Me, Et,  $\text{CH}_2\text{CH}_2\text{OH}$ , Ph

R3, R5 = H, Cl, F, Me,  $\text{NO}_2$ , OMe, OEt,  $\text{CH}_2\text{OMe}$  (**3a-8**)

R3, R5 = H (**9-14**)

R6, R7 =  $\text{CH}_2\text{OH}$ ,  $\text{CMe}_2\text{OH}$ ,  $\text{CPh}_2\text{OH}$ ,  $\text{C}(\text{O})\text{Me}$ ,  $\text{C}(\text{O})\text{CF}_3$ ,  $\text{C}(\text{O})\text{CHMe}_2$ ,  $\text{C}(\text{O})\text{CHMeEt}$ ,  $\text{C}(\text{O})\text{CHEt}_2$ ,  $\text{C}(\text{O})\text{CHMeCH}_2\text{CH}_2\text{OH}$ ,  $\text{C}(\text{O})\text{CHEtCH}_2\text{CH}_2\text{OH}$ ,  $\text{C}(\text{O})\text{Ph}$ ,  $\text{C}(\text{O})\text{CHMePh}$ ,  $\text{C}(\text{O})\text{CHEtPh}$

R8 = H, Me, OMe,  $\text{CF}_3$ , F, CN,  $\text{C}(\text{O})\text{NHOH}$ ,  $\text{CH}_2\text{NHCH}_2\text{Ph}$ ,  $\text{CH}_2\text{NH}(2-\text{C}_6\text{H}_4\text{OH})$ ,  $\text{CH}_2\text{NHC}(\text{O})(2-\text{C}_6\text{H}_4\text{OH})$ ,  $\text{C}(\text{O})\text{NHNHC}(\text{O})\text{Me}$ ,  $\text{C}(\text{O})\text{NHCH}_2\text{CH}_2\text{C}(\text{O})\text{NHOH}$

Рис. 2.1 — Каркасы соединений и функциональные группы, на основании которых собиралась база данных для тренировки модели

В процессе тренировки обучался ансамбль линейных регрессий, из этого ансамбля отбирались наиболее качественные модели, которые затем использовались для получения консенсусного предсказания.

При моделировании экспериментального фактора разделения  $SF_{Am/Eu}$  использовались квантовохимические дескрипторы для признакового описания лиганда и метод машинного обучения CatBoost [136]. База данных состоящая из 77 экстракционных систем и факторов разделения была собрана из литературы, преимущественно из работ, в которых жесткость лиганда упоминалась как основной или один из параметров при отборе экстрагентов [137–144] (Рис. 2.2).

Важным дескриптором комплексообразования является константа устойчивости комплекса металл-лиганд. Для моделирования этого свойства нами была собрана база данных комплексообразователей актининов и лантанинов, данные собирались, используя литературные источники [91; 105; 145; 146] и базу NIST46. Из собранных данных были убраны неорганические соединения и химически неправильные структуры. Построение моделей на основе собранных

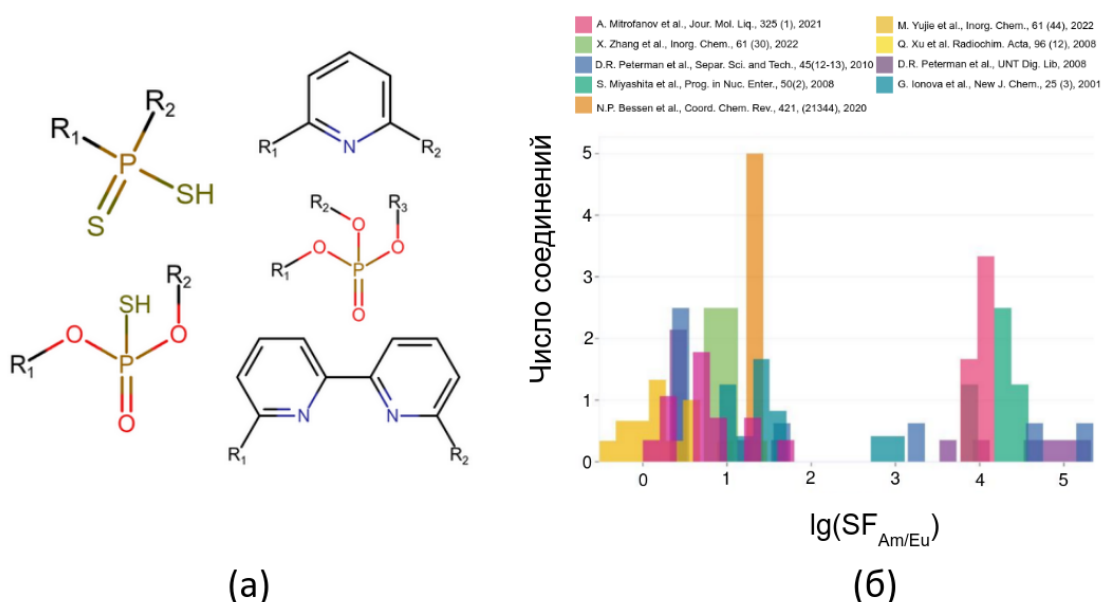


Рис. 2.2 — База данных факторов разделения. а - примеры каркасов соединений входящих в собранную базу данных, б - распределение факторов разделения для разных источников

данных велось двумя способами — используя условия при которых проходил эксперимент (температура и ионная сила) в качестве признаков модели и без использования этих данных. В случае, когда условия эксперимента не использовались, тренировочный набор состоял из значений, полученных при величине ионной силы от 0.1 до 0.3М и температурах от 20 до 25 °С. Общее число лигандов, для которых были собраны данные о константе устойчивости с различными металлами представлено на графике 2.3.

Основными классами соединений в собранных наборах данных были карбоновые кислоты, аминокислоты и нейтральные N-/O- донорные лиганды. В данных присутствовали как лиганды, специально разработанные для связывания металлов, так и лиганды с низкими значениями констант устойчивости. Распределение величин констант устойчивости представлено на графике 2.4, характер распределения для других лантанидов аналогичен распределению констант для Eu.

Поскольку для тренировки моделей машинного обучения была выбрана графовая архитектура нейронной сети, собранные лиганды не нуждались в отдельном способе представления. Используя информацию о связях между атомами из .sdf файлов, в которых хранились данные, молекулы были преобразованы в двумерные графы, где атомы являлись вершинами, а связи ребрами. Для описания вершин графа применялся стандартный признаковый вектор для описания

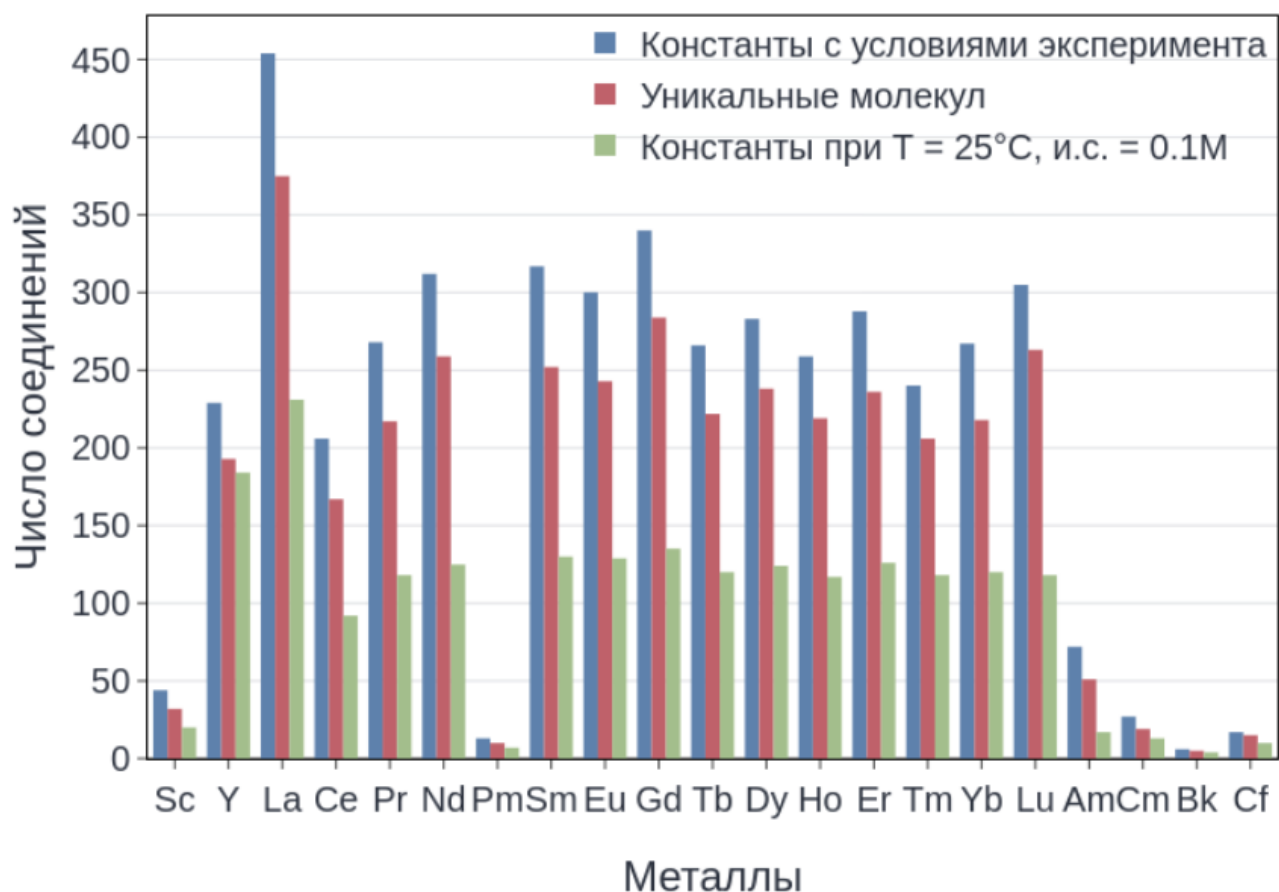


Рис. 2.3 — Композиция собранной базы данных комплексообразователей f- элементов

атомов в графовой сети — быстрая кодировка следующих признаков: номер элемента, валентность, заряд, число соседей и является ли атом частью ароматической системы.

Как видно из рисунка 2.3, количество доступных данных для разных металлов неоднородно — для большинства металлов из ряда лантанидов прямая тренировка модели машинного обучения возможна, в тренировочном наборе достаточное количество точек для построения модели высокого качества, однако нельзя сказать тоже самое для ряда актинидов. Для преодоления этой проблемы нами были разработаны специальные архитектуры нейронных сетей. Первым вариантом было применение методологии трансферного обучения для повышения качества моделей для актинидов. Эта методология заключается в тренировке двух моделей — донорной и акцепторной. Донорная модель тренируется для предсказания свойства, для которого доступно большое количество данных. После тренировки, часть параметров донорной модели (веса и баасы первых слоев нейронной сети) переносятся на акцепторную модель, которая уже тренируется для целевого свойства. Наилучший результат трансферного

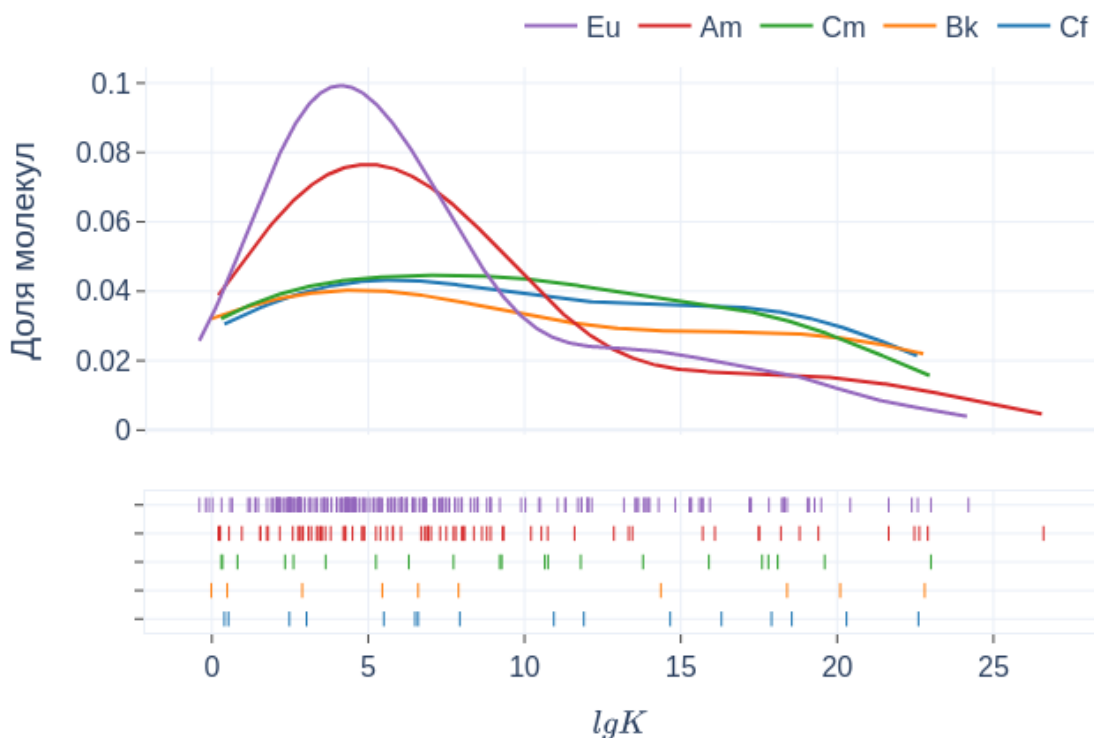


Рис. 2.4 — Распределение констант устойчивости в собранной базе данных для разных металлов

обучения достигается при максимальном сходстве донорного и акцепторного свойства. В этом случае мы передаем акцепторной модели начальную информацию о строении молекулярного графа и все, что остается в процессе тренировки, это подстроится под акцепторное свойство. Схема работы предложенной архитектуры представлена на рисунке 2.5.

Поскольку химические свойства актинилов и лантанидов похожи, то константы устойчивости их комплексов с лигандами также являются довольно сходными. В этой работе мы тренировали модели для констант устойчивости лантанидов и затем проводили перенос параметров на модель для актинилов, чем значительно повышали качество итоговой модели. Обучение моделей проводилось используя 5-кратный скользящий контроль — база данных сначала разбивалась на две части в соотношении 80:20, меньшая часть откладывалась как тестовый набор и не принимала участия в тренировке, а вторая использовалась как тренировочная. Чтобы обеспечить скользящий контроль, уже тренировочная выборка разбивалась на 5 равных частей и тренировка проходила следующим образом — 4 части выбирались в качестве обучающей выборки и оставшаяся одна часть оставалась как контрольная, модель тренировалась, используя контрольную выборку для предотвращения переобучения. Далее эта

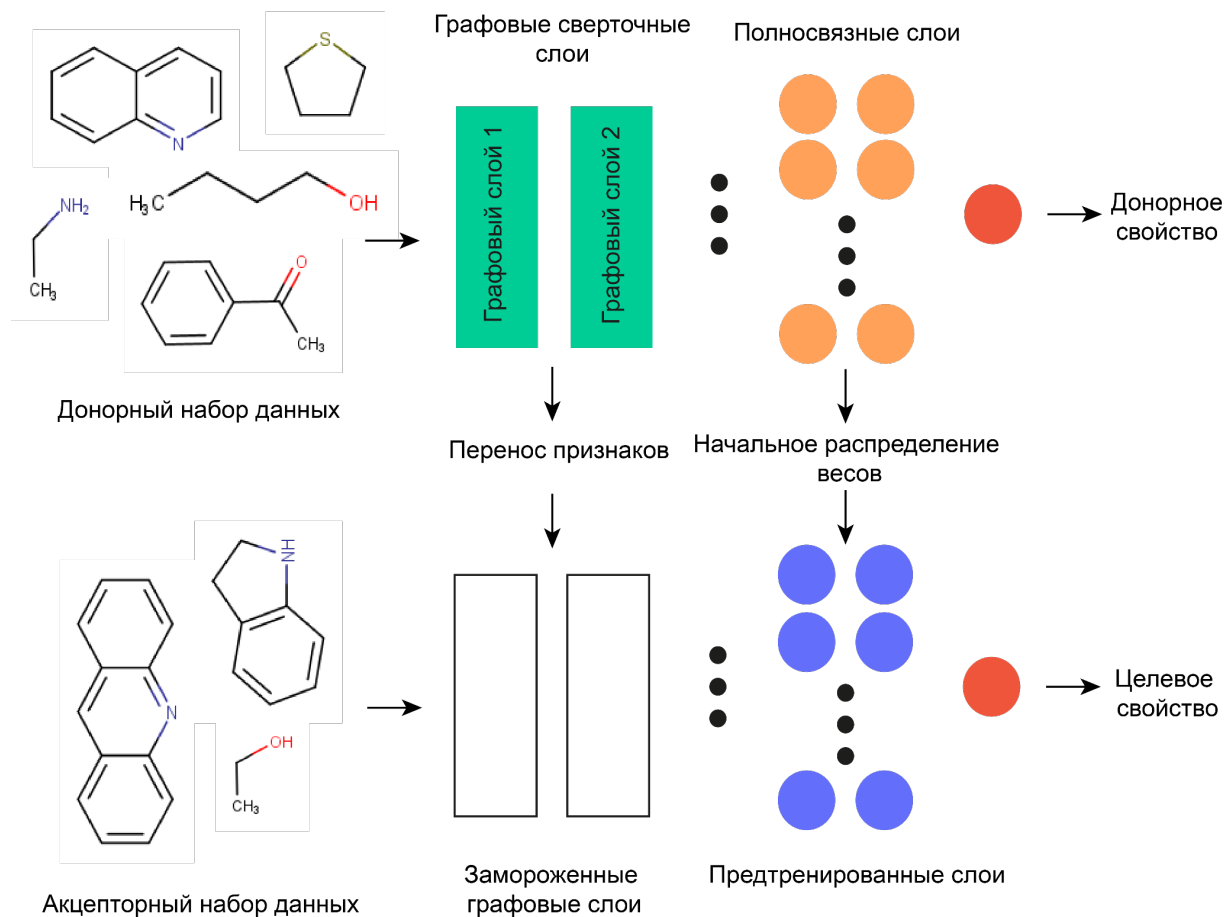


Рис. 2.5 — Схема предложенной архитектуры на основе трансферного обучения

операция проводилась еще 4 раза, с заменой контрольной выборки на одну из обучающих. Таким образом обучались 5 моделей и их консенсусное предсказание уже использовалось как итоговое (Рис. 2.6).

Трансферный подход к обучению позволил получить модели высокого качества. Однако, учитывая композицию собранных данных, и сходство в константах устойчивости, для оптимального использования собранной информации, была разработана еще одна архитектура, которая ценой более сложной структуры и более высокой вычислительной сложности позволила улучшить результаты для некоторых металлов. Особенность предложенной архитектуры заключается в том, что, в отличие от трансферной модели, она принимает на вход два объекта — помимо структуры лиганда на вход также подаётся вектор описывающий металл. Данная структура позволила нам использовать все собранные данные в одной модели и дополнительно учесть условия эксперимента (Рис. 2.7). Также такая структура позволит предсказывать константу устойчивости металла, который в принципе отсутствовал в обучающей выборке, подробнее об этом ниже.

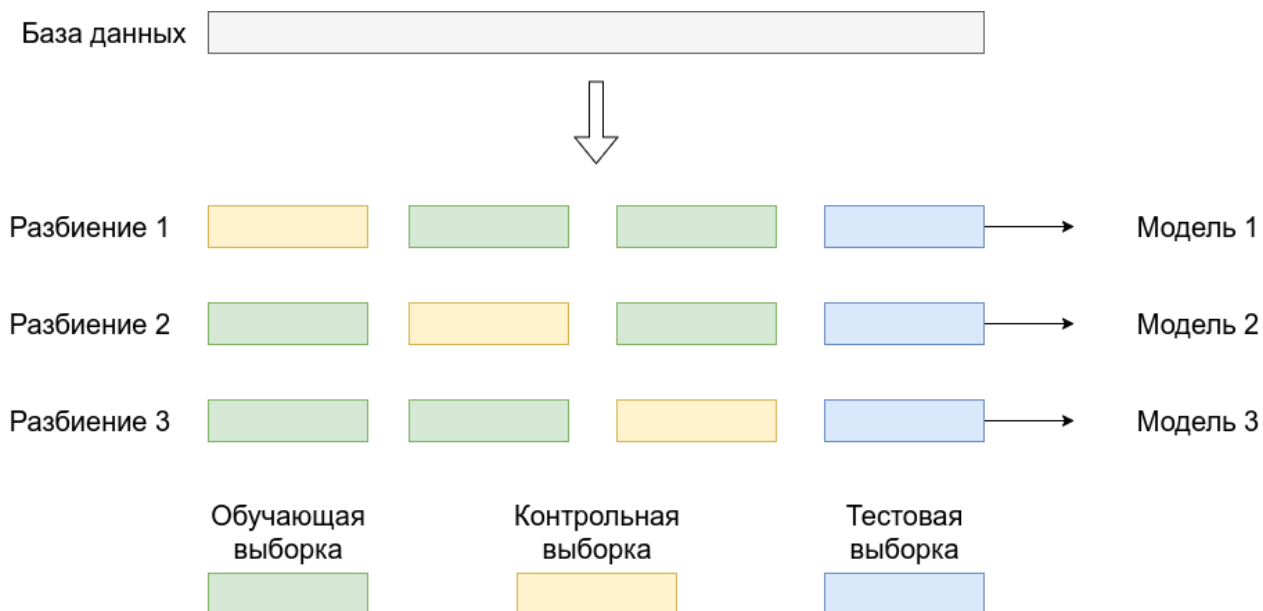


Рис. 2.6 — Схема n-кратного скользящего контроля

Важным элементом нового подхода является способ представления металла, его качество окажет непосредственное влияние на способность модели отличать элементы друг от друга и, соответственно, точно предсказывать константу устойчивости. Для решения этой задачи нами был использован подход *skipatom* [147], в котором методы обработки естественного языка были применены к задаче описания кристаллических структур. Предложенная авторами модель позволяет представить металл в виде вектора фиксированной длины, в котором в абстрактной форме закодированы химические и физические свойства элемента. Единственной проблемой использования этого метода было ограниченное число структур, на которых обучалась предложенная авторами модель. Используемые ими данные не содержали структур с минорными актинидами, соответственно мы не могли получить из неё представления для этих элементов. Для решения этой проблемы мы расширили базу данных, на которой обучался *skipatom*, соединениями с нужными нам атомами и переобучили модель. Данная модификация позволила нам применять этот подход для всех интересующих нас металлов.

