

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
имени М.В. ЛОМОНОСОВА

*На правах рукописи*

**Карпов Кирилл Викторович**

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

Специальность 1.4.13 Радиохимия

**АВТОРЕФЕРАТ**

диссертации на соискание ученой степени

кандидата химических наук

Москва - 2026

Диссертационная работа подготовлена в межкафедральной лаборатории интеллектуального химического дизайна на кафедре радиохимии химического факультета Московского государственного университета имени М.В. Ломоносова.

- Научный руководитель** - **Митрофанов Артем Александрович**, кандидат химических наук
- Официальные оппоненты**
- **Скорб Екатерина Владимировна**, доктор химических наук, профессор, Национальный исследовательский университет ИТМО, директор мегафакультета наук о жизни
  - **Успенская Ирина Александровна**, доктор химических наук, Московский государственный университет имени М.В. Ломоносова, кафедра физической химии, доцент
  - **Герман Константин Эдуардович**, доктор химических наук, Институт физической химии и электрохимии им. А.Н. Фрумкина Российской академии наук, главный научный сотрудник с возложением обязанностей заведующего лабораторией химии технеция

Защита диссертации состоится «10» июня 2026 г. в 14:00 на заседании диссертационного совета МГУ.014.6 Московского государственного университета имени М.В. Ломоносова по адресу: 119991, Москва, ГСП-1, Ленинские горы, 1, стр.10, ауд. 308.

E-mail: severin@radio.chem.msu.ru (А.В. Северин, ученый секретарь диссертационного совета МГУ.014.6).

С диссертацией можно ознакомиться в отделе диссертаций научной библиотеки МГУ имени М.В. Ломоносова (Ломоносовский просп., д. 27) и на портале: <https://dissovet.msu.ru/dissertation/3865>

Автореферат разослан «\_\_» апреля 2026 г.

Учёный секретарь  
диссертационного совета,  
кандидат химических наук

А. В. Северин

# Общая характеристика работы

## Актуальность темы исследования

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 энергетики 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 энергетики. 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 и задачи исследования

**Целью работы** является 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 и экспериментальных значений факторов разделения и констант устойчивости

2. Разработка и проверка на литературных данных предсказательных моделей машинного обучения для оценки устойчивости комплексов металл-лиганд
3. Проведение экстракционных экспериментов для проверки качества разработанных предсказательных моделей машинного обучения
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 f-элементов.

**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 структура-свойство, 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. 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 4 научных 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.

## Основное содержание работы

Во **введении** обоснована актуальность, сформулированы цели и задачи работы, определены научная новизна и практическая значимость, также перечислены положения, выносимые на защиту и способы апробации работы. **Обзор литературы** разделен на три части. В первой части рассмотрены общие сведения об обращении с отработанным ядерным топливом и о требованиях, предъявляемых перспективным экстрагентам для разделения радиоактивных отходов. Во второй части рассмотрены такие методы вычислительной химии как квантовохимические методы и методы искусственного интеллекта. Были рассмотрены их достоинства и ограничения при расчете разных химических систем. В третьей части рассмотрены конкретные применения вычислительных методов для расчета систем с f-элементами. **Расчетная часть** работы содержит подробное описание используемых вычислительных методов, механизма их работы способов имплементации, в случае если разработка метода осуществлялась в рамках данной работы. **Экспериментальная часть** содержит описание проведения экстракционных экспериментов и анализа полученных результатов. В **обсуждении результатов** рассматриваются результаты проведенных расчетов, качество полученных моделей, их соответствия экспериментальным результатам и анализ сгенерированных комплексообразователей.

## 1. Моделирование свойств лигандов и комплексов

Поскольку ключевые критерии эффективности экстрагента (факторы разделения и распределения) имеют экспериментальную природу, необходим поиск их расчетных дескрипторов. В качестве таких дескрипторов были проверены популярные в литературе свойства лиганда: жесткость, энергия предорганизации и электростатический потенциал в точке координации. Для анализа связи свойств лиганда с его экстракционной способностью по отношению к актинидам и лантанидам были рассмотрены корреляции между выбранными свойствами и фактором разделения Am(III)/Eu(III) на примере данных собранных из литературы [2–9]. Установлено, что ни один из этих популярных дескрипторов сам по себе не может служить основным показателем селективности, так как корреляция отсутствует как в целом по базе данных, так и в рамках отдельных работ (Рис. 1). При этом, предсказательная модель машинного обучения на основе алгоритма CatBoost [10], где в качестве признаков описания были выбраны величины рассчитанных дескрипторов вместе с экспериментальными условиями при экстракции, показала высокое качество (Рис. 2 (а)).

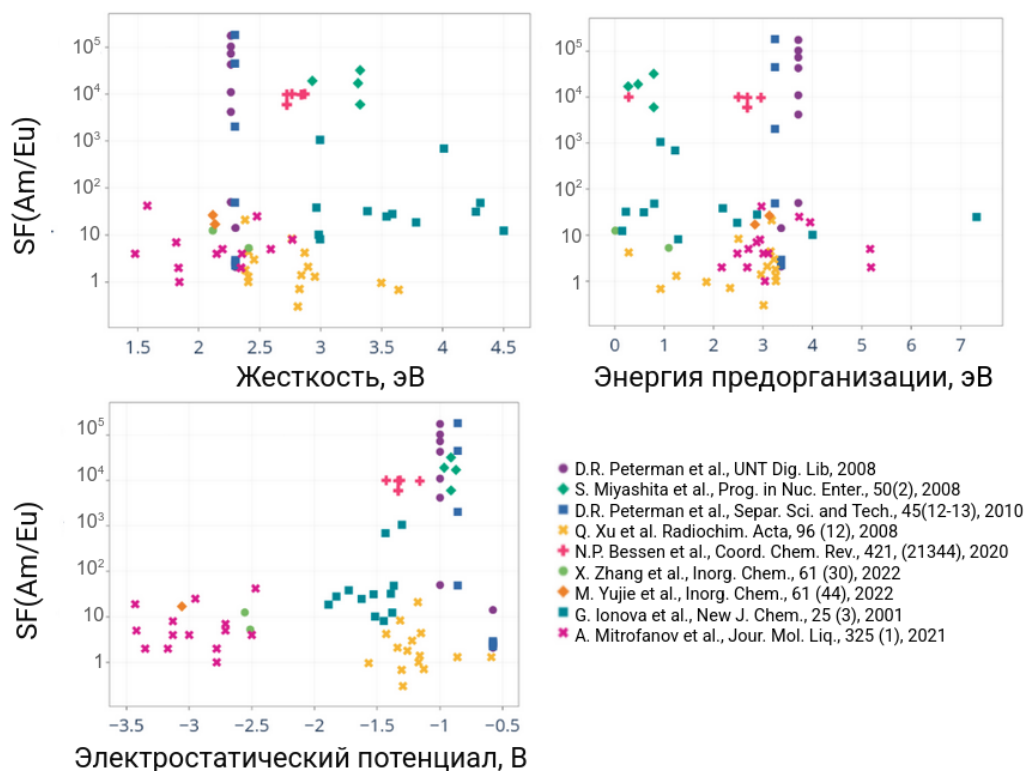


Рисунок 1 — Диаграммы рассеяния между фактором разделения и выбранными дескрипторами.

Используя полученную взаимосвязь между дескрипторами и селективностью мы получили значения важности дескрипторов (Рис. 2 (а)). Анализ важности признаков полученной модели показал, что вклад свойств лиганда в селективность превосходит вклад свойств растворителя, что позволяет проводить первичный дизайн комплексообразователей без учета среды. Наибольшее влияние среди дескрипторов оказывает электростатический потенциал в точке координации.

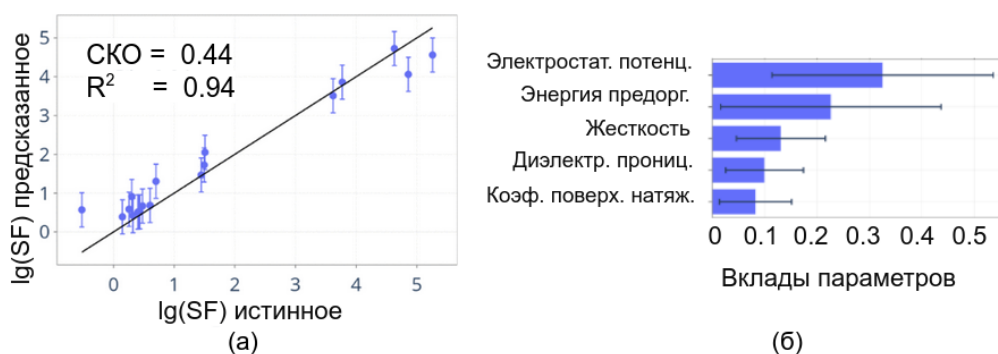


Рисунок 2 — Качество CatBoost-регрессии (а) и вклады наиболее важных параметров в  $SF_{Am(III)/Eu(III)}$  (б).