Одной из ключевых особенностей новой архитектуры является возможность предсказывать константу устойчивости комплекса даже при условии того, что в обучающей выборке не было целевого металла. Чтобы проверить данную возможность нашей модели мы применили особенный метод тренировки. Для каждого металла строилась отдельная модель, таким образом, что целевой металл

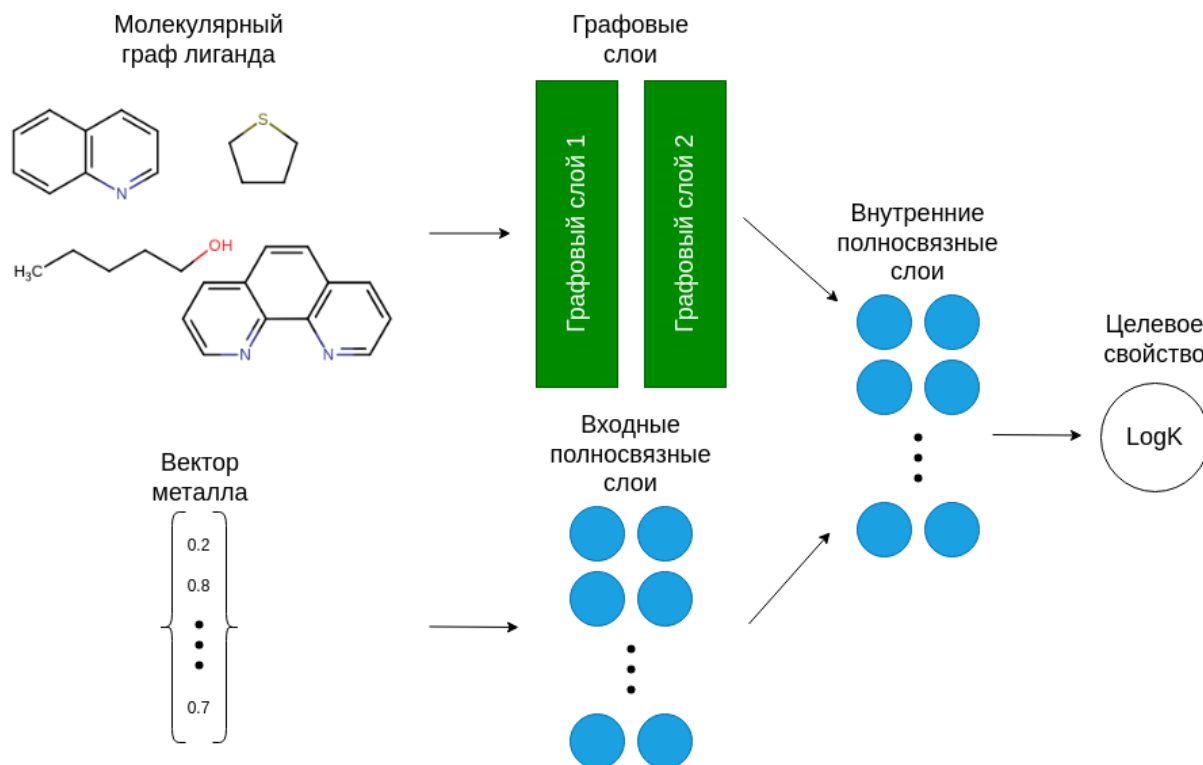


Рис. 2.7 — Схема предложенной гибридной нейронной сети

не принимал участия в тренировке, а участвовал только при оценке полученной модели. Предположим, мы уберем из обучающей базы данные комплексы с Am. Оставшиеся данные будут разбиты на обучающую и контрольную выборку так, чтобы и в обучающей и в контрольной выборке было пропорциональное количество комплексов с каждым металлом. Модель тренируется используя обучающую и контрольную выборку, а финальное тестирование проводится только на комплексах с Am. Таким образом можно проверить, получилось ли обучить модель «понимать» периодический закон и способна ли она предсказывать значения константы комплекса с металлом, которого до этого не видела. Для всех нейросетевых моделей проводилась оптимизация гиперпараметров с помощью баесовского метода с Парценовским оценщиком [148]. При построении нейронных сетей мы использовали библиотеки с открытым исходным кодом Pytorch [149], Scikit-learn [150] и RDKit [132], весь программный код написан на языке Python.

## 2.3 Анализ вкладов молекулярных фрагментов

Ключевым вопросом при построении моделей структура-свойство является последующая интерпретация их предсказаний. Разумеется, возможно использо-

вать модель только как «черный ящик» и ограничиваться только получением значений целевого свойства. Однако существует возможность проанализировать то, какие структурные фрагменты вносят наибольший вклад в свойство, это позволяет не только оценить разумность предсказания модели, но и даёт нам новую информацию о химии моделируемых процессов.

Для анализа вкладов молекулярных фрагментов в модели для жесткости лигандов, мы использовали встроенные возможности программного пакета ISIDA. Линейная модель в основе ISIDA предлагает естественную оценку наиболее важных вкладов в виде весов при соответствующих дескрипторах, а метод выбора подструктурных молекулярных фрагментов помогает уменьшить число дескрипторов для конкретной модели внутри ансамбля до единиц, что заметно облегчает итоговый анализ.

При поиске наиболее важных фрагментов в моделях, предсказывающих константу устойчивости, мы воспользовались подходом `exmol` [151; 152], который основан на описанном выше STONED [64] и использует алгоритмы теории возмущений. Суть метода заключается в построении суррогатной модели для объяснение предсказаний основной модели. Пусть  $f$  — оригинальная модель, а  $g$  — суррогатная,  $x$ , в свою очередь, является признаковым описанием молекулы, для которой сделано предсказание. Тогда задача суррогатной будет состоять в том, чтобы подстроить набор немного измененных, относительно  $x$ , признаковых описаний и соответствующих предсказаний от  $f$  таким образом, чтобы предсказания  $g$  были максимально похожи на предсказание  $f$  (Уравнение 2.1).

$$\xi(\vec{x}) = \arg \min_{g \in G} \zeta(f, g, w) + \Omega(g) \quad (2.1)$$

где суррогатная модель  $g$  минимизирует ошибку  $L$ ,  $w$  обозначает близость между оригинальным признаковым описанием и его вариантами, а  $\Omega$  — опциональная регуляризация. В качестве суррогатной модели используется линейная модель.

## 2.4 Методы глобальной оптимизации

Для дизайна соединения с заданными свойствами необходимо проводить интеллектуальный поиск по химическому пространству с параллельной оценкой целевых параметров. Для поиска как правило используются методы глобальной

оптимизации, которые позволяют находить наилучшие решения при заданной задаче в многомерных пространствах за ограниченное число измерений. Для исследования возможности этих методов мы начали с решения задачи поиска равновесной геометрии органических соединений. Для этого был проведен сравнительный анализ пяти методов глобальной оптимизации геометрии молекул: basin hopping (BH) [153] и minima hopping (MH) [154], генетического алгоритма (ГА), метода оценки Tree Parzen Estimator (Parzen) [155] как представителя байесовской оптимизации и дистанционно-геометрического метода ETKDG (DG) [156; 157]. Тестирование проводилось на специально собранной базе данных из 61 небольшой органической молекулы, для которых известна эталонная геометрия глобального минимума, установленная экспериментальными методами газовой электронной дифракции или микроволновой спектроскопии (Рис. 2.8).

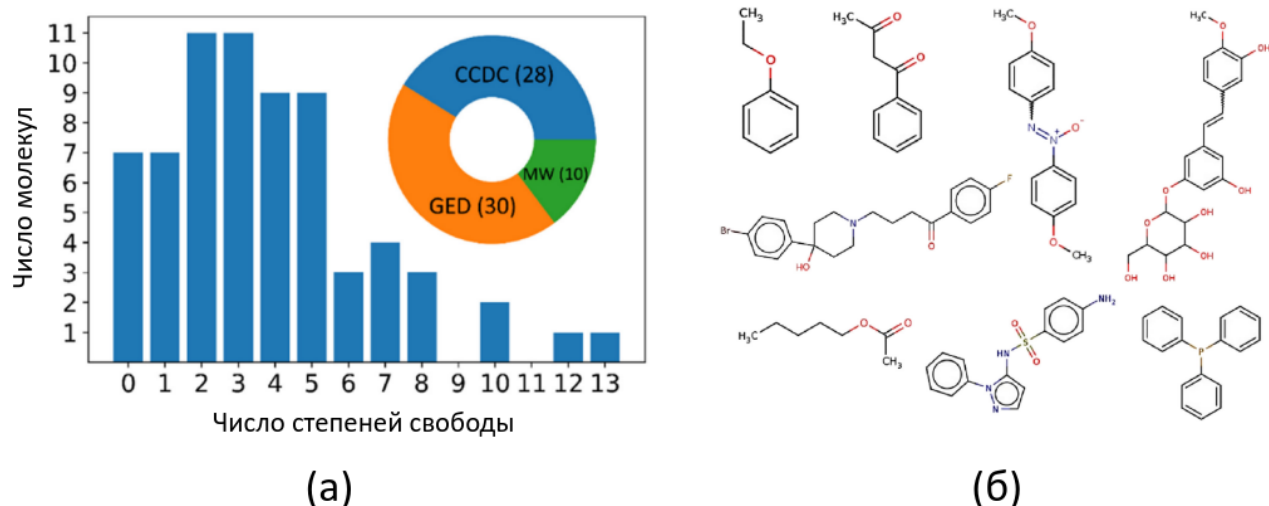


Рис. 2.8 — База данных для тестирования алгоритма глобальной оптимизации, (а) - распределение по числу степеней свободы, (б) - примеры молекул в базе данных

Для квантово-химических расчетов энергии использовался полуэмпирический метод GFN2-xTB, обеспечивающий оптимальный баланс между точностью и вычислительной стоимостью. Перед началом оптимизации для каждого метода были подобраны оптимальные гиперпараметры. Ключевой особенностью методики было представление молекулы в виде вектора торсионных углов, что сужало пространство поиска. Все алгоритмы выполняли фиксированное число итераций (500), а их эффективность оценивалась по двум основным метрикам: среднеквадратичному отклонению (RMSD) полученной геометрии от эталонной и разности энергий (dE), с последующей нормализацией этих значений на число степеней свободы.

## 2.5 Построение генеративной модели

Для построения генеративной модели мы применили вышеописанный подход JANUS [158] с дополнительными модификациями. Схема работы данного подхода приведена на Рис. 2.9, оригинальная модель была дополнена фильтрами синтетической доступности [159] и алгоритмом поиска глобального минимума для дальнейшего тестирования.

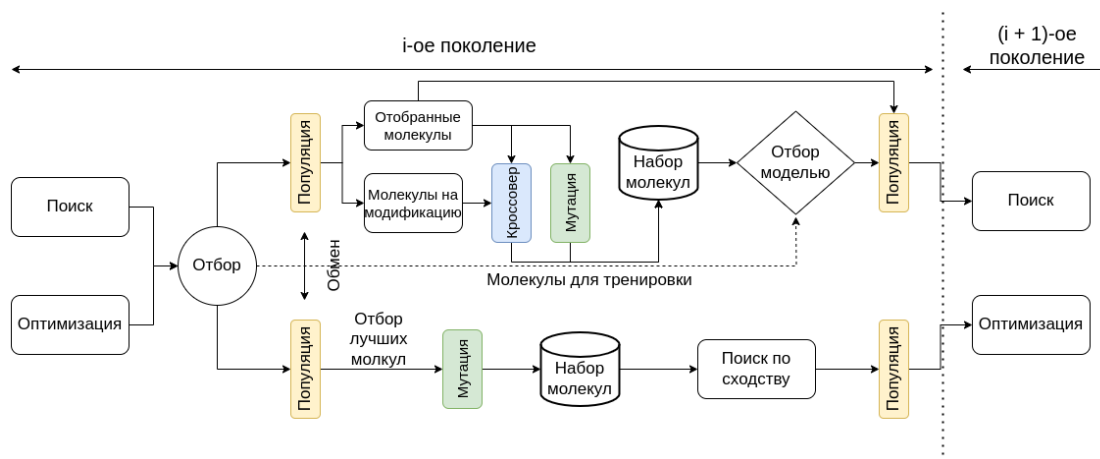


Рис. 2.9 — Схема работы генеративной модели

Работа генеративной модели является лишь одним из этапов дизайна нового химического соединения. Как правило, после получения набора перспективных структур, их дополнительно исследуют, сначала различными вычислительными методами (например методами квантовой химии), а наиболее подходящие соединения синтезируют и тестируют экспериментально. Для успешного исследования соединения методами квантовой химии, необходимо знать глобальный минимум этого соединения. Подавляющее большинство экстрагентов обладают высокой конформационной лабильностью, так, например, эффект аномального арильного упрочнения связывают с энергией предорганизации при образовании комплекса. Поиск глобального минимума для таких соединений является нетривиальной задачей и требует особого подхода, который мы создали и протестировали в рамках работы с методами глобальной оптимизации.

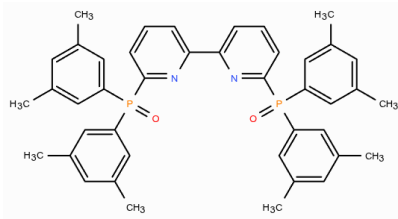
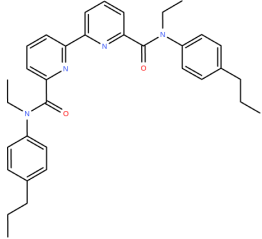
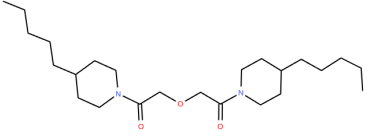
## Глава 3. Экспериментальная часть

Поскольку литературные данные, на которых мы тестировали полученные модели не содержат информации о экстракционных свойствах лигандов, мы решили провести серию экстракционных экспериментов и сравнить полученные экспериментально значения факторов разделения с отношениями констант устойчивости, предсказанными с помощью модели.

### 3.1 Реагенты и материалы

Исследованные в данной работе экстрагенты представлены в таблице 1. Соединения были синтезированы в лаборатории органического синтеза химического факультета МГУ.

Таблица 3.1 — Структурные формулы, названия и условные обозначения экстрагентов

Структурная формула	Название	Обозначение
	6,6'-Бис[бис(3,5-диметилфенил)фосфиноил]-2,2'-бипиридин	L1
	N,N'-Диэтил-N,N'-бис(4-пропилфенил)-2,2'-бипиридин-3,6'-дикарбоксамид	L2
	2,2'-Оксибис [1-(4-пентилпиперидин-1-ил)этан-1-он]	L3

В качестве растворителя использовался мета-нитробензотрифторид (F-3) (ОАО «ПИМ-ИНВЕСТ»). Для приготовления водных растворов использовали деиони-

зованную воду (MILLIPORE Simplicity), удельное сопротивление 18 МОм·см. В работе использовалась концентрированная азотная кислота (х.ч., Химмед). В работе использовались радионуклиды  $^{243}\text{Am}$ ,  $^{152}\text{Eu}$ ,  $^{244}\text{Cm}$ .

## 3.2 Жидкостная экстракция

Для определения максимальной растворимости лигандов в F-3 брали навеску диамида и добавляли растворитель до тех пор, пока диамид не растворится. После этого определяли объем раствора и рассчитывали концентрацию. В случае плохой растворимости, интенсивно перемешивали с помощью вортекс-шейкера.

### 3.2.1 Проведение экстракционных экспериментов

Для проведения экстракционных экспериментов в полипропиленовый Эппендорф объемом 1,5 мл помещали 0,5 мл органической фазы и 0,5 мл водной фазы, и интенсивно перемешивали с помощью вортекс-шейкера в термостате при постоянной температуре  $25 \pm 1^\circ\text{C}$  в течение 15 минут. В работах было показано, что этого времени достаточно для установления экстракционного равновесия [160; 161]. После этого для ускорения расслаивания образцы центрифугировали (5 минут, 9000 об/мин) и отбирали аликвоты по 0,4 мл для дальнейшего анализа. Все экстракционные эксперименты были воспроизведены трижды. При значениях коэффициентов распределения  $0,01 < D < 200$  относительная погрешность не превышала 15%.

### 3.2.2 Методы определения содержания металлов

#### Гамма-спектрометрия

Содержание  $^{243}\text{Am}$  ( $E_\gamma = 59,5$  кэВ) и  $^{152}\text{Eu}$  ( $E_\gamma = 121,1$  кэВ) определяли гамма-спектрометрическим методом с использованием высокочистого германиевого детектора GR 3020 (Canberra Ind.). Коэффициенты распределения рассчитывали, как отношение скорости счета в органической фазе к скорости счета в водной фазе. Содержание радионуклидов в исходной водной фазе составляло 1500-2500 Бк/мл.

## Альфа-спектрометрия

Содержание  $^{243}\text{Am}$  ( $E_\alpha = 5438,2$  кэВ),  $^{244}\text{Cm}$  ( $E_\alpha = 5901,6$  кэВ) в исходной водной фазе и в равновесной водной фазе после экстракции определяли с помощью альфа-спектрометрии с использованием альфа-спектрометра модели 7401 с кремниевым детектором (Canberra Ind.). Поскольку для альфа-спектрометрии важно наличие тонкослойного образца, предварительная пробоподготовка включала следующие стадии:

1. в отобранную аликвоту водной фазы (0,4 мл) добавляли 100 мкл 0,1 М раствора церия и 3 мл концентрированной фтороводородной кислоты; полученную смесь встряхивали и оставляли на 30 минут для осаждения фторидов;
2. подготавливали прибор для фильтрования (фильтр Resolve 0,1 мкм, диаметр 25 мм);
3. для активации фильтра пропускали несколько капель этилового спирта, после чего промывали 2-3 мл деионизированной воды;
4. пропускали через фильтр ранее подготовленный образец;
5. промывали образец 2-3 мл деионизированной воды и 1-2 мл этилового спирта;
6. с помощью пинцета извлекали фильтр и помещали на чашку Петри, приклеивая двусторонним скотчем;
7. полученный образец сушили на воздухе, после чего помещали в камеру альфа-спектрометра и измеряли скорость счета.

Коэффициенты распределения рассчитывали по следующей формуле:

$$D = \frac{I_0 - I}{I} \quad (3.1)$$

, где  $I_0$  – скорость счета исходной водной фазы,  $I$  – скорость счета равновесной водной фазы после экстракции. Содержание радионуклидов в исходной водной фазе составляло 1000-2000 Бк/мл. Фактор селективности рассчитывали как отношение коэффициентов распределения.

## Глава 4. Результаты и обсуждение

### 4.1 Моделирование свойств лигандов и комплексов

#### 4.1.1 Расчет жесткости лигандов

Был построен ряд моделей машинного обучения, используя разные подходы и методы для прогнозирования жесткости лигандов и определения константы устойчивости комплексов металл-лиганд.