Отсутствие прямой корреляции рассмотренных свойств с селективностью указывает на необходимость поиска более подходящих дескрипторов, при этом анализ важности признаков подтверждает возможность первичного дизайна комплексообразователей в рамках одной экспериментальной системы. Дальнейшее развитие подхода сдерживается ограниченностью экспериментальных данных по радиохимии, что требует привлечения инструментов, способных нивелировать влияние малых выборок.

## 2. Предсказательные модели для констант устойчивости комплексов

В качестве более подходящего дескриптора для описания комплексообразования и селективности была выбрана константа устойчивости комплекса, и для этих констант была собрана база данных экспериментальных значений (Рис. 3), данные собирались, используя литературные источники [11–14] и базу NIST46. Из собранных данных были убраны неорганические соединения и химически неправильные структуры. 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.

Для повышения точности моделей, предсказывающих свойства комплексов актинидов, была применена методология трансферного обучения. Суть подхода заключается в последовательном обучении двух нейросетевых моделей: сначала донорной — на обширном массиве данных по родственному свойству (в данном случае липофильности, база ChEMBL, 1.6 млн соединений), а затем акцепторной — для целевого свойства актинидов, с переносом весов первых слоев сети, кодирующих молекулярные графы. Тестирование на физико-химических и биомедицинских данных подтвердило эффективность метода: перенос обучения наиболее значим для малых выборок, значительно улучшая метрики качества для наборов из 10–20 соединений, что позволяет перейти от случайного гадания к пригодной для полуквантитативной оценки

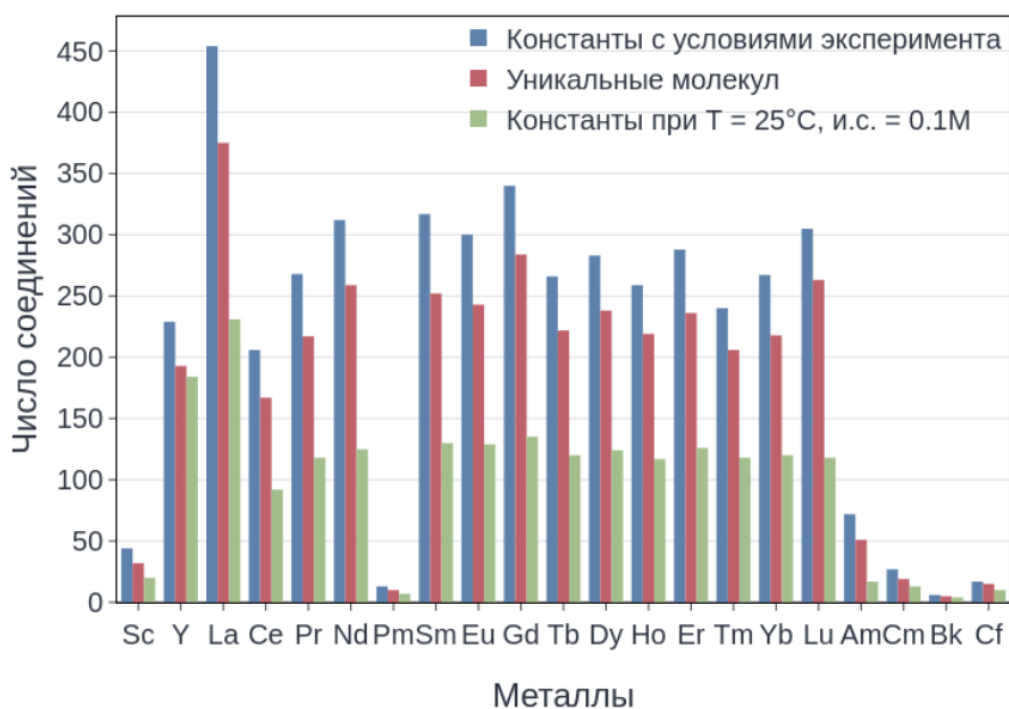


Рисунок 3 — Композиция собранной базы данных комплексообразователей f-элементов.

модели (Рис. 4).

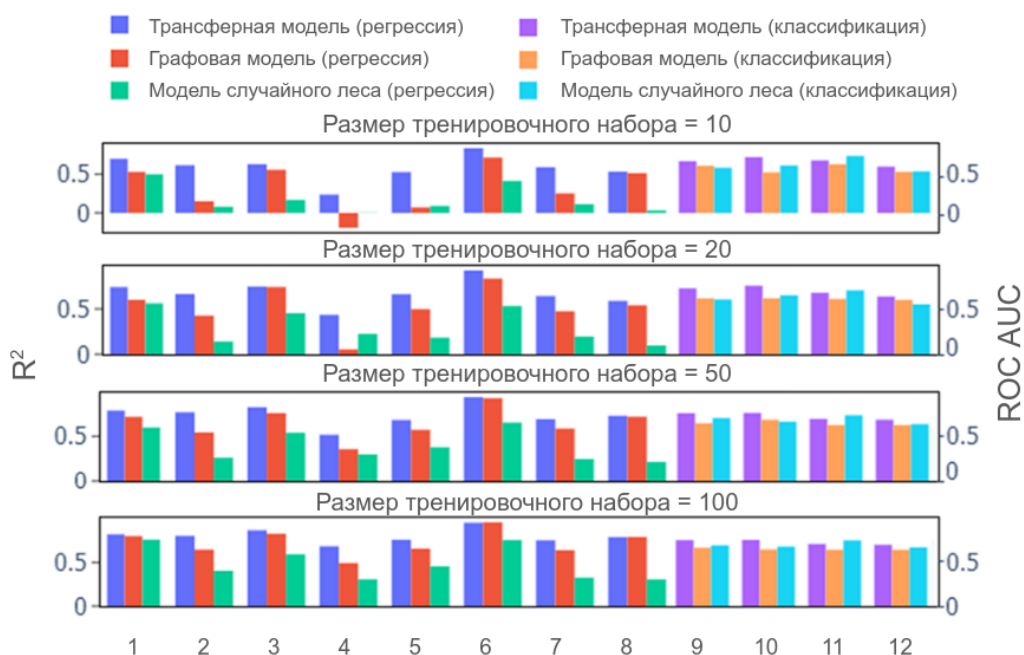


Рисунок 4 — Качество моделей для набора физико-химических и биологических свойств при использовании трансферного обучения.

Далее были построены предсказательные модели для лантанидов и впервые для америция, обеспечивающие качественную оценку потенциальных комплексообразователей с ошибкой на уровне  $2[\lg K]$  (Рис. 5). Однако такой уровень качества является недостаточным, а также не получились модели для других актинидов. Причиной этого является то, что текущая архитектура использует лишь часть доступной

информации: при обучении задействованы только донорный и целевой наборы, тогда как данные по комплексообразованию с другими металлами игнорируются. Это ограничивает точность и область применимости моделей. Следовательно, для полного использования всех накопленных данных необходима разработка новой архитектуры нейронной сети, способной учитывать свойства различных металлов и соответствующие экспериментальные выборки.

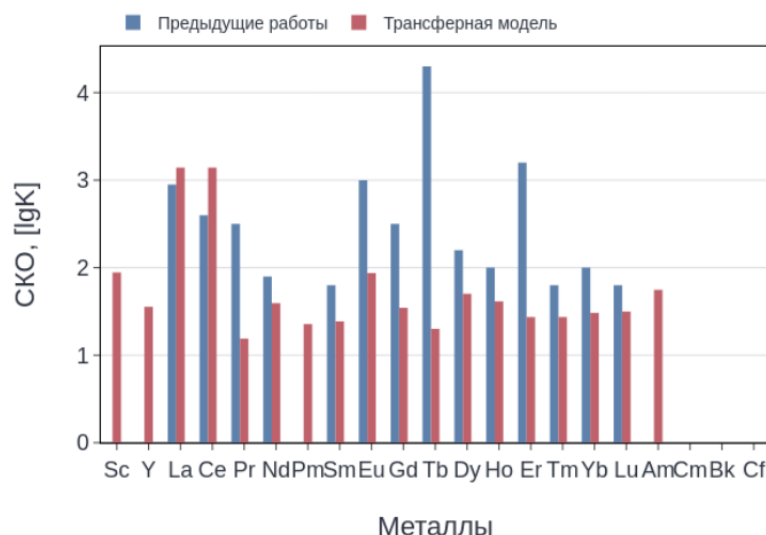


Рисунок 5 — Качество моделей для константы устойчивости при использовании трансферного обучения.