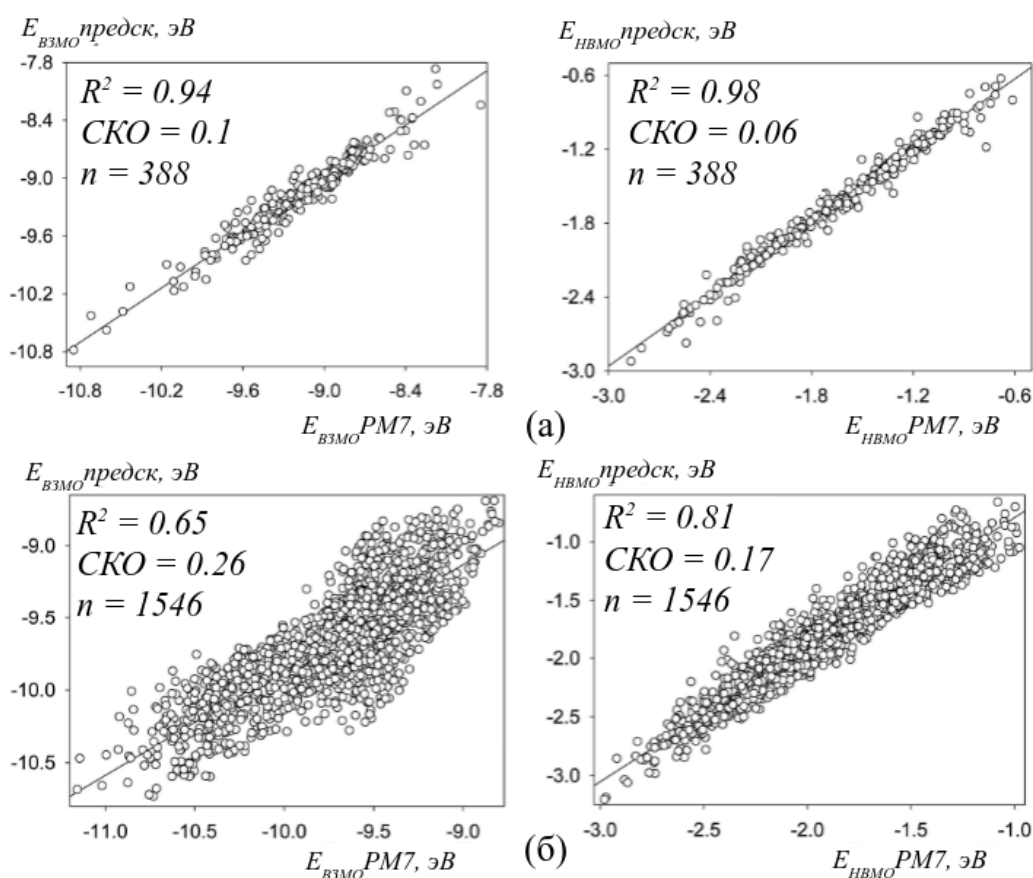


Рис. 4.1 — Параметры качества моделей для расчета энергий граничных орбиталей на (а) внутренней контрольной и (б) внешней тестовой выборке

Для определения жесткости лигандов были построены ансамблевые модели по предсказанию энергий граничных орбиталей методом множественной линейной регрессии (Рис. 4.1).

Несмотря на относительно неплохое качество моделей, при дальнейшем анализе мы обнаружили явное несоответствие между предсказанием и уже извест-

ными нам данными. При исследовании вкладов наиболее важных фрагментов мы увидели следующую картину (Рис. 4.2).

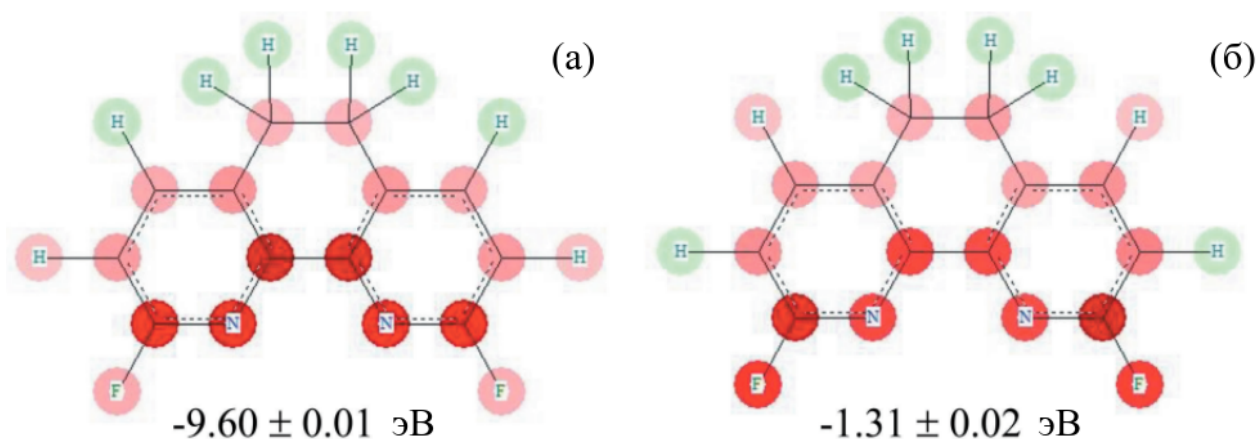


Рис. 4.2 — Вклады атомов в значения (а) ВЗМО и (б) НСМО, красным отмечены фрагменты, вносящие положительный вклад в энергии граничных орбиталей, зеленым - отрицательный.

Из изображения видно, что исходя из фрагментов наибольший вклад и в ВЗМО и НСМО вносят одни и те же атомы. Этот факт входит в противоречие и с полученными нами и с известными из литературы квантово-химическими оценками распределения ВЗМО и НСМО.

Чтобы проверить действительно ли жесткость лиганда является определяющим параметром для его комплексообразования, мы проанализировали корреляцию между константами устойчивости комплексов из собранных нами данных со значениями жесткости этих лигандов (Рис. 4.3).

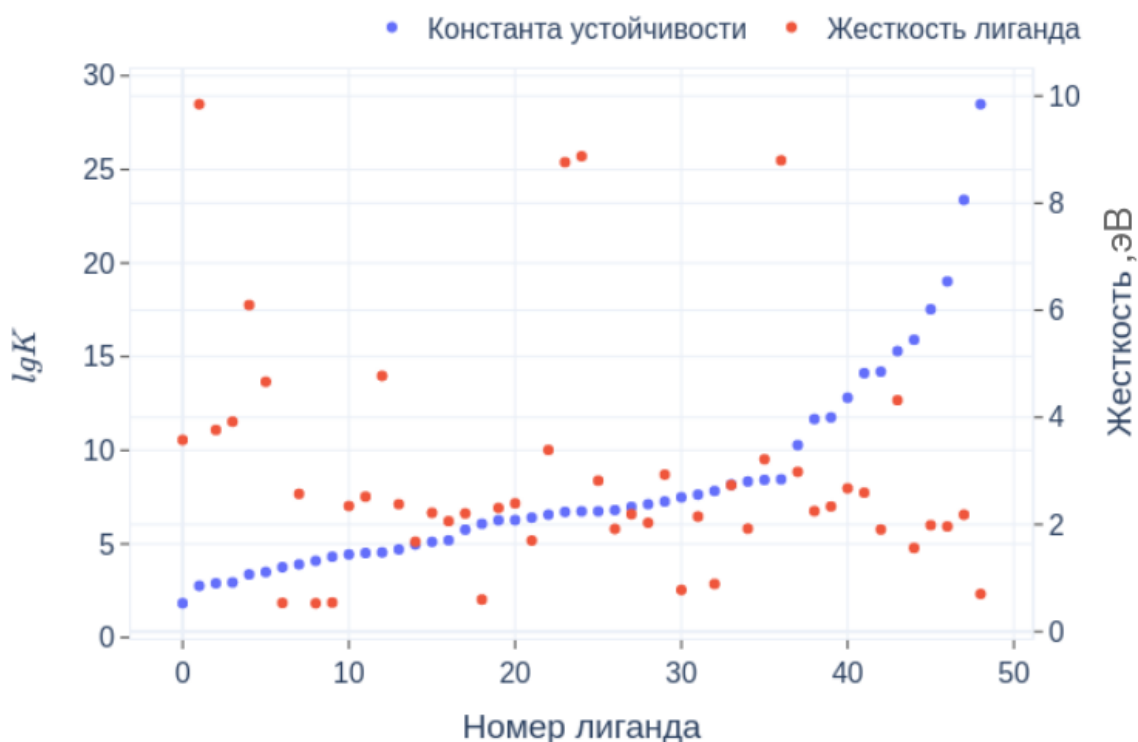


Рис. 4.3 — Взаимосвязь между жесткостью лигандов и константами устойчивости их комплексов с Eu

Жесткость лиганда определяли как полуразность энергий граничных орбиталей, рассчитанных с помощью метода DFT (B3LYP, 6-31G<sup>\*</sup>). Из полученных результатов видно, что корреляции между этими величинами не прослеживается, что ставит под вопрос целесообразность использования жесткости в качестве основного дескриптора комплексообразования. В литературе этот вопрос неоднократно обсуждался, так в работах [94; 96; 97], авторы исследовали вклады 5-f оболочек в граничные орбитали и обсуждали большую ковалентность связей в комплексах с актинидами, чем в комплексах с лантанидами. В обзоре [162] также упоминалось, что более «мягкие» атомы актинидов должны проявлять к лигандам с «мягкими» донорными центрами большее сродство чем лантаниды. Созданные на основе этой концепции высокоселективные лиганды эту гипотезу подтверждают. В тоже время, в уже обсуждавшейся работе [96] авторы привели уменьшение жесткости лиганда как одну из причин повышения селективности, что входит в противоречие с полученными нами результатами.

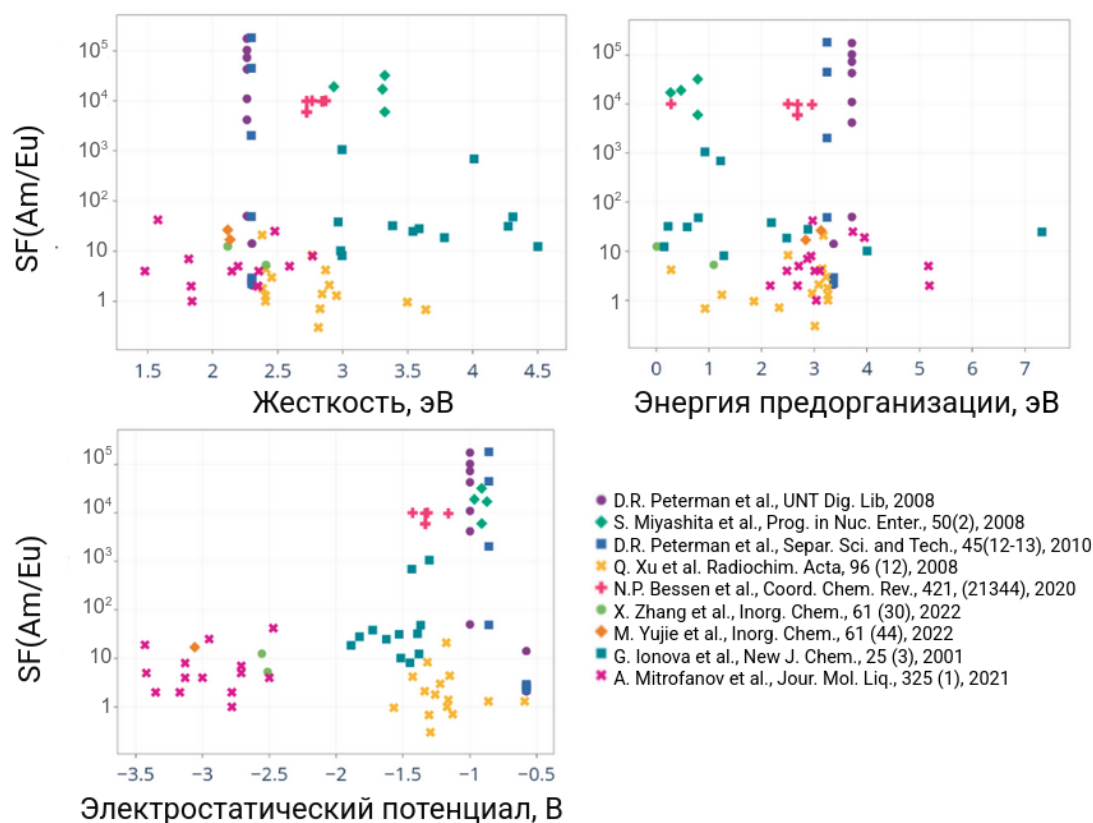


Рис. 4.4 — Диаграммы рассеяния между фактором разделения и wybranными дескрипторами

Для дальнейшего анализа связи между свойствами лиганда и его пригодности на роль экстрагента актининов и лантанинов мы рассмотрели корреляции между такими параметрами лиганда как жесткость, энергия предорганизации и электростатический потенциал в точке координации с фактором разделения Am(III)/Eu(III). Так, на рисунке 4.4 представлены зависимости экспериментально полученного фактора разделения от рассчитанных методами квантовой химии свойств лиганда. Из полученных результатов можно сделать вывод, что ни одно из этих популярных в литературе свойств не подходит как основной дескриптор селективности экстрагента. При этом корреляция отсутствует как для базы данных в целом, что ожидаемо, поскольку условия в которых проводилась экстракция отличаются от работы к работе, так и для каждой работы в отдельности (Таблица 4.1).

И хотя в отдельности выбранные параметры не могут быть основными дескрипторами селективности лиганда, возможно их сочетание с учетом параметров используемых экстракционных систем, может быть использовано для создания численной модели селективности. Чтобы проверить эту гипотезу, мы

Таблица 4.1 — Коэффициенты корреляции ( $\rho$ ) между фактором селективности и выбранными дескрипторами для отдельных работ

Источник	Жесткость, эВ	Энергия предорганизации, эВ	Электростатический потенциал, В
A. Mitrofanov et al.[144]	-0,2	0,05	0,26
D. R. Peterman et al.[137]	0,14	0,66	-0,18
G. Ionova et al. et al.[143]	-0,16	-0,25	0,43
N. P. Bessen et al.[140]	0,6	-0,33	0,09
Q. Xu et al.[139]	-0,34	0,16	-0,05
S. Miyashita et al.[138]	-0,03	0,08	-0,07

выбрали величины рассчитанных дескрипторов как признаковое описание лиганда, а такие свойства как концентрация лиганда, концентрация высаливателя, концентрация азотной кислоты, а также диэлектрическую проницаемость и коэффициент поверхностного натяжения органического растворителя для описания экстракционной системы. Полученный вектор мы использовали для построения линейной регрессионной модели фактора разделения. Обычная линейная модель показала неудовлетворительный результат ( $R^2 = 0.35$ ), что совпадает с нашим предположением о более сложной взаимосвязи селективности и выбранных дескрипторов. Для того чтобы учесть возможные нелинейные вклады дескрипторов в селективность мы построили более сложную модель используя алгоритм CatBoost. Полученная модель показала более высокое качество (Рис. 4.5 (а)), что подтверждает нашу гипотезу о нелинейной взаимосвязи рассматриваемых свойств. Теперь имея на руках надежную модель, связывающую свойства лиганда и параметры экстракционной системы с фактором разделения, можно проанализировать вклады различных дескрипторов модели на целевое свойство (Рис. 4.5 (б)).

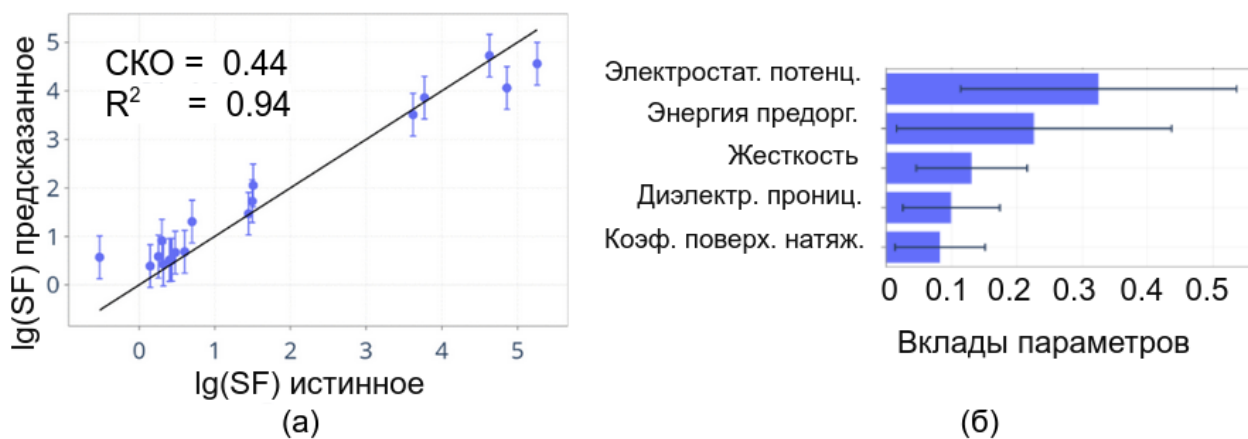


Рис. 4.5 — а - Качество CatBoost регрессии, б - вклады наиболее важных параметров в SF

Из результатов анализа наиболее важных свойств можно сделать несколько выводов. Во-первых, вклады свойств лиганда в селективность превосходят вклады вносимые свойствами растворителя, отсюда можно сделать вывод, что первичный дизайн комплексообразователя можно проводить в одной системе, пренебрегая влиянием растворителя. Также видно, что из выбранных дескрипторов наибольший вклад вносит именно электростатический потенциал в точке координации, что можно использовать в будущем при формировании признакового описания для более сложных моделей машинного обучения. Вопрос вызывает большая погрешность для вклада энергии предорганизации в селективность лиганда, но это можно объяснить тем, что в обучающей выборке было недостаточно конформационно жестких соединений, из-за чего влияние этого дескриптора было недостаточно хорошо учтено. Еще один важный вывод, который можно сделать из этого этапа работы, это малое количество данных доступное нам для исследований. Радиохимические исследования, ввиду ограниченного списка возможных приложений, а также ограничений, обусловленных мерами безопасности, немногочисленные по сравнению с другими областями химии, а следовательно и количество доступных нам данных для анализа существенно ограничено. Таким образом, для получения предсказательных моделей, пригодных для оценки свойств широкого круга соединений, нам необходимы инструменты чтобы компенсировать этот недостаток.

#### 4.1.2 Апробация методологии трансферного обучения на примере физико-химических и биомедицинских свойств

Для моделирования констант устойчивости металл лиганд, нами предварительно была проведена оптимизация гиперпараметров модели. Так, для трансферной модели результаты оптимизации представлены на Рис. 4.6. Мы подбирали оптимальное число и размеры графовых и полносвязных слоев сети.

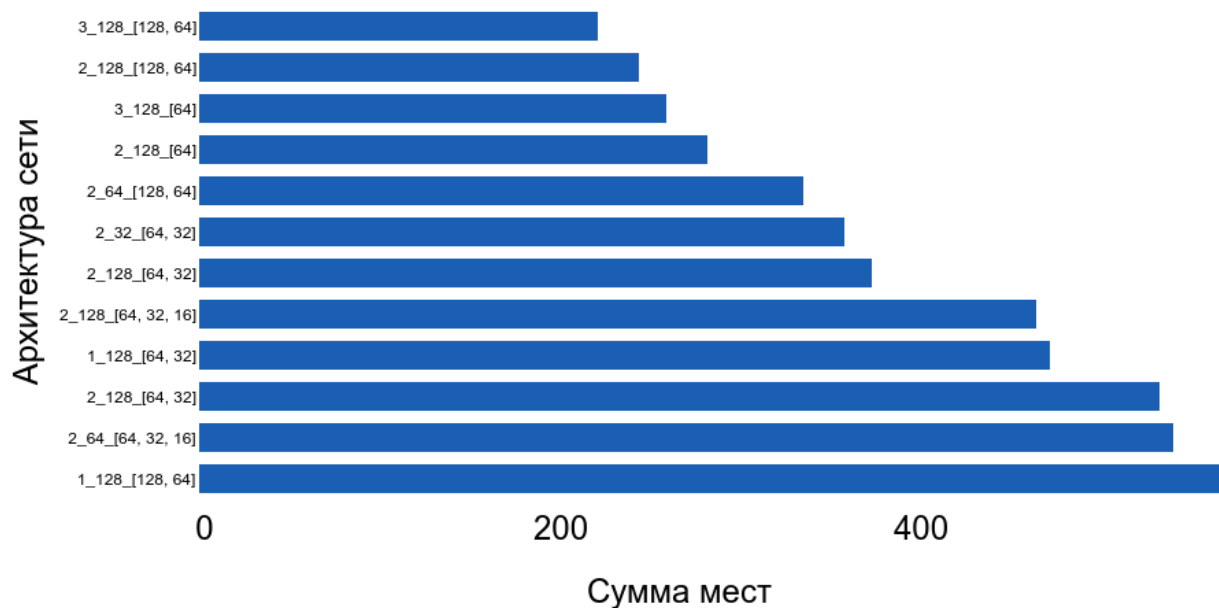


Рис. 4.6 — Результаты оптимизации структуры модели (число графовых слоев, размер графовых слоев, композиция полносвязной части модели) отсортированные по сумме мест, которые занимали указанные архитектуры в общем списке всех построенных тестовых моделей

Наиболее удачные архитектуры оценивались с помощью сумм мест, которым соответствовали архитектуры моделей построенных в процессе оптимизации. Для этого было построено множество моделей варьируя более десятка параметров, все построенные модели были отсортированы в соответствии с финальной ошибкой и каждой композиции графовых и сверточных слоев было присвоено число, соответствующее сумме мест, которые занимала каждая конкретная архитектура. Из результатов видно, что наилучшее качество показали модели с 3 полносвязными слоями размером в 128 нейронов, и двумя графовыми слоями с размерами 128 и 64 нейрона.

После проведения оптимизации мы дополнительно протестировали надежность метода на распространённых физико-химических и биомедицинских свойствах, качество полученных моделей представлено на рисунке 4.7. Для этого бы-

ла натренирована донорная модель для липофильности (База данных ChEMBL 1.6 миллиона соединений) и на ее основе построены акцепторные модели. Из результатов видно, что применение трансферной модели значительно улучшает качество модели, и чем меньше размер обучающей выборки, тем более заметен этот эффект. Так, для обучающей выборки в 10-20 соединений, применение трансферного обучения позволяет получить увеличение метрики качества на 0.5, что является разницей между нетренированной моделью и моделью, подходящей для полуколичественной оценки.