### 3. Гибридные модели для предсказания констант устойчивости

Разработана новая архитектура нейронной сети, которая, несмотря на более высокую вычислительную сложность, позволила улучшить результаты предсказания констант устойчивости. Ключевое отличие от трансферного подхода заключается в том, что модель принимает на вход два объекта — структуру лиганда и векторное представление металла. Важнейшим элементом подхода стало адекватное представление металла, от которого напрямую зависит способность модели различать элементы. Для этого был адаптирован метод *skipatom* [15], изначально применявшийся для описания кристаллических структур. Он позволяет представить любой металл в виде вектора фиксированной длины, кодирующего его физико-химические свойства. Поскольку оригинальная модель не содержала данных по минорным актинидам, мы расширили обучающую базу соединениями с нужными элементами и переобучили *skipatom*, что обеспечило получение векторных представлений для всех интересующих нас металлов. Такая архитектура дает возможность использовать все накопленные экспериментальные данные в рамках единой модели. Качество новой модели значительно превышает трансферные модели и литературные аналоги (Рис. 6).

Новая архитектура также предоставляет возможность предсказывать константу устойчивости для металлов, которых не было в обучающей выборке. Для проверки способности модели предсказывать свойства незнакомых металлов была применена особая

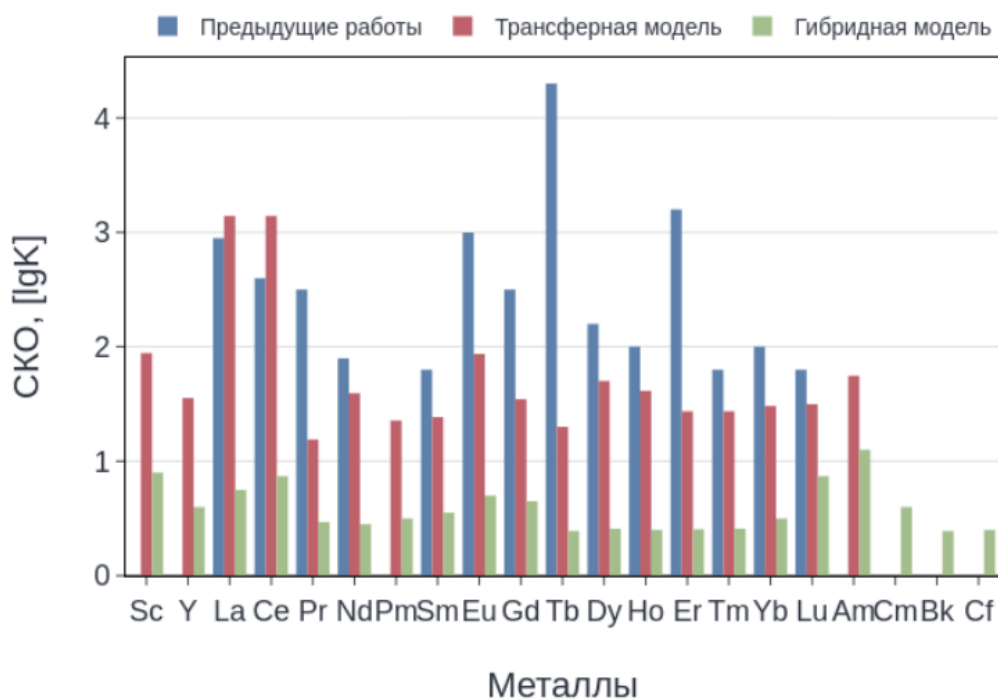


Рисунок 6 — Сравнение предсказательных моделей для константы устойчивости

процедура валидации: каждый целевой элемент полностью исключался из обучающей выборки и использовался только для финального тестирования. Результаты подтвердили, что модель действительно «понимает» периодический закон и способна экстраполировать знания на новые металлы, причем качество предсказаний превосходит все известные литературные модели (Рис. 7). Гибридный подход ожидаемо показал более высокую точность по сравнению с трансферным обучением. При этом для лантана, церия и тория точность оказалась ниже, чем для остальных f-элементов, что объясняется как отличием их свойств, так и большим разнообразием имеющихся данных, создающим более строгие условия оценки. Особого внимания заслуживают высокие показатели моделей для кюрия, берклия и калифорния. Однако эта точность, вероятно, завышена из-за крайне ограниченного набора доступных данных для этих металлов, как по количеству точек, так и по разнообразию структур лигандов.