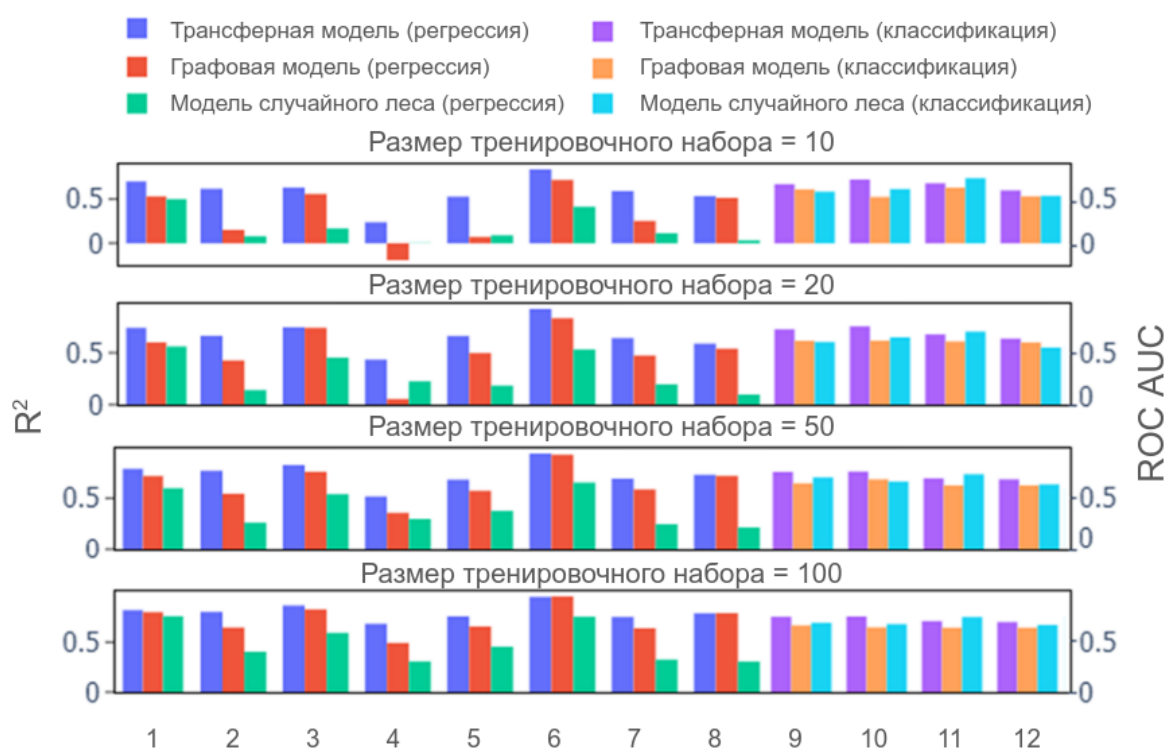


Рис. 4.7 — Качество моделей для набора физико-химических и биологических свойств. (1 — растворимость в воде, 2 — фактор биоконцентрации, 3 — фактор биоразложения, 4 — константа Генри, 5 — фактор биотрансформации в рыбе, 6 — коэффициент разделения октанол-вода, 7 — коэффициент поглощения почвой, 8 — давление насыщенных паров; далее ингибиторы изоморфов цитохрома P450, 9 — U 1851 2c19, 10 — U 1851 2c9, 11 — U 1851 3a4, 12 — U 1851 2d6)

Как мы убедились, сходство между донорным и акцепторным свойством играет решающую роль при тренировке трансферной модели. Следующим интересующим нас вопросом было то, влияет ли сходство молекулярных графов между донорным и акцепторным набором на эффективность. Чтобы проверить эту гипотезу мы натренировали ряд моделей для пары  $\log P/\log S$ , с разным размером обучающей выборки. Но сами обучающие выборки донора подбирали та-

ким образом, чтобы они в разной степени пересекались с составом акцепторной обучающей выборки. Таким образом мы хотели проследить влияние сходства составов наборов на итоговую модель. Результаты представлены на Рис. 4.8. Слева изображены обучающие выборки для Chembl logP и OPERA logS, визуализированные с помощью метода понижения размерности TSNE. Этот метод сохраняет расстояния между объектами, так что опираясь на данные TSNE мы смогли выделить 3 обучающих набора для донорной модели так, чтобы один из них совсем не был похож на акцепторный, другой бы имел какую-то степень пересечения, и еще один должен иметь очень схожие с акцептором соединения. Также мы ввели еще один случайный донорный набор. Из полученных результатов хорошо видно, что композиция обучающих выборок не влияет на итоговое качество трансферной модели. Таким образом, сходство свойств — единственное необходимое условие успешного применения такой модели.

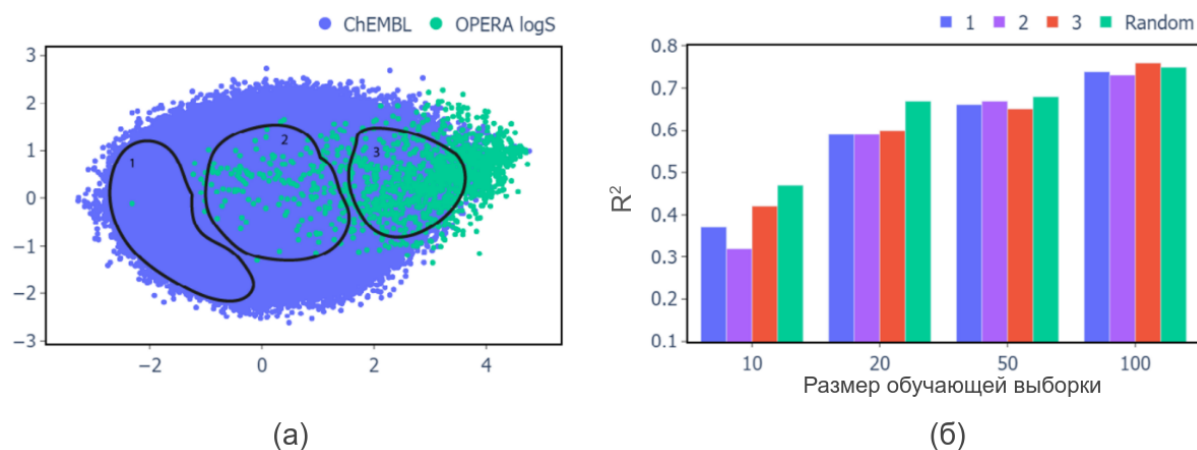


Рис. 4.8 — (а) t-SNE разложение донорного и акцепторного базисных наборов и выделенные области для выбора обучающих выборок, (б) качество получившихся трансферных моделей

Проанализировав полученные результаты, мы пришли к выводу о пригодности этого метода для построения моделей машинного обучения с целью предсказания констант устойчивости комплексов.

### 4.1.3 Построение трансферных моделей для предсказания констант устойчивости комплексов

Поскольку метод показал высокое качество на физико-химических и биомедицинских свойствах молекул, мы применили его для моделирования констант устойчивости комплексов "металл-лиганд". Близость химических свойств лантанидов и актинидов в этом случае дополнительно способствует моделирова-

нию константы устойчивости, поскольку чем ближе донорное и акцепторное свойство, тем выше качество акцепторной модели. Важным предварительным шагом при применении трансферного подхода является подбор оптимальной донорной модели. Для этого мы построили ряд моделей для лантанидов (Рис. 4.9). Важно понимать, что помимо того, что свойства моделей должны быть похожи, качество донорной модели должно быть настолько высоким, насколько это возможно. На Рис. 4.9 видно, что модели для части металлов (Ho, Er) показали качество заметно ниже чем для остальных.

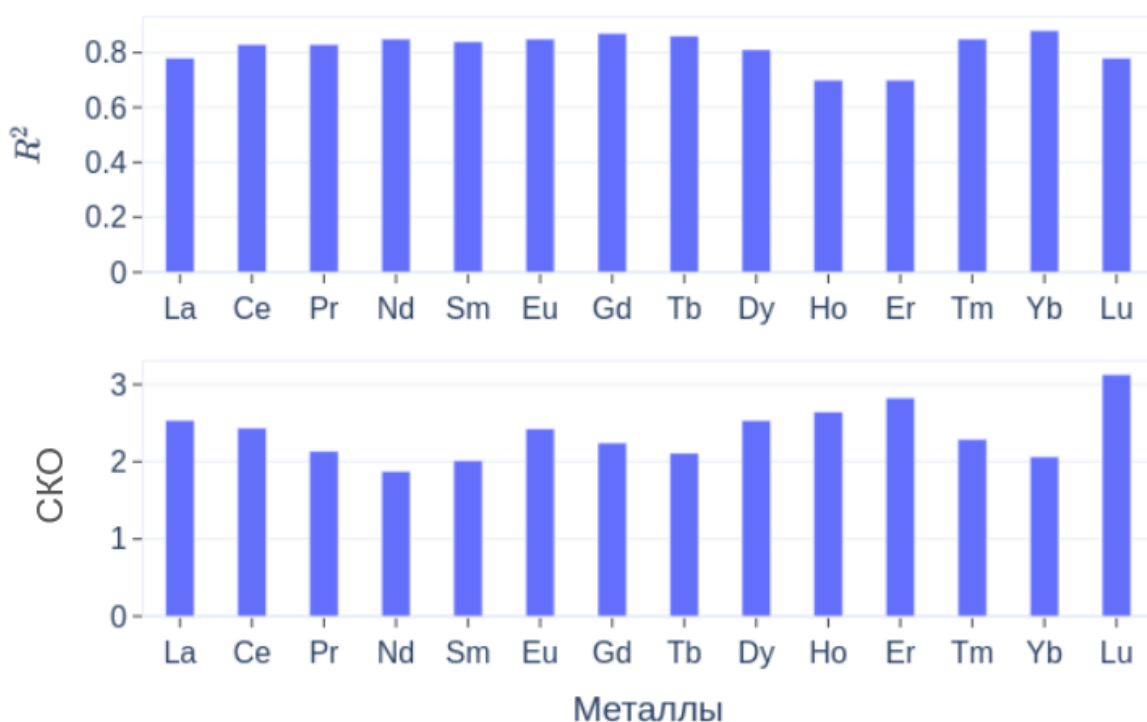


Рис. 4.9 — Метрики качества трансферной моделей для лантанидов

Также ошибка модели для Lu оказалась довольно большой в сравнении с другими. Анализ тренировочных наборов и хода обучения этих моделей не привел к однозначным выводам о причинах худшего качества для этих моделей, возможно экспериментальные данные для этих металлов оказались получены с большей ошибкой. Независимо от причин, из-за низкой точности, выбор этих моделей на роль донора нецелесообразен. Далее мы нашли пересечения между базами данных для актинидов и лантанидов и построили корреляции изменения констант устойчивости этих металлов для пересекающихся молекул (Рис. 4.10).

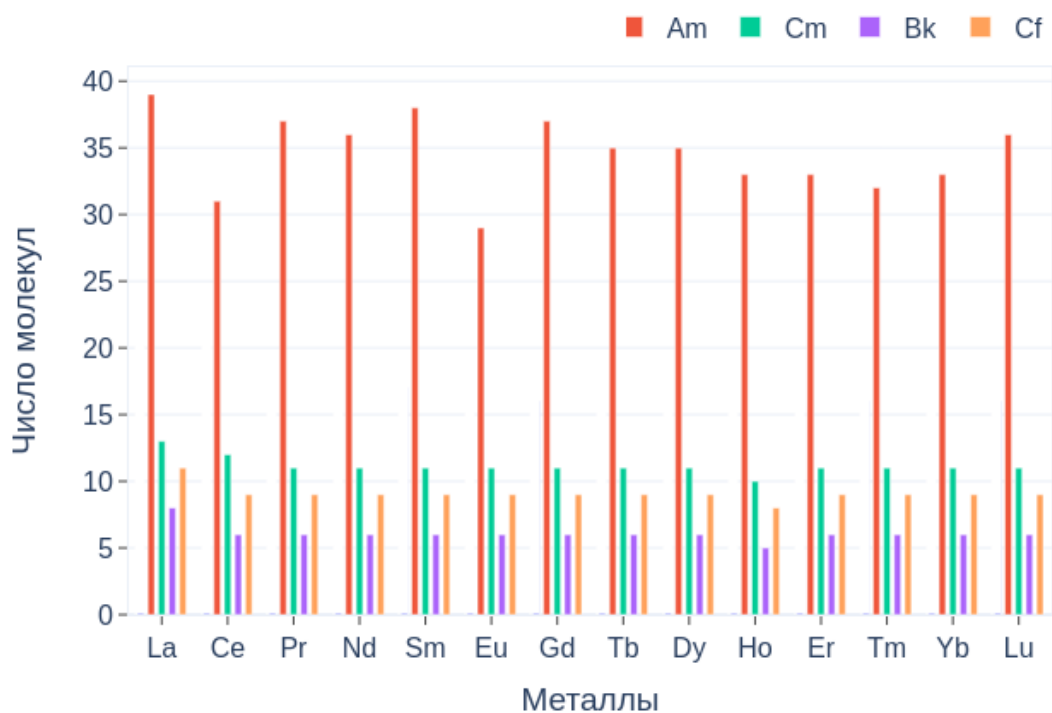


Рис. 4.10 — Пересечения между наборами данных актинилов и лантанидов

Из полученных результатов можно сделать несколько наблюдений. Первое, это более сильный разброс в константах устойчивости при сравнении лантана и актинилов. Это достаточно легко объяснить — лантан обладает несколькими отличающимися от остальных лантанидов свойствами и в целом проявляет себя скорее как d-элемент. Следующая особенность, характерная для всех полученных зависимостей, это меньшая корреляция с константами устойчивости для тория - разброс значений для него значительно выше чем для других актинилов. Это связано прежде всего с иной наиболее устойчивой степенью окисления тория (+4) по сравнению с остальными рассматриваемыми металлами, которые имеют степень окисления (+3). Также на некоторых графиках заметны отчетливые выбросы, что не совсем характерно для рассматриваемых объектов, особенно выделяются два выброса на графике 4.11. Молекулы — выбросы выделены на графиках, а их структуры и значения констант представлены в Таблице 4.2.

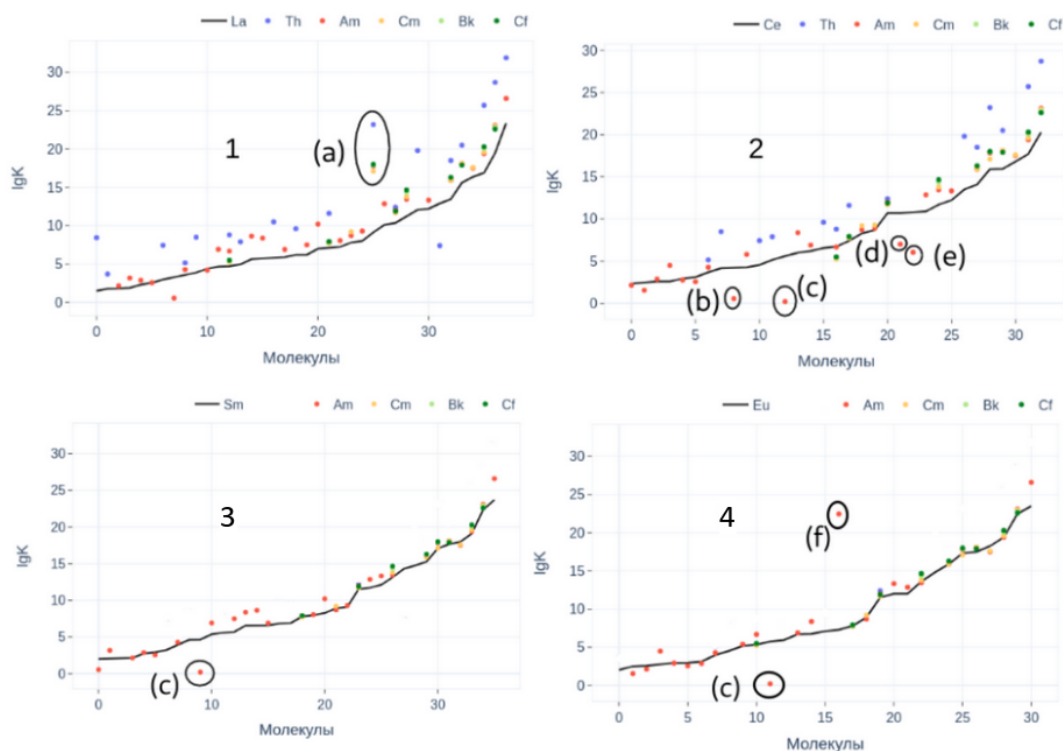
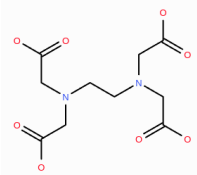
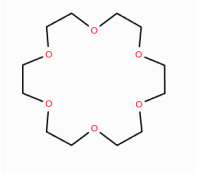
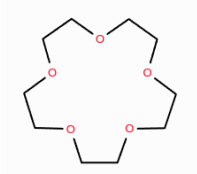
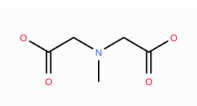
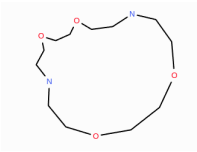
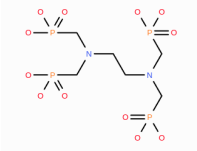


Рис. 4.11 — Зависимости констант устойчивости для разных лигандов между лантанидами (1) La(III), (2) Ce(III), (3) Sm(III), (4) Eu(III) и актинидами (Th(IV), Am(III), Cm(III), Bk(III), Cf(III)) с выделенными выбросами (a, b, c, d, e, f).

Некоторые из выбросов вполне объяснимы, так EDTA (a) — популярный комплексообразователь и не только для разделения ОЯТ, показывает большие отличия в константе только для La(III) и Th(IV), чья непохожесть на остальные металлы уже обсуждалась. Более интересной является ситуация с краунэфирами (b, c, e), константа устойчивости которых с америцием является аномально низкой. В литературе известны работы, в которых подобные соединения рассматриваются как потенциальные экстрагенты [84; 146; 163—165] и эти соединения показывают высокую селективность, в том числе и для разделения Am(III)/Ln(III). В частности, такая картина хорошо соотносится с разностью в жесткости актинидов и лантанидов, представленные краунэфиры, обладающие только «жесткими» -O донорными центрами лучше связывают именно лантаниды. Что касается лиганда (d), наличие в нем мягкого координационного центра делает разницу в константах устойчивости не такой большой. Отличие величин констант устойчивости лиганда (f), с большим количеством жестких координационных кислородов, легко объяснить разной степенью окисления Ce и Th.

При рассмотрении оставшихся лигандов становится понятно, что разница в

Таблица 4.2 — Соединения с выбросами в значениях констант устойчивости

Индекс	Структура	La	Ce	Sm	Eu	Th	Am	Cm	Bk	Cf
a		9.13	15.9	17.1	17.3	23.2	17.8	17.1	17.9	18.0
b		3.29	4.43	2.05	-	-	0.55	-	-	-
c		5.60	4.65	5.73	-	0.21	-	-	-	-
d		-	10.7	-	-	-	7.0	-	-	-
e		-	10.2	-	-	-	6.1	-	-	-
e		-	-	7.3	-	22.5	-	-	-	-

разбросе не особо велика и значения констант устойчивости, в целом, достаточно хорошо между собой коррелируют. Соответственно наиболее важным параметром остается качество донорной модели. Следующим шагом мы провели серию тренировок для определения констант устойчивости актинидов. Для этой цели, используя донорную модель для липофильности, были натренированы модели для лантанидов и уже с них, как с донорных моделей, были построены модели для актинидов (Рис. 4.12). Как видно, применение трансферного обучения значительно улучшило качество моделей для актинидов. Мы также проанализировали то, как обучаются трансферные модели. Так как в роли до-

норного набора данных для каждого актинида выступал Eu (поскольку модель на Eu обладала хорошим качеством), было интересно, сможем ли мы увидеть сходство между металлами донорного и акцепторного наборов по тому, как шла тренировка модели. Для этого мы вынесли на график 4.13 то, как изменяется функция ошибки в процессе тренировки для Am(III), Cm(III) и Th(IV), результат подтвердил изначальную задумку. Как видно из построенных кривых обучения для выбранных металлов, характер обучения сильно отличается.