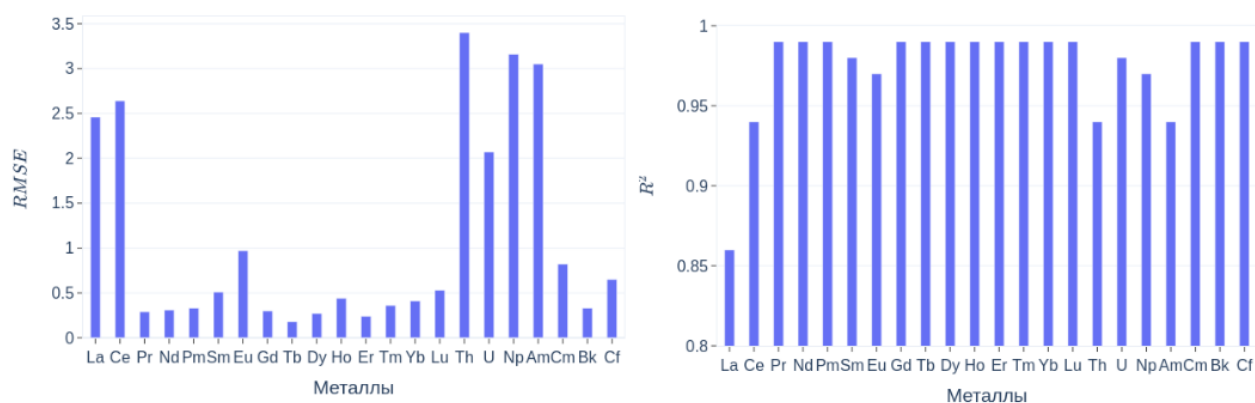


Рисунок 7 — Качество предсказательных моделей для константы устойчивости

Также для построенных моделей провели анализ наиболее важных структурных фрагментов, вносящих наибольший вклад в константу устойчивости. Анализ (табл. 1) подтвердил, что фрагменты с «жесткими» кислородными донорами вносят больший вклад в константы для европия, тогда как для америция и кюрия важны также фрагменты с «мягкими» азотными центрами.

Таблица 1 — Наиболее важные фрагмента при анализе результатов для моделей Eu(III), Am(III), Cm(III)

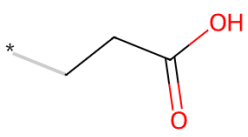
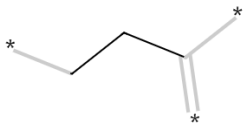
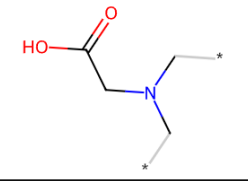
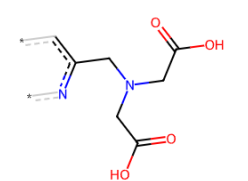
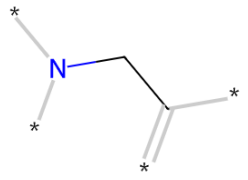
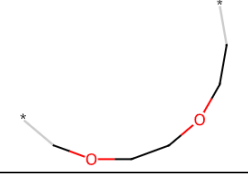
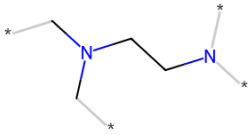
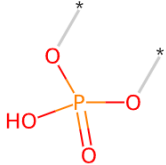
Структура фрагменты	Вклад в lgK
Eu(III)	
	$0.67 \pm 0.06$
	$-1.16 \pm 0.07$
	$3.02 \pm 0.21$
Am(III)	
	$2.26 \pm 0.60$
	$3.64 \pm 1.90$
	$-1.39 \pm 0.80$
Cm(III)	

Таблица 1 – продолжение

Структура фрагменты	Вклад в lgK
	$2.24 \pm 0.92$
	$-0.53 \pm 0.16$

#### 4. Алгоритм поиска равновесной геометрии

Для тестирования алгоритмов глобальной оптимизации с целью их использования в составе генеративной модели мы разработали собственный метод поиска равновесной геометрии и сравнили его с литературными аналогами. Проведено сравнение пяти методов глобальной оптимизации (basin hopping, minima hopping, генетический алгоритм, Tree Parzen Estimator, ETKDG) для задачи поиска равновесной геометрии органических молекул. Показано, что байесовский метод Tree Parzen Estimator при представлении молекулы в виде вектора торсионных углов обеспечивает наилучший баланс между точностью и вычислительными затратами (таблица 2). Генетический алгоритм, несмотря на меньшую эффективность в данном тесте, был выбран как основа для генеративной модели благодаря своей универсальности.