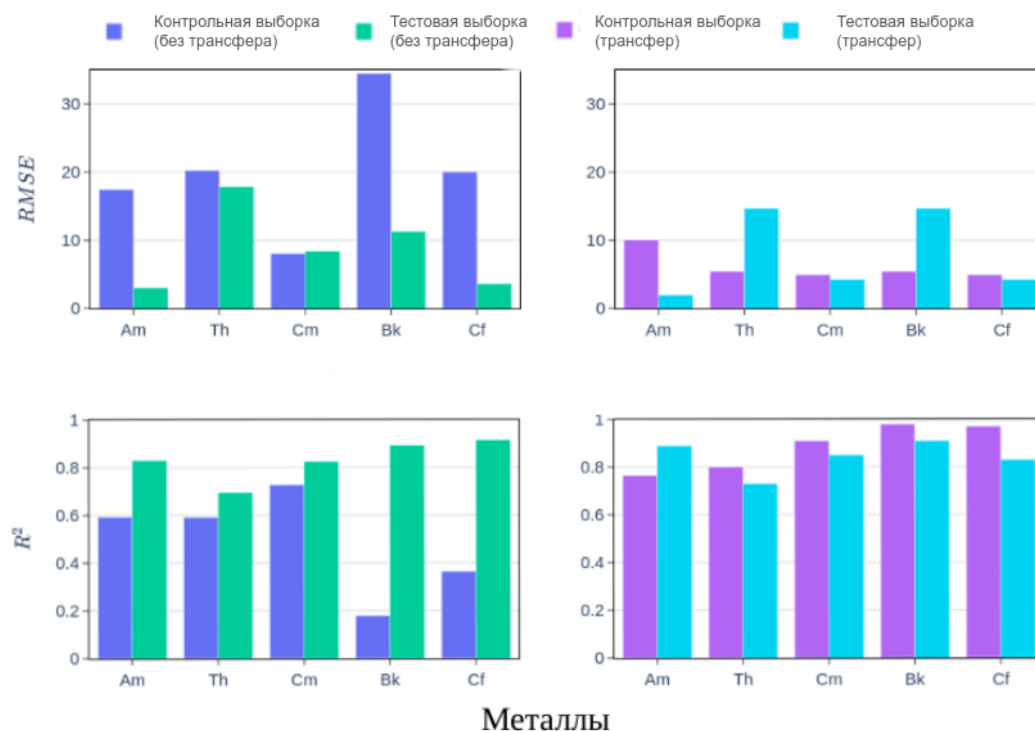


Рис. 4.12 — Качество трансферных моделей актинидов

Быстрее всего обучилась модель на Am(III), хватило нескольких тренировочных шагов, чтобы получить модель с наименьшей ошибкой. При этом тренировка моделей для Cm(III) и Th(IV) заняла на один порядок и на два порядка больше времени, соответственно. И эти результаты легко объяснить — то как долго обучалась трансферная модель, зависит от того, насколько сильно необходимо изменить параметры модели для предсказания акцепторного свойства. Следовательно время обучения напрямую связано с тем, насколько сильно отличаются между собой свойства донорной и акцепторной модели. Таким образом, можно заключить, что европий имеет большее сходство с америцием, чем с кюрием, а свойства кюрия ближе к европию чем свойства тория. Мы дополнительно проанализировали то, какой вклад использование трансферного

обучения делает в область применимости модели. Для оценки этого вклада мы применили следующий метод — мы выбрали тренировочный набор данных и натренировали на нем модель с трансфером и без и записали величины ошибок этих моделей на тестовой выборке.

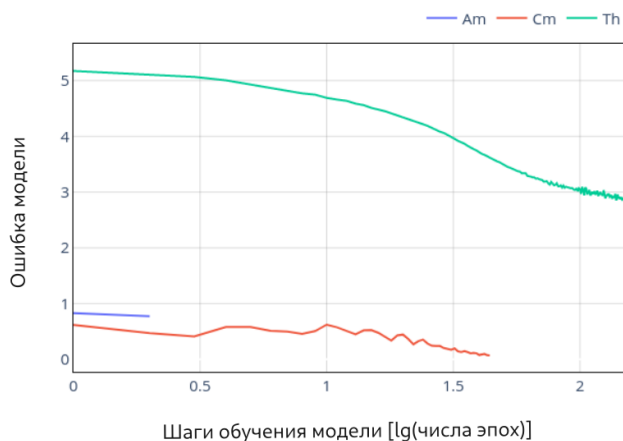


Рис. 4.13 — Кривые обучения моделей для разных акцепторных наборов

Далее, чтобы определить эффективное увеличение обучающей выборки от использования трансфера, мы стали по очереди добавлять к тренировочному набору по одной случайной молекуле и тренировать модель без использования трансферного обучения (для каждого добавления тренировалось несколько моделей). И в тот момент, когда ошибка модели без трансфера, но на увеличившемся тренировочном наборе, сравнивается с ошибкой трансферной модели, мы достигли эффективного увеличения выборки, которое дает использования трансферного обучения. На Рис. 4.14 представлена такая зависимость для пары Am(III)/Eu(III), при изначальном размере ацепторного тренировочного набора в 20 соединений. Как видно, модель на расширенном наборе начинает превосходить трансферную модель только при добавлении 8 дополнительных соединений в обучающую выборку. Можно сделать вывод, что трансферное обучение эффективно увеличивает размер обучающей выборки в 1.5 раза, соразмерно с ним также растет и область применимости модели. Можно заключить, что применение технологии трансферного обучения позволило строить предсказательные модели на наборах данных, которые до этого считались непригодными по причине малого объёма.

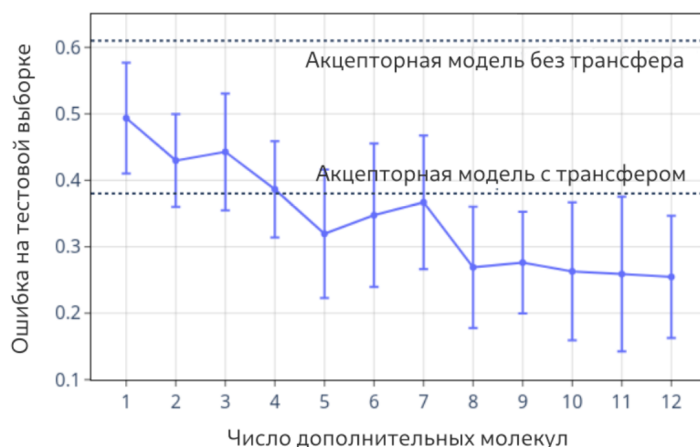


Рис. 4.14 — Оценка эффективного увеличения тренировочной выборки от использования трансферного обучения

Построенные модели для минорных (Th(IV), Cm(III), Bk(III), Cf(III)) актинидов получены впервые и обладают достаточной точностью для качественной оценки лиганда на роль комплексообразователя f-элементов. В тоже время, структура трансферного обучения имеет свои недостатки. Как было показано и как известно из литературы, свойства актинидов и лантанидов в целом довольно похожи друг на друга, соответственно любая модель для предсказания константы устойчивости комплекса с лантанидом, может применяться как донорная модель для аналогичной модели с актинидом. Однако в представленном варианте модель не способна утилизировать всю информацию из собранных нами данных, поскольку при обучении используются только два набора молекул — донорная и акцепторная выборки, все другие данные о комплексообразовании с другими металлами не используются. Это уменьшает потенциальное качество модели, поскольку в оставшихся данных есть дополнительная информация. Во-первых, так как выборки для разных металлов не полностью пересекаются, там присутствуют дополнительные структуры лигандов для обучения, что потенциально расширяет область применимости модели. Во-вторых, использование данных о связи уже известных лигандов с другими металлами, при учете свойств металла, способно дать модели больше информации о комплексообразовании как таковом, что дополнительно увеличивает потенциальную точность модели. Соответственно необходимо создать такую архитектуру нейронной сети, которая смогла бы учесть эту дополнительную информацию.

#### 4.1.4 Построение гибридных моделей для предсказания констант устойчивости комплексов

Следующим этапом было проведение тренировок на основе гибридной модели, структура которой описана в пункте 4.1., но предварительно было необходимо провести оптимизацию гиперпараметров этой сети. Для этого база данных была разделена на обучающую и контрольную выборки таким образом, что данные для актинидов, не применялись для обучения, а использовались только для оценки. При этом ошибка для каждого из актинидов считалась отдельно, и наибольшая из них, на каждом шагу, была критерием отбора. В качестве алгоритма оптимизации был выбран байесовский процесс с Парценовским оценщиком. Для поиска наиболее важных параметров мы применили PCA [166] разложение результатов оптимизации и взяли коэффициент первого компонента как меру важности гиперпараметра (Рис. 4.15).

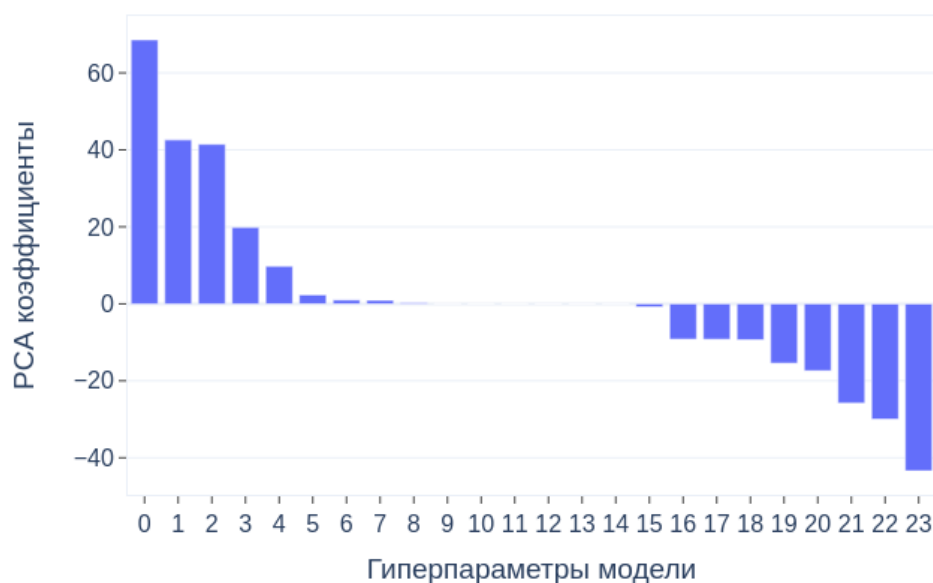


Рис. 4.15 — Важность гиперпараметров модели по результатам оптимизации. Размер третьего скрытого слоя (0), размер первого скрытого слоя (1), размер четвертого скрытого слоя (2), размер третьего сверточного слоя (3), размер четвертого сверточного слоя (4), размер пятого сверточного слоя (5), размер шестого слоя металла (6), число скрытых слоев (7), функция активации сверточного слоя (8), исключение в слое металла (9), исключение в полносвязном слое (10), алгоритм объединения (11), функция подвыборки (12), исключение в сверточном слое (13), число сверточных слоев (14), число полносвязных слоев в блоке металла (15), размер первого полносвязного слоя, описывающего металл (16), размер второго слоя металла (17), размер первого сверточного слоя (18), размер четвертого слоя металла (19), размер второго сверточного слоя (20), размер пятого слоя металла (21) Размер третьего слоя металла (22), размер второго скрытого слоя (23)

Как видно, наиболее важными являются именно размерные характеристики модели, соответственно их анализ и стал следующим шагом (Рис. 4.16).

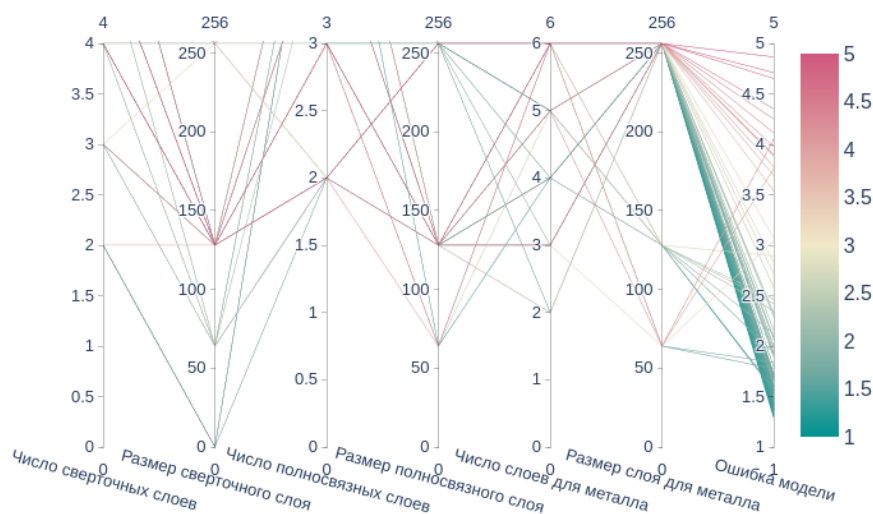


Рис. 4.16 — Влияние архитектуры модели на финальную ошибку, конкретные модели представлены линиями, а цветом показана ошибка модели на контрольной выборке

Из представленных на графике результатов можно сделать несколько выводов. Первой особенностью результатов является, с одной стороны широкий разброс, а с другой стороны высокое качество итоговых моделей для самого большого слоя металла. Очевидно, что моделей с этим значением параметра также построено больше, т.е в процессе оптимизации алгоритм отдавал предпочтение именно этому параметру. Такой результат может означать, что вектор металла потенциально содержит больше информации, чем текущая архитектура модели способна обработать. Это может быть полезно при дальнейшем расширении применения этой архитектуры для моделирования констант устойчивости с металлами, которые не относятся к f- элементам. Аналогичная картина наблюдается для размера полносвязного слоя, большая часть успешных моделей имеют большое количество нейронов на внутреннем слое, возможно, частичная причина этому заключается в предыдущем вопросе — неспособность модели полностью реализовать информацию о металле. При этом важно упомянуть, что у оригинального подхода skipatom есть несколько вариантов размеров латентного пространства и в нашей работы мы используем вектор длиной 200 — самый большой из доступных. Все попытки применить меньшие варианты сталкивались с ухудшением качества. Возможно, повторная перетренировка skipatom, но уже для увеличения латентного пространства, позволит тренировать модели для других металлов, при адекватной адаптации архитектуры под

новый вектор металла.

Следующим шагом была проведена непосредственно тренировка модели для предсказания константы устойчивости. Тут мы применили описанный выше подход по проверке модели на комплексах металлов, которые не присутствовали в обучающей выборке (Рис. 4.17).

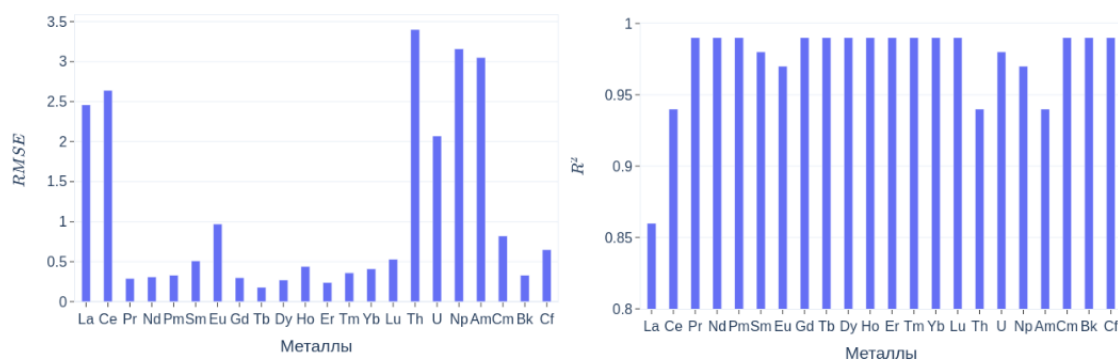


Рис. 4.17 — Качество гибридной модели при исключении тестового металла из обучающей выборки

Результаты показывают, что предложенная модель действительно способна предсказывать константы устойчивости не имея предварительно информации о тестовом металле, качество предсказаний при этом превосходит все остальные модели, представленные в литературе. Рассмотрим получившиеся модели, первое, что можно отметить, это более высокое качество гибридных моделей по сравнению с трансферными. Это подтверждает высказанные ранее предположения о том, что дополнительная информация о комплексообразовании с другими металлами позволит улучшить итоговую точность предсказания. Вторым необычным результатом является меньшая, по сравнению с другими лантанидами, точность моделей для лантана и церия. Получившиеся модели все равно превосходят трансферные в качестве, но при этом, для трансферных моделей такой большой разницы в качестве не наблюдалось. Это можно объяснить разницей в свойствах между лантаном и остальными f- элементами, также набор данных для лантана шире и разнообразнее, что в условиях ограниченной области применимости, также влияет на результат. Аналогичная картина наблюдается для тория, и причины разной точности также заключаются в отличии свойств тория и других рассматриваемых металлов. Заслуживает внимание высокое качество моделей для кюрия, берклия и калифорния, ошибки этих моделей не сильно отличаются от таковых у моделей для лантанидов. И хотя этот

результат является положительным, и предложенные модели можно будет использовать для оценки свойств комплексов с этими металлами, вероятнее всего ошибки при сравнении с реальными данными будут выше. Причина такой аномально высокой точности, особенно в сравнении с америцием, заключается в очень ограниченном наборе данных, на которых рассчитывались метрики качества. Наборы данных для этих металлов ограничены как по количеству точек, так и по разнообразию лигандов, которые там представлены. Также из графика 4.10 можно увидеть, что наборы лигандов для этих металлов, в отличие от Am(III) и Cm(III), полностью пересекаются с наборами для лантанидов, что делает предсказания констант для этих лигандов более простой задачей.

И хотя модель показывает хорошие результаты, нам необходим способ определять, для каких молекул предсказаниям модели можно доверять, а на какие область ее применимости уже не распространяется. Для того, чтобы найти метод, который был бы способен определять, входит ли молекула в область применимости модели, мы протестировали несколько популярных в литературе способов и сравнили их качество для наших задач. Суть всех трех методов заключается в расчете неопределенности модели и в вынесении вердикта на основе этой неопределенности. Мы сравнили 3 метода — ансамблевый, в котором неопределенность оценивается по разнице в предсказании между моделями ансамбля, в качестве ансамбля мы использовали модели, тренирующиеся в рамках 10-кратного скользящего контроля. Следующий метод заключается в поиске расстояния между выходом одного из скрытых слоев модели для тестируемой молекулы с такими же выходами для молекул тренировочного набора. По расстоянию между ними мы можем определить входит ли эта молекула в область применимости. И последний метод, который мы тестировали заключается в расчете определенности используя измененный выход модели. Архитектура модели меняется таким образом, что ее выходной слой формируется из двух нейронов вместо одного. На выходе этих двух нейронов мы моделируем среднее значение целевого свойства и его отклонение. По величине этого отклонения мы рассчитываем неопределенность модели. Качество каждого из методов оценивалось с помощью площади перекалибровки (Рис. 4.18). Как видно из результатов, метод с оценкой отклонения показал себя наилучшим образом и используя измененную архитектуру мы можем оценивать, доверять ли предсказанию модели для конкретного комплексообразователя или нет. В целом результат подтвер-

дид литературные данные, ансамблевый подход склонен переоценивать область применимости модели, а мера расстояния в латентном пространстве сильно зависит от конкретной архитектуры нейронной сети, возможно для успешного применения этого метода требуется оптимизировать композицию слоев модели.

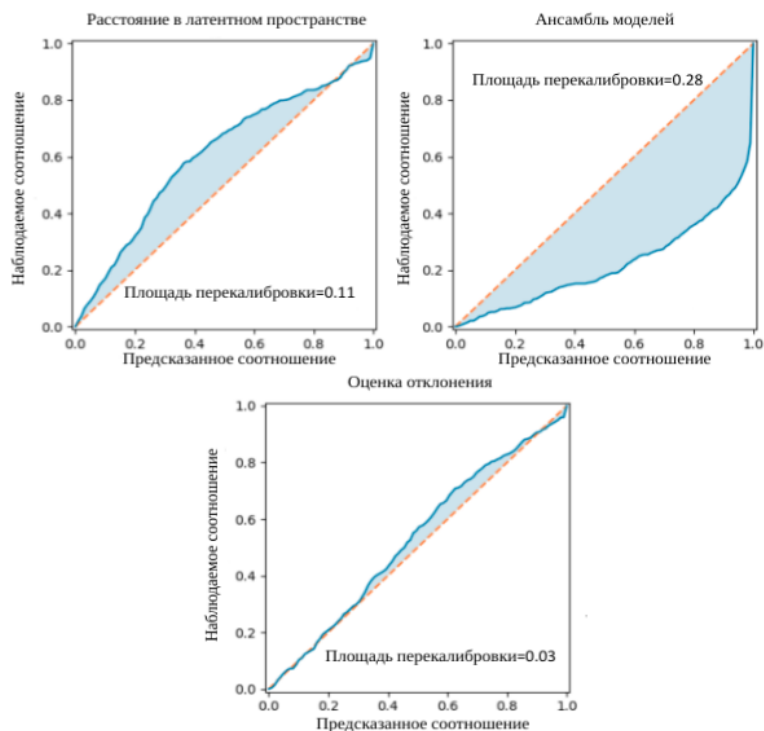


Рис. 4.18 — Выбор метода оценки области применимости, используя метод неопределенности

Хорошо видно, что гибридная модель превосходит по качеству трансферную, это может быть связано с тем, что архитектура гибридной модели позволяет использовать больше данных, что благоприятно сказывается на качестве. Еще одним преимуществом гибридной архитектуры является то, что предсказание констант устойчивости для всех металлов лежат в одной области применимости относительно структуры лиганда. С целью оценить эту область применимости, нами был выбран набор молекул из базы данных ZINC [167], в которой выложены коммерчески доступные соединения, и с помощью описанного выше метода, мы рассчитали величины неопределенности при предсказании молекул из этой выборки. В качестве пороговой величины неопределенности мы использовали среднее значение для тестовой выборки целевого металла, притом для каждого металла проверялась отдельная модель, при тренировке которой данные о целевом металле не использовались. Такой подход был применён, чтобы найти

оценку снизу для области применимости, для полной модели эта область будет шире. Для рассматриваемых в этой работе актинидов, доля молекул, входящих в область применимости варьировалась от 57 до 63%. Учитывая, что это оценка снизу, можно считать, что модель покрывает значительную часть химического пространства.

Мы также проанализировали ключевые фрагменты молекул, вносящие наибольший вклад в константу устойчивости. Алгоритм по которому проходил анализ состоял из нескольких шагов.

- Сначала мы рассчитали ключевые фрагменты для каждой молекулы тренировочного набора для исследуемых металлов. Все металлы рассматривались отдельно друг от друга.
- Следующим шагом данные о вкладах всех фрагментов были собраны и для каждого фрагмента были рассчитаны среднее значение и стандартное отклонение его вклада в каждую молекулу, где он присутствовал.
- Далее мы отбрасывали те фрагменты, для которых стандартное отклонение превосходило по модулю среднее значение, а также редкие фрагменты, встречавшиеся менее чем в 3 соединениях обучающей выборки.

По итогам анализа нами был получен набор наиболее важных фрагментов, а также вклады в константу устойчивости, которые вносили эти фрагменты (Таблица 4.3).

Таблица 4.3 — Наиболее важные фрагменты при анализе результатов для моделей Eu(III), Am(III), Cm(III)

Структура фрагменты	Вклад в lgK
	Eu(III)
	$0.67 \pm 0.06$

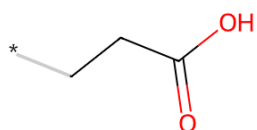
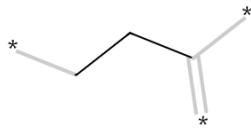
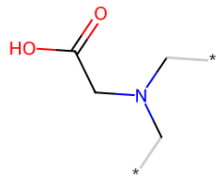
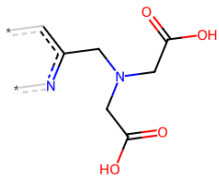
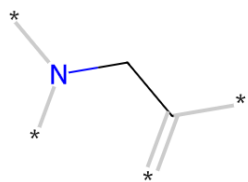
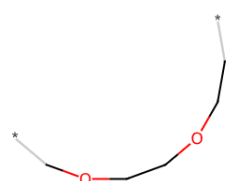
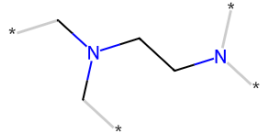
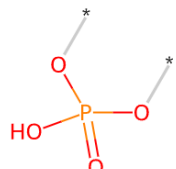


Таблица 4.3 – продолжение

Структура фрагменты	Вклад в lgK
	$-1.16 \pm 0.07$
	$3.02 \pm 0.21$
Am(III)	
	$2.26 \pm 0.60$
	$3.64 \pm 1.90$
	$-1.39 \pm 0.80$
Cm(III)	
	$2.24 \pm 0.92$
	$-0.53 \pm 0.16$

Фрагменты с «жесткими» кислородными координационными центрами при-

сутствуют, как одни из основных фрагментов, для каждого металла, но при этом вносят разный вклад, больший для Eu и меньший для актинидов, что согласуется с теорией Пирсона [129]. Интересный результат связан с уменьшением константы устойчивости в присутствии фрагмента 2 для Eu. Можно предположить, что такой эффект связан с изменением относительного расположения координационных центров. Похожий эффект описывался в работе [168].

#### 4.1.5 Алгоритм поиска равновесной геометрии малых органических молекул

Поскольку в составе большинства генеративных моделей используются методы глобальной оптимизации, то подбор такого метода - важная задача при создании системы de-novo дизайна. Чтобы решить эту задачу мы предварительно провели тестирование разных алгоритмов глобальной оптимизации на более распространенной задаче вычислительной химии - определение равновесной геометрии малых органических молекул. Для этой задачи существует большое количество тестовых данных и отработанных подходов, таким образом нам будет проще оценить пригодность различных методов глобальной оптимизации для химических задач.

Результаты тестирования алгоритмов глобальной оптимизации для поиска равновесной геометрии органических соединений показали разную эффективность рассматриваемых методов. Алгоритм *minima hopping* продемонстрировал наивысшую структурную точность, однако он же оказался одним из наиболее требовательных к вычислительным ресурсам наряду с методом *basin hopping*. Напротив, метод *Tree Parzen Estimator* и дистанционно-геометрический алгоритм *ETKDG* показали наименьшее время расчета и количество обращений к расчету энергии. Генетический алгоритм в данном тесте показал наихудшие результаты как по точности, так и по вычислительной эффективности. Важным выводом стало то, что добавление процедуры локальной оптимизации на каждом шаге или в конце расчета значительно повышало точность методов *Parzen* и *DG*, практически приближая ее к точности МН, но с существенно меньшими затратами. Сравнение данных газовой фазы и кристаллографических данных подтвердило, что последние могут вносить систематическую погрешность из-за влияния кристаллического поля и менее надежны в качестве эталона для задач глобальной оптимизации (Таблица 4.4).

Таблица 4.4 — Метрики эффективности различных алгоритмов глобальной оптимизации для получения равновесной геометрии молекул

Метод		RMSD, Å	dE, кДж/моль	Время расчета	Число расчетов
Minima (MH)	Hopping	Наименьшее	Наименьшее	Наибольшее	Наибольшее
Basin Hopping (BH)		Среднее	Среднее	Высокое	Высокое
Tree Estimator (Parzen)	Parzen	Низкое	Низкое	Наименьшее	Низкое
ETKDG (DG)		Высокое	Высокое	Низкое	Наименьшее
Генетический алгоритм (GA)		Наибольшее	Наибольшее	Среднее	Высокое

Результаты демонстрируют, что как генетический алгоритм, так и методы байесовской оптимизации применимы для задачи глобальной оптимизации геометрии. Ключевым фактором, способствующим эффективной работе байесовского подхода (Tree Parzen Estimator), является возможность компактного представления молекулы в виде вектора торсионных углов. Это представление резко сокращает размерность пространства поиска, что позволяет байесовским методам, которые эффективно исследуют и используют вероятностную модель этого пространства, быстро находить глобальный минимум с оптимальным балансом между точностью и вычислительными затратами. Генетический алгоритм, в свою очередь, также может использовать это представление, однако в рамках данного исследования он показал меньшую эффективность, что может указывать на необходимость более тонкой настройки его операторов для задач конформационного поиска.

Можно сделать вывод, что применение таких методов в химических задачах оправдано и применение такого универсального метода как генетический алгоритм для дальнейшей разработки генеративной модели рационально.

## 4.2 Жидкостная экстракция

Для представленных выше лигандов были определены коэффициенты распределения и факторы разделения пар Am(III)/Eu(III) и Am(III)/Cm(III) при экстракции в F-3 из 3 М HNO<sub>3</sub> (таблицы 4.5, 4.6).

Таблица 4.5 — Коэффициенты распределения Am(III) и Eu(III) при экстракции в F-3 из 3 М HNO<sub>3</sub>.

Лиганд	C(L), М	D(Am(III))	D(Eu(III))	SF(Am(III)/Eu(III))
L1	0,05	5,1	5,8	0,9
L2	0,05	1,7	0,2	8,5
L3	0,01	1,2	6,8	0,2

Таблица 4.6 — Коэффициенты распределения Am(III) и Cm(III) при экстракции в F-3 из 3 М HNO<sub>3</sub>.

Лиганд	C(L), М	D(Am(III))	D(Cm(III))	SF(Am(III)/Cm(III))
L1	0,01	7,0	7,3	1,0
L2	0,05	3,7	1,8	2,1
L3	0,01	1,6	1,8	0,9

#### 4.2.1 Сравнение результатов эксперимента и расчетов

Полученные экспериментальные результаты были использованы для тестирования качества предлагаемых моделей. Для этого были предсказаны константы устойчивости металл-лиганд для соответствующих структур с Eu(III), Am(III) и Cm(III). Фактор разделения оценивали как отношение соответствующих констант. Результаты сравнения представлены в таблице 4.7.

Таблица 4.7 — Коэффициенты распределения Am(III) и Cm(III) при экстракции в F-3 из 3 М HNO<sub>3</sub>.

Лиганд	$lgK_{Eu(III)}$	$lgK_{Am(III)}$	$lgK_{Cm(III)}$	$K_{Am(III)}/K_{Eu(III)}$	$K_{Am(III)}/K_{Cm(III)}$	$SF(Am(III)/Eu(III))$	$SF(Am(III)/Cm(III))$
L1	5,4	5,4	5,4	1,0	1,0	0,9	1,0
L2	5,8	5,9	5,8	1,4	1,4	8,5	2,1
L3	6,5	6,4	6,5	0,8	0,8	0,2	0,9

Как видно из приведенных результатов, модель смогла корректно проранжировать рассматриваемые экстрагенты по их селективности для обеих пар металлов. В то же время, абсолютные значения факторов селективности не совпадают с отношением констант устойчивости, полученными из модели. Это

соответствует принятым в начале работы упрощениям, по которым мы тренируем модель без учета растворителя, используя только значения констант устойчивости, полученные для водной фазы. Отсюда также можно сделать вывод о том, какое значение для разделения конкретной пары металлов имеет природа лиганда, а какой вклад вносится самой экстракционной системой. Мы также определили наиболее важные фрагменты для селективности этих лигандов для пар Am(III)/Eu(III) и Am(III)/Cm(III) (Рис. 4.19).

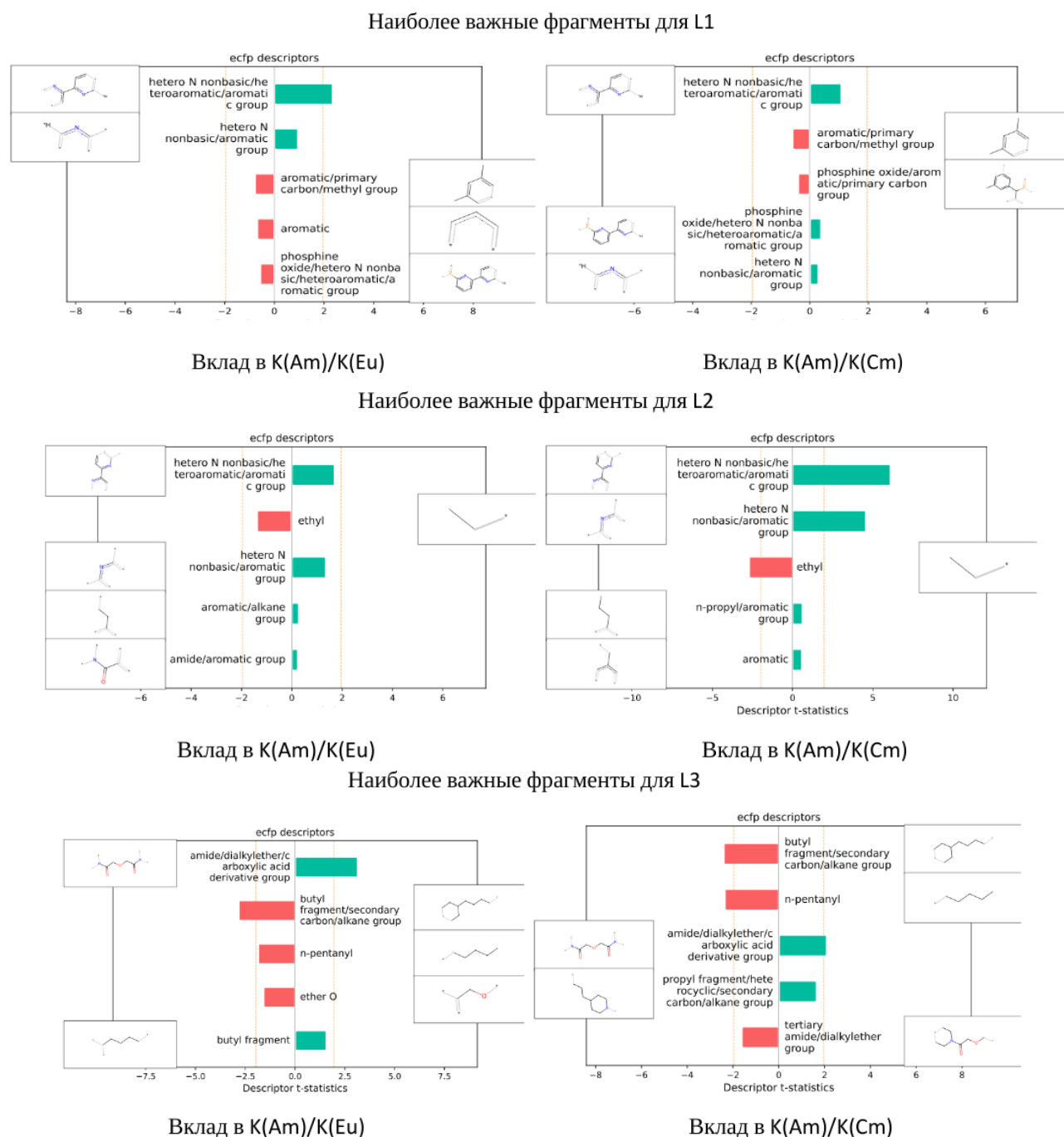


Рис. 4.19 — Наиболее важные фрагменты для экспериментально исследованных лигандов

Можно заметить, что для L2 вклады одних и тех же фрагментов с коорди-

национными азотами в отношении констант отличается, и для пары актинидов они оказываются больше чем для разделения Am(III) и Eu(III). Вероятно это связано с большим сродством актинидов к мягким координационным центрам. Для фрагментов лиганда L1, кажется необычным малый вклад фосфоросодержащих фрагментов в селективность. Что касается фрагментов для L3, то для обеих пар металлов наибольший положительный вклад вносит фрагмент содержащий координационные кислороды. При этом азотосодержащий фрагмент 4 вносит ощутимый вклад только для пары актинидов, что также может объясняться теорией Пирсона.

### 4.3 Генеративная модель

Поскольку модели для констант устойчивости комплексов показали высокое качество, на их основе были проведены поиски перспективных комплексообразователей с помощью генеративной модели. При подборе молекулы по эффективности, в качестве целевой функции использовался результат предсказания модели, а для оценки эффективности отношение предсказанных констант устойчивости. Были проведены генерации для разных металлов, а обсуждаться далее будут результаты для пар Am(III)/Eu(III) и Am(III)/Cm(III) как наиболее интересных с практической точки зрения, при этом предложенный инструмент может использоваться для любых других сочетаний.

#### 4.3.1 Генерация на основе случайных выборок

Одним из ключевых этапов работы генеративной модели является создание начальной выборки соединений, для которых потом происходят операции кроссовера и мутации. Состав этой начальной выборки может достаточно сильно влиять на направление поиска, который будет осуществлять модель, а ее размер скажется на скорости генерации и, соответственно, на том, сколько популяций получится изучить. Соответственно первым этапом работы с генеративной моделью была проверка того, насколько хорошо предложенный подход работает для случайной начальной выборки. Как было упомянуто выше, в процессе работы генеративной модели формируются две выборки — глобального и локального поиска. Для того, чтобы оценить как меняются эти выборки в процессе генерации, были построены распределения по отношениям констант для проме-

жуточных популяций для Am(III)/Eu(III) (Рис. 4.20) и Am(III)/Cm(III) (Рис. 4.21).

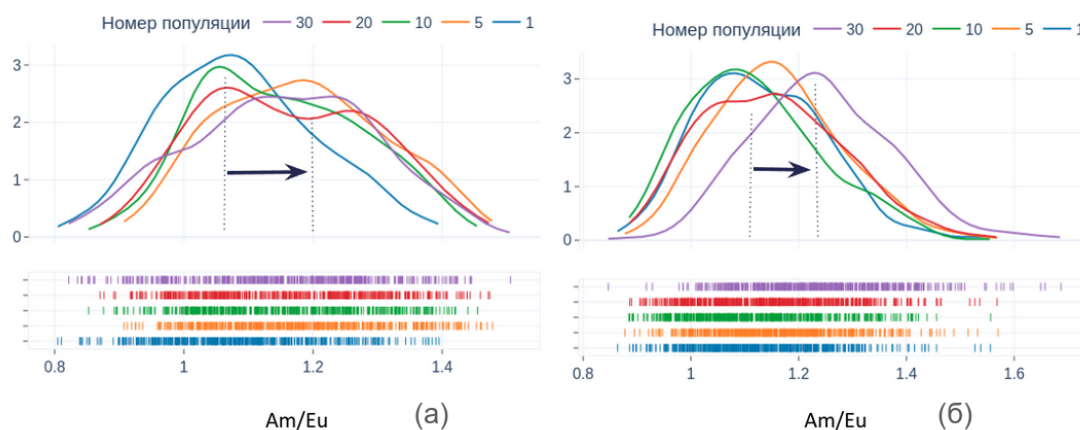


Рис. 4.20 — Распределения отношения констант устойчивости Am(III)/Eu(III) для (а) глобальной и (б) локальной выборки генеративной модели

Из распределений хорошо видно, как с ростом номера популяции, целевой параметр — отношение констант устойчивости, сдвигается в сторону увеличения, что и было ожидаемым поведением. Особенно сильно это заметно именно для локальной выборки, где и происходила тонкая настройка структуры соединений под целевое свойство.

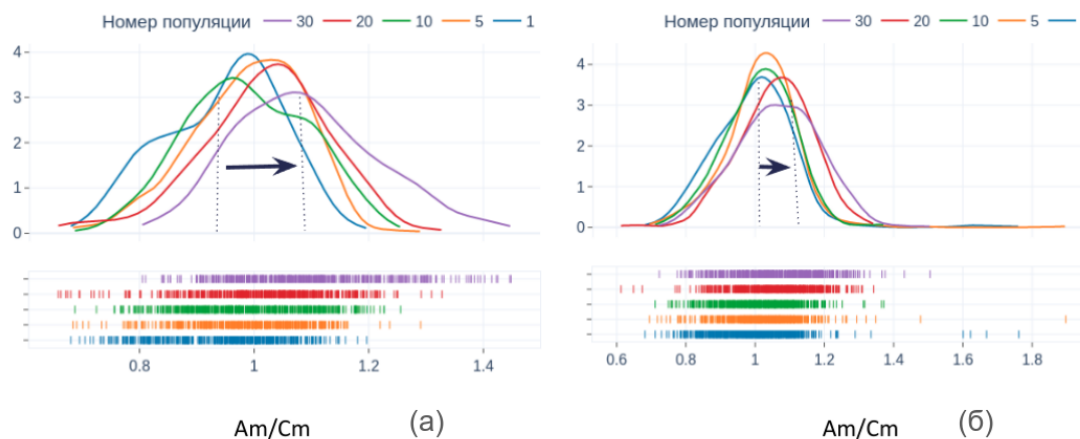
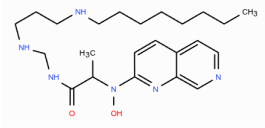
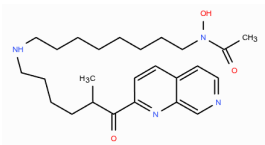
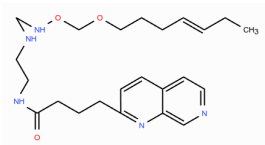
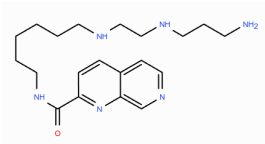
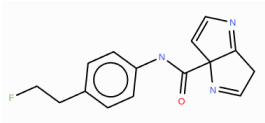
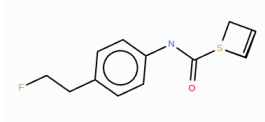
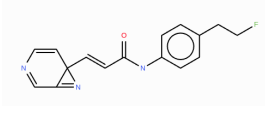
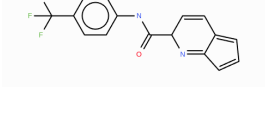


Рис. 4.21 — Распределения отношения констант устойчивости Am(III)/Cm(III) для (а) глобальной и (б) локальной выборки генеративной модели

Генерации проходили на одинаковых параметрах (размер популяции, число мутаций, обмен между выборками и т.д.), а так как вопрос разделения пары Am(III)/Cm(III) является более сложным с точки зрения подбора лиганда, следовательно для пары Am(III)/Cm(III) требуется более широкий поиск. Далее

мы проанализировали некоторые соединения, которые предложила генеративная модель (Таблицы 4.8). Мы отобрали серии лигандов с разными значениями констант устойчивости, по которым можно как оценить правильность работы модели и понять ее ограничения, так и узнать потенциальные способы влиять на селективность.

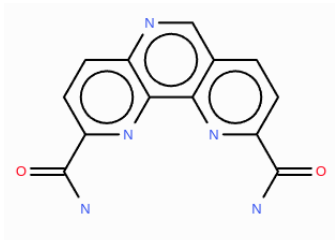
Таблица 4.8 — Результаты генераций на основе случайных начальных выборок

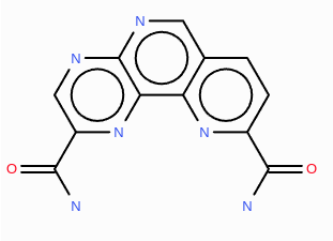
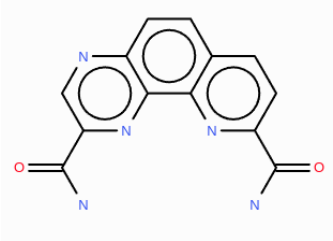
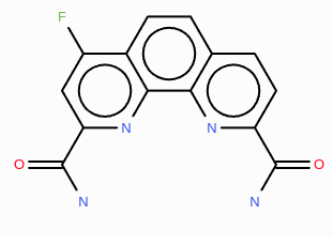
№	Лиганд	$\lg K_{Eu(III)}$	$\lg K_{Am(III)}$	$\lg K_{Cm(III)}$	$K_{Am(III)}/K_{Eu(III)}$	$K_{Am(III)}/K_{Cm(III)}$
1		9.81	5.89	9.59	0.0001	0.0002
2		6.80	5.27	6.89	0.03	0.0004
3		6.51	5.06	8.84	0.04	0.0002
4		7.34	6.43	11.20	0.12	$1.70 \times 10^{-5}$
5		7.27	10.1	7.04	671.33	1134.58
6		5.27	8.67	5.24	2484.40	2687.19
7		6.81	8.44	5.85	42.62	382.31
8		8.26	11.27	8.38	1026.86	274.12

Из результатов, приведенных в таблице 4.8 видно, что модель предложила соединения с разнообразной структурой и подходящими значениями целевых

свойств, однако становятся очевидными ограничения такого подхода. Поскольку при генерации алгоритм опирался преимущественно на значения целевого свойства, неучтенными оказывались другие важные свойства молекулы, в особенности синтетическая доступность. Так, соединения 1-4, похожих на которые в итоговой библиотеке данных достаточно много, видно, что модель достраивает длинный алифатический фрагмент с гетероатомами в его составе. Как видно из значений констант устойчивости, модель предсказывает высокую селективность таких соединений к паре Am(III)/Cm(III), но возможность синтеза таких соединений и их стабильность вызывает сомнения. Более многообещающими молекулами являются соединения 5-8, однако и здесь мы видим трех- и четырехзвенные гетероциклы в составе соединения, которые ограничивают как синтетическую доступность, так и стабильность подобных соединений. Из результатов случайных генераций можно сделать вывод о том, что учет синтетической доступности соединений - один из важнейших шагов при создании генеративной модели для дизайна новых соединений. В литературе неоднократно поднимался вопрос создания методов для решения этой задачи, но универсального решения пока предложено не было. Поскольку создание такого метода выходит за рамки текущего исследования мы опираемся на литературные данные при генерации новых соединений. Для этого мы провели ряд генераций, добавляя в начальную выборку каркасы известных в литературе экстрагентов, несколько примеров полученных при этих генерациях молекул приведены в таблице 4.9.

Таблица 4.9 — Примеры генераций на основе фенантролинов

№	Лиганд	$\lg K_{Eu(III)}$	$\lg K_{Am(III)}$	$\lg K_{Cm(III)}$	$K_{Am(III)}/K_{Eu(III)}$	$K_{Am(III)}/K_{Cm(III)}$
1		3.48	6.62	6.61	1372.40	1.01

2		3.59	6.74	6.8	1414.33	0.871
3		3.47	6.63	6.61	1455.12	1.05
4		3.68	6.93	7.31	1775.67	0.42

В приведенных структурах можно увидеть, как модель для увеличения селективности производила замещения в фенантролиновом ядре. Переходя от 1 соединения к 4 модель добавляет электроноакцепторную группу сначала в само фенантролиновое ядро, а затем и как заместитель в положение 4, при этом наблюдается увеличение селективности в обеих парах металлов. Похожий эффект увеличения селективности, особенно в паре Am/Cm был показан нашими коллегами в работах [169; 170] и этот эффект был переоткрыт нашей моделью не имея подобных соединений ни в обучающей выборке предсказательной модели ни в начальной выборке генеративной модели.

Таким образом, в результате работы над генеративной моделью мы показали, что применение генетического алгоритма в сочетании с предсказательной моделью для константы устойчивости комплекса позволяет получать перспективные комплексообразователи. Также мы подтвердили необходимость учета синтетической доступности при дизайне генеративных моделей. Используя разработанную модель, мы смогли переоткрыть тонкий химический эффект увеличения селективности разделения Am(III)/Cm(III) при добавлении электроноакцепторной группы в фенантролиновое ядро.

## Заключение

В работе разработана и протестирована система de-novo дизайна комплексообразователей для трехвалентных f-элементов. Для этого был решен комплекс задач по сбору баз данных, анализу свойств соединений и созданию вычислительных методов.

Показано, что отдельные свойства лиганда (жесткость, энергия предорганизации, электростатический потенциал) недостаточно точно описывают эффективность комплексообразования. Однако их нелинейная комбинация с учетом условий экстракции позволяет предсказывать фактор разделения с точностью  $0.44 \lg(SF)$ . В условиях дефицита экспериментальных данных впервые применено трансферное обучение для малых молекул, подтвердившее высокую эффективность при прогнозировании физико-химических свойств. Создана наиболее полная база экспериментальных констант устойчивости комплексов f-элементов. На ее основе с помощью специальной архитектуры нейросети построена предиктивная модель, определяющая константы устойчивости с рекордной точностью (ошибка до  $0.4 \lg K$ ). Модель валидирована в экстракционных экспериментах по разделению Am(III)/Eu(III) и Am(III)/Cm(III) с ранее неизученными лигандами, подтвердившими ее применимость для оценки селективности. Для автоматизации дизайна разработан генеративный подход, сочетающий генетический алгоритм и созданную предсказательную модель. Анализ результатов работы этого подхода показал потенциал таких систем при дизайне новых соединений, попутно переоткрыв известный в литературе эффект повышения селективности путем введения определенных заместителей в структуру фенантролиновых экстрагентов.

Работа демонстрирует перспективность вычислительных методов в радиохимии и предлагает набор готовых инструментов для разработки новых комплексообразователей.

## Выводы

1. Показано, что такие свойства лигандов, как жёсткость, энергия предорганизации и электростатический потенциал, по отдельности не коррелируют с селективностью в паре Am(III)/Eu(III). Однако их нелинейная комбинация с учётом условий экстракции позволяет предсказывать фактор разделения с точностью 0,44 [lg(SF)].
2. Разработана методология трансферного обучения для создания предсказательных моделей для малых органических молекул. На примере 12 физико-химических свойств показано, что трансферное обучение эффективно увеличивает область применимости модели и позволяет строить качественные модели при малом объёме данных.
3. Создана гибридная графовая нейронная сеть, использующая два входных слоя - для структуры лиганда и векторного представления металла. Модель предсказывает константы устойчивости для трехвалентных лантанидов и актинидов со средней ошибкой 0,4 [lgK], а также способна оценивать неопределённость предсказания.
4. Экспериментально определены коэффициенты распределения и факторы разделения для трёх новых лигандов для пар Am(III)/Eu(III) и Am(III)/Cm(III). Показано, что предсказанные моделью отношения констант устойчивости корректно ранжируют лиганды по селективности.
5. Создана генеративная модель на основе генетического алгоритма, интегрированная с предсказательной моделью констант устойчивости. Модель способна генерировать структуры перспективных комплексообразователей, а также она успешно переоткрыла известный эффект повышения селективности при введении электроноакцепторных заместителей в фенантролиновое ядро.

## Список литературы

1. *Agency I. A. E.* Energy, Electricity and Nuclear Power Estimates for the Period up to 2050 : тех. отч. / International Atomic Energy Agency. — 2022. — С. 1—137.
2. *Irish E., Reas W.* THE PUREX PROCESS-A SOLVENT EXTRACTION REPROCESSING METHOD FOR IRRADIATED URANIUM : тех. отч. / Hanford Atomic Products Operation. — 1957. — HW—49483 A, 4341712. — HW—49483 A.
3. Long-term aqueous alteration kinetics of an alpha-doped SON68 borosilicate glass / M. Tribet [и др.] // MRS Online Proceedings Library. — 2013. — Т. 1518, № 1. — С. 9—14.
4. *Nilsson M., Nash K.* Review Article: A Review of the Development and Operational Characteristics of the TALSPEAK Process // Solvent Extraction and Ion Exchange. — 2007. — Т. 25. — С. 665—701.
5. Impact Limits of Partitioning and Transmutation Scenarios on Radiotoxicity of Actinides in Radioactive Waste / J. Magill [и др.] // JRC Publications Repository. — 2003. — Т. 42. — С. 263—277.
6. *Romanovsky V.* R&D ACTIVITIES ON PARTITIONING IN RUSSIA // Proceedings of the 5th International Information Exchange Meeting on Actinide and Fission Product Partitioning and Transmutation. — St. Petersburg, Russia, 1998.
7. *Modolo G., Vijgen H., Serrano-Purroy D.* DIAMEX Counter-Current Extraction Process for Recovery of Trivalent Actinides from Simulated High Active Concentrate // Separation Science and Technology. — 2007. — Т. 42. — С. 439—452.
8. N,N,N',N'-Tetraoctyl Diglycolamide (TODGA): A Promising Extractant for Actinide - Partitioning from High - Level Waste (HLW) / S. Ansari [и др.] // Solvent Extraction and Ion Exchange. — 2005. — Т. 23, № 4. — С. 463—479.
9. Inhibiting radiolysis of BTP molecules by addition of nitrobenzene / M. Nilsson [и др.] // Radiochimica Acta. — 2006. — Т. 94, № 2. — С. 103—106.
10. ГОСТ 12.1.044-89 (ИСО 4589-84) Система стандартов безопасности труда (ССБТ). Пожаровзрывоопасность веществ и материалов. Номенклатура показателей и методы их определения (с Изменением N 1) - docs.cntd.ru.
11. First-Principles Study of the Separation of AmIII/CmIII from EuIII with Cyanex301 / X. Cao [и др.] // Inorganic Chemistry. — 2010. — Т. 49, № 22. — С. 10307—10315.
12. *Slater J.* A Simplification of the Hartree-Fock Method // Physical Review. — 1951. — Т. 81, № 3. — С. 385—390.
13. *McQuarrie D., Simon J.* Physical Chemistry: A Molecular Approach. — University Science Books, 1997.
14. *Fock V.* „Selfconsistent field“ mit Austausch für Natrium // Zeitschrift für Physik. — 1930. — Ноябрь. — Т. 62, № 11. — С. 795—805. — ISSN 0044-3328.

15. *Hartree D. R.* The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part II. Some Results and Discussion // Mathematical Proceedings of the Cambridge Philosophical Society. — 1928. — Т. 24, № 1. — С. 111–132.
16. *Bohr F., Henon E.* Comparison of Various Quantum Chemistry Methods for the Computation of Equilibrium Constants // Journal of Physical Chemistry A. — 1998. — Т. 102, № 25. — С. 4857–4862.
17. *David Sherrill C., Schaefer H.* The Configuration Interaction Method: Advances in Highly Correlated Approaches // Advances in Quantum Chemistry. Т. 34 / под ред. P.-O. L'owdin [и др.]. — Academic Press, 1999. — С. 143–269.
18. *Møller C., Plesset M.* Note on an Approximation Treatment for Many-Electron Systems // Physical Review. — 1934. — Т. 46, № 7. — С. 618–622.
19. *Cremer D.* Møller–Plesset perturbation theory: from small molecule methods to methods for thousands of atoms // WIREs Computational Molecular Science. — 2011. — Т. 1, № 4. — С. 509–530.
20. Chapter 10 - CORRELATION OF THE ELECTRONIC MOTIONS // Ideas of Quantum Chemistry / под ред. L. Piela. — Amsterdam : Elsevier, 2007. — С. 498–566.
21. *Bartlett R., Musiał M.* Coupled-cluster theory in quantum chemistry // Reviews of Modern Physics. — 2007. — Т. 79, № 1. — С. 291–352.
22. *Bartlett R.* Many-Body Perturbation Theory and Coupled Cluster Theory for Electron Correlation in Molecules // Annual Review of Physical Chemistry. — 1981. — Т. 32, № 1. — С. 359–401.
23. *Hohenberg P., Kohn W.* Inhomogeneous Electron Gas // Physical Review. — 1964. — Т. 136, 3B. — B864–B871.
24. *Kohn W., Sham L.* Self-Consistent Equations Including Exchange and Correlation Effects // Physical Review. — 1965. — Т. 140, 4A. — A1133–A1138.
25. *Kohn W.* Nobel Lecture: Electronic structure of matter—wave functions and density functionals // Reviews of Modern Physics. — 1999. — Т. 71, № 5. — С. 1253–1266.
26. *Sahni V., Bohnen K.-P., Harbola M.* Analysis of the local-density approximation of density-functional theory // Physical Review A. — 1988. — Т. 37, № 6. — С. 1895–1907.
27. *Perdew J., Burke K., Ernzerhof M.* Generalized Gradient Approximation Made Simple // Physical Review Letters. — 1996. — Т. 77, № 18. — С. 3865–3868.
28. *Smiga S., Constantin L.* Unveiling the Physics Behind Hybrid Functionals // Journal of Physical Chemistry A. — 2020. — Т. 124, № 27. — С. 5606–5614.
29. *Goodfellow I., Bengio Y., Courville A.* Deep Learning. — Illustrated. — Cambridge, Massachusetts The MIT Press, 2016.
30. *Rumelhart D., Hinton G., Williams R.* Learning representations by back-propagating errors // Nature. — 1986. — Т. 323, № 6088. — С. 533–536.

31. *Balaban A.* Neural Networks in QSAR and Drug Design // Journal of Chemical Information and Computer Sciences. — 1997. — Т. 37, № 3. — С. 628–629.
32. Informatics for Materials Science and Engineering: Data-driven Discovery for Accelerated Experimentation and Application / под ред. К. Rajan. — 1st. — Amsterdam : Butterworth-Heinemann, 2013.
33. *Todeschini R., Consonni V.* Handbook of Molecular Descriptors. — John Wiley & Sons, 2008.
34. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions / L. Alzubaidi [и др.] // Journal of Big Data. — 2021. — Т. 8, № 1. — С. 53.
35. A survey of transformers / Т. Lin [и др.] // AI Open. — 2022. — Т. 3. — С. 111–132.
36. Graph neural networks: A review of methods and applications / J. Zhou [и др.] // AI Open. — 2020. — Т. 1. — С. 57–81.
37. *Reiser P., Goscinski A., Lilienfeld O. von.* Graph neural networks for materials science and chemistry // Communications Materials. — 2022. — Т. 3, № 1. — С. 1–18.
38. Message Passing Neural Networks / J. Gilmer [и др.] // Machine Learning Meets Quantum Physics / под ред. К. Schutt, А. Suzuki, О. von Lilienfeld. — Cham : Springer International Publishing, 2020. — С. 199–214.
39. *Flam-Shepherd D., Chen J., Aspuru-Guzik A.* Neural Message Passing on High Order Paths // arXiv. — 2020.
40. *Aouichaoui A., Boulanger J., Varnek A.* Application of interpretable group-embedded graph neural networks for pure compound properties // Computers & Chemical Engineering. — 2023. — Т. 176. — С. 108291.
41. Molecular graph convolutions: moving beyond fingerprints / S. Kearnes [и др.] // Journal of Computer-Aided Molecular Design. — 2016. — Т. 30, № 8. — С. 595–608.
42. A compact review of molecular property prediction with graph neural networks / O. Wieder [и др.] // Drug Discovery Today: Technologies. — 2020. — Т. 37. — С. 1–12.
43. Chemi-Net: A Molecular Graph Convolutional Network for Accurate Drug Property Prediction / K. Liu [и др.] // International Journal of Molecular Sciences. — 2019. — Т. 20, № 14. — С. 3389.
44. Could graph neural networks learn better molecular representation for drug discovery? A comparison study of descriptor-based and graph-based models / D. Jiang [и др.] // Journal of Cheminformatics. — 2021. — Т. 13, № 1. — С. 12.
45. Deep learning modeling strategy for material science: from natural materials to metamaterials / W. Li [и др.] // Journal of Physics: Materials. — 2022. — Т. 5, № 1. — С. 014003.
46. *Sanchez-Lengeling B., Aspuru-Guzik A.* Inverse molecular design using machine learning: Generative models for matter engineering // Science. — 2018. — Т. 361, № 6400. — С. 360–365.

47. *Anstine D., Isayev O.* Generative Models as an Emerging Paradigm in the Chemical Sciences // Journal of the American Chemical Society. — 2023. — Т. 145, № 16. — С. 8736—8750.
48. *Segler M., Preuss M., Waller M.* Generating Focused Molecule Libraries for Drug Discovery with Recurrent Neural Networks // ACS Central Science. — 2018. — Т. 4, № 1. — С. 120—131.
49. De Novo Design of Bioactive Small Molecules by Artificial Intelligence / D. Merk [и др.] // Molecular Informatics. — 2018. — Т. 37, № 1/2. — С. 1700153.
50. Generative and reinforcement learning approaches for the automated de novo design of bioactive compounds / M. Korshunova [и др.] // Communications Chemistry. — 2022. — Т. 5, № 1. — С. 1—11.
51. Deep learning enables rapid identification of potent DDR1 kinase inhibitors / A. Zhavoronkov [и др.] // Nature Biotechnology. — 2019. — Т. 37, № 9. — С. 1038—1040.
52. A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications / J. Gui [и др.] // IEEE Transactions on Knowledge and Data Engineering. — 2023. — Т. 35, № 4. — С. 3313—3332.
53. Generative adversarial networks (GAN) based efficient sampling of chemical composition space for inverse design of inorganic materials / Y. Dan [и др.] // npj Computational Materials. — 2020. — Т. 6, № 1. — С. 1—7.
54. Generative Adversarial Networks for Crystal Structure Prediction / S. Kim [и др.] // ACS Central Science. — 2020. — Т. 6, № 8. — С. 1412—1420.
55. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules / R. G'omez-Bombarelli [и др.] // ACS Central Science. — 2018. — Т. 4, № 2. — С. 268—276.
56. *Kobyzev I., Prince S., Brubaker M.* Normalizing Flows: An Introduction and Review of Current Methods // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2021. — Т. 43, № 11. — С. 3964—3979.
57. *Kuznetsov M., Polykovskiy D.* MolGrow: A Graph Normalizing Flow for Hierarchical Molecular Generation // Proceedings of the AAAI Conference on Artificial Intelligence. Т. 35. — 2021. — С. 8226—8234.
58. *Katoch S., Chauhan S., Kumar V.* A review on genetic algorithm: past, present, and future // Multimedia Tools and Applications. — 2021. — Т. 80, № 5. — С. 8091—8126.
59. *Terfloth L., Gasteiger J.* Neural networks and genetic algorithms in drug design // Drug Discovery Today. — 2001. — Т. 6. — С. 102—108.
60. *Venkatasubramanian V., Chan K., Caruthers J.* Computer-aided molecular design using genetic algorithms // Computers & Chemical Engineering. — 1994. — Т. 18, № 9. — С. 833—844.
61. *Venkatasubramanian V., Chan K., Caruthers J.* Evolutionary Design of Molecules with Desired Properties Using the Genetic Algorithm // Journal of Chemical Information and Computer Sciences. — 1995. — Т. 35, № 2. — С. 188—195.

62. *Jensen J.* A graph-based genetic algorithm and generative model/Monte Carlo tree search for the exploration of chemical space // *Chemical Science*. — 2019. — Т. 10, № 12. — С. 3567–3572.
63. Guiding Deep Molecular Optimization with Genetic Exploration / S. Ahn [и др.] // *Advances in Neural Information Processing Systems*. Т. 33. — Curran Associates, Inc., 2020. — С. 12008–12021.
64. *Kwon Y., Lee J.* MolFinder: an evolutionary algorithm for the global optimization of molecular properties and the extensive exploration of chemical space using SMILES // *Journal of Cheminformatics*. — 2021. — Т. 13, № 1. — С. 24.
65. SELFIES and the future of molecular string representations / M. Krenn [и др.] // *Patterns*. — 2022. — Т. 3, № 10. — С. 100588.
66. Beyond generative models: superfast traversal, optimization, novelty, exploration and discovery (STONED) algorithm for molecules using SELFIES / A. Nigam [и др.] // *Chemical Science*. — 2021. — Т. 12, № 20. — С. 7079–7090.
67. *Steinmann C., Jensen J.* Using a genetic algorithm to find molecules with good docking scores // *PeerJ Physical Chemistry*. — 2021. — Т. 3. — e18.
68. Mechanism of the reaction between atomic tritium and the functional groups of aminoacids. Quantum chemical calculations and experiment. / K. German [и др.] // *Radiochemistry*. — 1982. — Ноябрь. — Т. 24. — С. 567–569.
69. *German K., Kryutchkov S.* Polynuclear technetium halide clusters // *Russian Journal of Inorganic Chemistry*. — 2002. — Апрель. — Т. 47. — С. 578–583.
70. *Plekhanov Y. V., German K. E., Sekine R.* Electronic Structure of Binuclear Technetium Chloroacetate Cluster: Quantum Chemical Calculation and Assignment of Optical and XPE Spectra // *Radiochemistry*. — 2003. — Май. — Т. 45, № 3. — С. 243–249. — ISSN 1608-3288.
71. Speciation and Complex Formation of Zirconium(IV) with Tc(VII) and Re(VII) in Liquid–Liquid Extraction Systems / A. P. Novikov [и др.] // *ACS Omega*. — 2025. — Август. — Т. 10, № 33. — С. 38064–38072. — Publisher: American Chemical Society.
72. Intramolecular Re...O Nonvalent Interactions as a Stabilizer of the Polyoxorhenate(VII) / M. A. Volkov [и др.] // *Inorganic Chemistry*. — 2023. — Август. — Т. 62, № 33. — С. 13485–13494. — ISSN 0020-1669. — Publisher: American Chemical Society.
73. Speciation of heptavalent technetium in sulfuric acid: structural and spectroscopic studies / F. Poineau [и др.] // *Dalton Trans*. — 2010. — Т. 39, № 37. — С. 8616–8619. — Publisher: The Royal Society of Chemistry.
74. *Plekhanov Y., German K.* Quantum Chemical Model of Technetium Carbide // *Journal of Nuclear and Radiochemical Sciences*. — 2005. — Декабрь. — Т. 6. — С. 215–216.
75. *Front Matter* // *Computational Methods in Lanthanide and Actinide Chemistry*. — John Wiley & Sons, Ltd, 2015. — С. i–xviii.

76. *Liu W., Franke R., Dolg M.* Relativistic ab initio and density functional theory calculations on the mercury fluorides: Is HgF<sub>4</sub> thermodynamically stable? // *Chemical Physics Letters*. — 1999. — T. 302, № 3/4. — C. 231–239.
77. *Schreckenbach G., Shamov G.* Theoretical Actinide Molecular Science // *Accounts of Chemical Research*. — 2010. — T. 43, № 1. — C. 19–29.
78. *Aebbersold L., Hesselmann A., Knecht S.* Efficacy of Density Functionals and Relativistic Effective Core Potentials for Lanthanide-Containing Species: The Ln<sub>54</sub> Molecule Set // *Journal of Chemical Theory and Computation*. — 2017. — T. 13, № 6. — C. 2831–2839.
79. *Vetere V., De Saint Jean C., Pouchan C.* A theoretical study of the bonding in trivalent americium complexes // *Chemical Physics Letters*. — 2004. — T. 396, № 4. — C. 452–457.
80. *Chai J.-D., Head-Gordon M.* Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections // *Physical Chemistry Chemical Physics*. — 2008. — T. 10, № 44. — C. 6615–6620.
81. *Zhao Y., Truhlar D.* The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals // *Theoretical Chemistry Accounts*. — 2008. — T. 120, № 1. — C. 215–241.
82. *Grimme S.* Semiempirical hybrid density functional with perturbative second-order correlation // *Journal of Chemical Physics*. — 2006. — T. 124, № 3. — C. 034108.
83. A search for a DFT functional for actinide compounds / A. Mitrofanov [и др.] // *The Journal of Chemical Physics*. — 2021. — T. 155, № 16. — C. 161103.
84. Benzoazacrown compound: a highly effective chelator for therapeutic bismuth radioisotopes / E. Matzova [и др.] // *Medicinal Chemistry Communications*. — 2019. — T. 10, № 9. — C. 1641–1645.
85. *Pykkö P.* Relativistic effects in structural chemistry // *Chemical Reviews*. — 1988. — T. 88, № 3. — C. 563–594.
86. *Chang C., Pelissier M., Durand P.* Regular Two-Component Pauli-Like Effective Hamiltonians in Dirac Theory // *Physica Scripta*. — 1986. — T. 34, № 5. — C. 394.
87. Relativistic regular two-component Hamiltonians / E. Lenthe [и др.] // *International Journal of Quantum Chemistry*. — 1996. — Февр. — T. 57. — C. 281–293.
88. *Douglas M., Kroll N.* Quantum electrodynamical corrections to the fine structure of helium // *Annals of Physics*. — 1974. — T. 82. — C. 89–155.
89. *Hess B.* Applicability of the no-pair equation with free-particle projection operators to atomic and molecular structure calculations // *Physical Review A*. — 1985. — T. 32, № 2. — C. 756–763.
90. *Hong G., Dolg M., Li L.* A comparison of scalar-relativistic ZORA and DKH density functional schemes: monohydrides, monooxides and monofluorides of La, Lu, Ac and Lr // *Chemical Physics Letters*. — 2001. — T. 334, № 4. — C. 396–402.

91. Probing the difference in covalence by enthalpy measurements: a new heterocyclic N-donor ligand for actinide/lanthanide separation / Y. Yang [и др.] // Dalton Transactions. — 2015. — Т. 44, № 19. — С. 8959–8970.
92. Theoretical insights into the separation of Am(III) over Eu(III) with PhenBHPPA / H. Wu [и др.] // Dalton Transactions. — 2015. — Т. 44, № 38. — С. 16737–16745.
93. Solvent extraction of americium(III) and europium(III) with tridentate N,N-dialkyl-1,10-phenanthroline-2-amide-derived ligands: extraction, complexation and theoretical study / S. Cao [и др.] // New Journal of Chemistry. — 2016. — Т. 40, № 12. — С. 10560–10568.
94. On the Origin of Covalent Bonding in Heavy Actinides. / M. P. Kelley [и др.] // Journal of the American Chemical Society. — 2017. — Июль. — Т. 139, № 29. — С. 9901–9908.
95. *Altmaier M., Gaona X., Fangh"anel T.* Recent Advances in Aqueous Actinide Chemistry and Thermodynamics // Chemical Reviews. — 2013. — Т. 113, № 2. — С. 901–943.
96. *Bhattacharyya A., Mohapatra P.* The separation and complexation behavior of trivalent actinides and lanthanides // BARC Newsletter. — 2021. — № 378. — С. 19–24.
97. Relativistic quantum calculations to understand the contribution of f-type atomic orbitals and chemical bonding of actinides with organic ligands / A. Zapata-Escobar [и др.] // Physical Chemistry Chemical Physics. — 2023. — Т. 25, № 7. — С. 5592–5601.
98. *Solov'ev V., Varnek A., Wipff G.* Modeling of Ion Complexation and Extraction Using Substructural Molecular Fragments // Journal of Chemical Information and Computer Sciences. — 2000. — Т. 40, № 3. — С. 847–858.
99. Benchmarking of Linear and Nonlinear Approaches for Quantitative Structure-Property Relationship Studies of Metal Complexation with Ionophores / I. Tetko [и др.] // Journal of Chemical Information and Modeling. — 2006. — Т. 46, № 2. — С. 808–819.
100. ISIDA - Platform for Virtual Screening Based on Fragment and Pharmacophoric Descriptors / A. Varnek [и др.] // Current Computer - Aided Drug Design. — 2008. — Сент. — Т. 4. — С. 191–198.
101. Stability constants of complexes of Zn<sup>2+</sup>, Cd<sup>2+</sup>, and Hg<sup>2+</sup> with organic ligands: QSPR consensus modeling and design of new metal binders / V. Solov'ev [и др.] // Journal of Inclusion Phenomena and Macrocyclic Chemistry. — 2012. — Т. 72, № 3. — С. 309–321.
102. QSPR Modeling of the AmIII/EuIII Separation Factor: How Far Can we Predict? / A. Varnek [и др.] // Solvent Extraction and Ion Exchange. — 2007. — Т. 25, № 1. — С. 1–26.
103. *Quang N., Nhung N., Van Tat P.* An insight QSPR-based prediction model for stability constants of metal-thiosemicarbazone complexes using MLR and ANN methods // Vietnam Journal of Chemistry. — 2019. — Т. 57, № 4. — С. 500–506.
104. *Kanahashi K., Urushihara M., Yamaguchi K.* Machine learning-based analysis of overall stability constants of metal–ligand complexes // Scientific Reports. — 2022. — Т. 12, № 1. — С. 11159.
105. Deep Learning Insights into Lanthanides Complexation Chemistry / A. Mitrofanov [и др.] // Molecules. — 2021. — Т. 26, № 11. — С. 3237.

106. *Reymond J.-L.* The Chemical Space Project // *Accounts of Chemical Research*. — 2015. — Т. 48, № 3. — С. 722–730.
107. *Korn M., Schröder M., Schneider G.* Navigating large chemical spaces in early-phase drug discovery // *Current Opinion in Structural Biology*. — 2023. — Т. 80. — С. 102578.
108. Designing catalysts with deep generative models and computational data. A case study for Suzuki cross coupling reactions / O. Schilter [и др.] // *Digital Discovery*. — 2023. — Т. 2, № 3. — С. 728–735.
109. *Lee Z.-H., Lin P., Yang T.* Inverse design of ligands using a deep generative model semi-supervised by a data-driven ligand field strength metric // *Journal of the Chinese Chemical Society*. — 2023. — Т. 70, № 5. — С. 1095–1101.
110. *Duan C., Rudorff G. von, Lilienfeld O. von.* Active Learning Exploration of Transition-Metal Complexes to Discover Method-Insensitive and Synthetically Accessible Chromophores // *JACS Au*. — 2023. — Т. 3, № 2. — С. 391–401.
111. *Laplaza R., Gallarati S., Corminboeuf C.* Genetic Optimization of Homogeneous Catalysts // *Chemistry–Methods*. — 2022. — Т. 2, № 6. — e202100107.
112. *Karulin B., Kozhevnikov M.* Ketcher: web-based chemical structure editor // *Journal of Cheminformatics*. — 2011. — App. — Т. 3, № 1. — P3. — ISSN 1758-2946.
113. *Simmons E.* Markush structure searching over the years // *World Patent Information*. — 2003. — Т. 25, № 3. — С. 195–202.
114. *Spicher S., Grimme S.* Robust Atomistic Modeling of Materials, Organometallic, and Biochemical Systems // *Angewandte Chemie International Edition*. — 2020. — Т. 59, № 36. — С. 15665–15673. — ISSN 1521-3773.
115. Extended tight-binding quantum chemistry methods / C. Bannwarth [и др.] // *WIREs Computational Molecular Science*. — 2021. — Т. 11, № 2. — e1493. — ISSN 1759-0884.
116. *Stewart J.* Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters // *Journal of Molecular Modeling*. — 2013. — Т. 19, № 1. — С. 1–32.
117. *Moussa J. E., Stewart J. J. P.* OpenMOPAC. — 2020. — URL: <http://openmopac.net/> ; v23.2.2.
118. *Becke A. D.* Density-functional thermochemistry. III. The role of exact exchange // *Journal of Chemical Physics*. — 1993. — App. — Т. 98. — С. 5648–5652. — ISSN 0021-9606.
119. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields / P. J. Stephens [и др.] // *The Journal of Physical Chemistry*. — 1994. — Ноябрь. — Т. 98, № 45. — С. 11623–11627. — ISSN 0022-3654.
120. 6-31G\* basis set for atoms K through Zn / V. Rassolov [и др.] // *The Journal of Chemical Physics*. — 1998. — Т. 109, № 4. — С. 1223–1229.

121. *Weigend F., Ahlrichs R.* Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy // *Phys. Chem. Chem. Phys.* — 2005. — Т. 7, вып. 18. — С. 3297–3305.
122. A generally applicable atomic-charge dependent London dispersion correction / E. Caldeweyher [и др.] // *The Journal of Chemical Physics.* — 2019. — Апр. — Т. 150, № 15. — С. 154122. — ISSN 0021-9606.
123. *Caldeweyher E., Bannwarth C., Grimme S.* Extension of the D3 dispersion coefficient model // *The Journal of Chemical Physics.* — 2017. — Июль. — Т. 147, № 3. — С. 034112. — ISSN 0021-9606.
124. *Neese F.* The SHARK integral generation and digestion system // *Journal of Computational Chemistry.* — 2023. — Т. 44, № 3. — С. 381–396.
125. An improved chain of spheres for exchange algorithm / B. Helmich-Paris [и др.] // *The Journal of Chemical Physics.* — 2021. — Сент. — Т. 155, № 10. — С. 104109. — ISSN 0021-9606.
126. *Neese F.* An improvement of the resolution of the identity approximation for the formation of the Coulomb matrix // *Journal of Computational Chemistry.* — 2003. — Т. 24, № 14. — С. 1740–1747.
127. *Neese F.* Software update: The ORCA program system—Version 5.0 // *WIREs Computational Molecular Science.* — 2022. — Т. 12, № 5. — e1606.
128. *Grimme S., Ehrlich S., Goerigk L.* Effect of the damping function in dispersion corrected density functional theory // *Journal of Computational Chemistry.* — 2011. — Т. 32, № 7. — С. 1456–1465. — ISSN 1096-987X.
129. *Drago R. S.* Pearson's quantitative statement of HSAB [hard-soft acid-base] // *Inorganic Chemistry.* — 1973. — Сент. — Т. 12, № 9. — С. 2211–2212. — ISSN 0020-1669.
130. *Koopmans T.* Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den Einzelnen Elektronen Eines Atoms // *Physica.* — 1934. — Т. 1, № 1. — С. 104–113. — ISSN 0031-8914.
131. ConfGen: A Conformational Search Method for Efficient Generation of Bioactive Conformers / K. S. Watts [и др.] // *Journal of Chemical Information and Modeling.* — 2010. — Апр. — Т. 50, № 4. — С. 534–546. — ISSN 1549-9596.
132. RDKit: Open-source Cheminformatics / G. Landrum [и др.]. — 2006. — URL: <https://www.rdkit.org>; DOI: 10.5281/zenodo.591637.
133. *Lu T.* A comprehensive electron wavefunction analysis toolbox for chemists, Multiwfn. // *The Journal of chemical physics.* — 2024. — Авг. — Т. 161, № 8. — С. 082503. — ISSN 1089-7690 0021-9606.
134. *Lu T., Chen F.* Multiwfn: A multifunctional wavefunction analyzer // *Journal of Computational Chemistry.* — 2012. — Т. 33, № 5. — С. 580–592.
135. ISIDA Property-Labelled Fragment Descriptors. / F. Ruggiu [и др.] // *Molecular informatics.* — 2010. — Дек. — Т. 29, № 12. — С. 855–868. — ISSN 1868-1743.

136. CatBoost: unbiased boosting with categorical features / L. Prokhorenkova [и др.]. — 01.2019. — arXiv:1706.09516 [cs].
137. Selective Extraction of Minor Actinides from Acidic Media Using Symmetric and Asymmetric Dithiophosphinic Acids / D. R. Peterman [и др.] // Separation Science and Technology. — 2010. — Июль. — Т. 45, № 12/13. — С. 1711–1717. — ISSN 0149-6395.
138. Separation of Am(III) from Eu(III) by extraction based on in situ extractant formation of dithiocarbamate derivatives / S. Miyashita [и др.] // Progress in Nuclear Energy. — 2008. — Март. — Т. 50, № 2. — С. 499–503. — ISSN 0149-1970.
139. Extraction of Am(III) and lanthanides(III) with organo dithiophosphinic acids / Q. Xu [и др.] // Radiochimica Acta. — 2008. — Дек. — Т. 96, № 12. — С. 771–779. — ISSN 2193-3405.
140. Sulfur donating extractants for the separation of trivalent actinides and lanthanides / N. P. Bessen [и др.] // Coordination Chemistry Reviews. — 2020. — Окт. — Т. 421. — С. 213446. — ISSN 0010-8545.
141. Advancing the Am Extractant Design through the Interplay among Planarity, Preorganization, and Substitution Effects / X. Zhang [и др.] // Inorganic Chemistry. — 2022. — Авг. — Т. 61, № 30. — С. 11556–11570. — ISSN 0020-1669.
142. Separation and Complexation of Trivalent Actinides and Lanthanides by Two Novel Asymmetric N,O-Hybrid Pyridyl Ligands: A Combination of Phosphoryl and Triazinyl Groups / Y. Miao [и др.] // Inorganic Chemistry. — 2022. — Ноябрь. — Т. 61, № 44. — С. 17911–17923. — ISSN 0020-1669.
143. Mechanism of trivalent actinide/lanthanide separation using synergistic mixtures of di(chlorophenyl)phosphinic acid and neutral O-bearing co-extractants / G. Ionova [и др.] // New Journal of Chemistry. — 2001. — Янв. — Т. 25, № 3. — С. 491–501. — ISSN 1369-9261.
144. An(III)/Ln(III) solvent extraction: Theoretical and experimental investigation of the role of ligand conformational mobility / A. Mitrofanov [и др.] // Journal of Molecular Liquids. — 2021. — Март. — Т. 325. — С. 115098. — ISSN 0167-7322.
145. BTBPs versus BTPPhens: some reasons for their differences in properties concerning the partitioning of minor actinides and the advantages of BTPPhens / F. Lewis [и др.] // Inorganic Chemistry. — 2013. — Т. 52, № 9. — С. 4993–5005.
146. Complexation of Bi<sup>3+</sup>, Ac<sup>3+</sup>, Y<sup>3+</sup>, Lu<sup>3+</sup>, La<sup>3+</sup> and Eu<sup>3+</sup> with benzo-diaza-crown ether with carboxylic pendant arms / B. Egorova [и др.] // Radiochimica Acta. — 2016. — Т. 104, № 8. — С. 555–565.
147. Antunes L., Grau-Crespo R., Butler K. Distributed representations of atoms and materials for machine learning // npj Computational Materials. — 2022. — Т. 8, № 1. — С. 1–9.
148. Optuna: A Next-generation Hyperparameter Optimization Framework / T. Akiba [и др.]. — 2019. — arXiv: 1907.10902 [cs.LG].

149. PyTorch: An Imperative Style, High-Performance Deep Learning Library / A. Paszke [и др.] // Advances in Neural Information Processing Systems. Т. 32. — Curran Associates, Inc., 2019.
150. Scikit-learn: Machine Learning in Python / F. Pedregosa [и др.] // Journal of Machine Learning Research. — 2011. — Т. 12. — С. 2825–2830.
151. Wellawatte G., Seshadri A., White A. Model agnostic generation of counterfactual explanations for molecules // Chemical Science. — 2022. — Т. 13, № 13. — С. 3697–3705.
152. Gandhi H. A., White A. D. Explaining molecular properties with natural language // Т. 2022. — 2022. — DOI: 10.26434/chemrxiv-2022-v5p6m-v3.
153. Wales D. J., Doye J. P. K. Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms // The Journal of Physical Chemistry A. — 1997. — Июль. — Т. 101, № 28. — С. 5111–5116. — ISSN 1089-5639.
154. Goedecker S. Minima hopping: An efficient search method for the global minimum of the potential energy surface of complex molecular systems // The Journal of Chemical Physics. — 2004. — Т. 120, № 21. — С. 9911–9917.
155. Algorithms for hyper-parameter optimization / J. Bergstra [и др.] // Proceedings of the 25th International Conference on Neural Information Processing Systems. — Granada, Spain, 2011. — С. 2546–2554. — (NIPS'11). — ISBN 9781618395993.
156. Ebejer J.-P., Morris G. M., Deane C. M. Freely Available Conformer Generation Methods: How Good Are They? // Journal of Chemical Information and Modeling. — 2012. — Май. — Т. 52, № 5. — С. 1146–1158. — ISSN 1549-9596.
157. Riniker S., Landrum G. A. Better Informed Distance Geometry: Using What We Know To Improve Conformation Generation // Journal of Chemical Information and Modeling. — 2015. — Дек. — Т. 55, № 12. — С. 2562–2574. — ISSN 1549-9596.
158. Nigam A., Pollice R., Aspuru-Guzik A. Parallel tempered genetic algorithm guided by deep neural networks for inverse molecular design // Digital Discovery. — 2022. — Т. 1, № 4. — С. 390–404.
159. Ertl P., Schuffenhauer A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions // Journal of Cheminformatics. — 2009. — Т. 1, № 1. — С. 8.
160. New Diamides of 2,2'-dipyridyl-6,6'-dicarboxylic Acid for Actinide-Lanthanide Separation / M. Alyapyshev [и др.] // Solvent Extraction and Ion Exchange. — 2014. — Т. 32, № 2. — С. 138–152.
161. Babain V., Alyapyshev M., Kiseleva R. Metal extraction by N,N'-dialkyl-N,N'-diaryl-dipicolinamic from nitric acid solutions // Radiochimica Acta. — 2007. — Т. 95, № 4. — С. 217–223.
162. Alyapyshev M., Babain V., Ustynyuk Y. Recovery of minor actinides from high-level wastes: modern trends // Russian Chemical Reviews. — 2016. — Т. 85, № 9. — С. 943.

163. Complexation and Separation of Trivalent Actinides and Lanthanides by a Novel DGA Derived from Macrocyclic Crown Ether: Synthesis, Extraction, and Spectroscopic and Density Functional Theory Studies / Y. Fan [и др.] // ACS Omega. — 2021. — Т. 6, № 3. — С. 2156—2166.
164. *Rogers R., Bauer C., Bond A.* Crown ethers as actinide extractants in acidic aqueous biphasic systems: partitioning behavior in solution and crystallographic analyses of the solid state // Journal of Alloys and Compounds. — 1994. — Т. 213/214. — С. 305—312.
165. *Wenji W., Xiaoming G., Wenming L.* Extraction of several lanthanide and actinide radionuclides with crown ethers // Journal of Radioanalytical Chemistry. — 1983. — Т. 76, № 1. — С. 49—62.
166. *Wold S., Esbensen K., Geladi P.* Principal component analysis // Chemometrics and Intelligent Laboratory Systems. — 1987. — Т. 2, № 1. — С. 37—52.
167. *Irwin J., Shoichet B.* ZINC – A Free Database of Commercially Available Compounds for Virtual Screening // Journal of Chemical Information and Modeling. — 2005. — Т. 45, № 1. — С. 177—182.
168. *Wang Z.-M., Burgt L. van de, Choppin G.* Spectroscopic study of lanthanide(III) complexes with aliphatic dicarboxylic acids // Inorganica Chimica Acta. — 2000. — Т. 310, № 2. — С. 248—256.
169. Unravelling the mechanism of f-element extraction by phenanthroline-diamides: A case of 4,7-substituted 1,10-phenanthroline-2,9-diamides / M. V. Evsiunina [и др.] // Separation and Purification Technology. — 2024. — Т. 339. — С. 126621. — ISSN 1383-5866.
170. 4-Oxo-7-fluoro-1,10-phenanthroline-2,9-diamides: Synthesis, Structural Features, Lanthanide Complexes, and Am(III)/Ln(III) Solvent Extraction / N. A. Avagyan [и др.] // Inorganic Chemistry. — 2023. — Окт. — Т. 62, № 43. — С. 17721—17735. — ISSN 0020-1669.