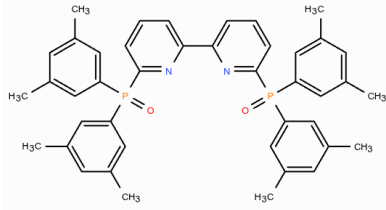
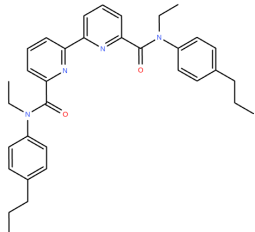
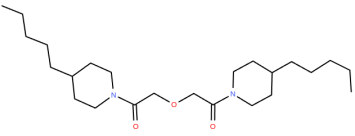
Таблица 2 — Метрики эффективности различных алгоритмов глобальной оптимизации для получения равновесной геометрии молекул

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

## 5. Экспериментальная проверка

Для трёх лигандов (L1–L3, табл. 3) определены коэффициенты распределения и факторы разделения пар Am/Eu и Am/Cm при экстракции из 3 М HNO<sub>3</sub> в F-3. Экстракционные эксперименты проводили в полипропиленовых эппендорфах объемом 1,5 мл. В каждый эппендорф помещали по 0,5 мл органической и водной фаз, после чего смесь интенсивно перемешивали на вортекс-шейкере в термостате при 25±1°C в течение 15 минут (достаточных для установления равновесия). Для ускорения расслаивания образцы центрифугировали (5 мин, 9000 об/мин), после чего отбирали аликвоты по 0,4 мл для анализа. Каждый эксперимент воспроизводили трижды; при значениях коэффициентов распределения (D) в интервале 0,01 < D < 200 относительная погрешность не превышала 15%.

Таблица 3 — Исследованные лиганды

Обозначение	Структура
L1	
L2	
L3	

Содержание радионуклидов определяли методами гамма- и альфа-спектрометрии. Для <sup>243</sup>Am ( $E_\gamma = 59,5$  кэВ) и <sup>152</sup>Eu ( $E_\gamma = 121,1$  кэВ) использовали гамма-спектрометр с высокочистым германиевым детектором GR 3020. Коэффициенты распределения рассчитывали как отношение скорости счета в органической фазе к скорости счета в водной фазе; исходная активность составляла 1500–2500 Бк/мл.

Определение <sup>243</sup>Am и <sup>244</sup>Cm ( $E_\alpha = 5438,2$  и 5901,6 кэВ соответственно) выполняли на альфа-спектрометре 7401 с кремниевым детектором. Для получения тонкослойных образцов проводили пробоподготовку: к аликвоте водной фазы (0,4 мл) добавляли раствор церия и концентрированную фтороводородную кислоту, после осаждения фторидов смесь фильтровали через мембранный фильтр (0,1 мкм), который затем высушивали

и помещали в камеру спектрометра. Коэффициенты распределения рассчитывали по формуле  $D = (I_0 - I)/I$ , где  $I_0$  и  $I$  — скорости счета исходной и равновесной водной фазы (исходная активность 1000–2000 Бк/мл). Фактор селективности находили как отношение коэффициентов распределения. Результаты приведены в табл. 4 и 5.

Таблица 4 — Экстракция Am(III) и Eu(III) (3 М HNO<sub>3</sub>, F-3)

Лиганд	C(L), М	D(Am)	D(Eu)	SF(Am/Eu)
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 5 — Экстракция Am(III) и Cm(III) (3 М HNO<sub>3</sub>, F-3)

Лиганд	C(L), М	D(Am)	D(Cm)	SF(Am/Cm)
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

Предсказанные гибридной моделью отношения констант устойчивости  $K_{Am}/K_{Eu}$  и  $K_{Am}/K_{Cm}$  корректно ранжируют лиганды по селективности (табл. 6), хотя абсолютные значения отличаются от экспериментальных факторов разделения, что объясняется влиянием органической фазы, не учтённым в модели.

Таблица 6 — Предсказанные константы устойчивости и отношения констант

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

Результаты экспериментальной проверки подтвердили наши предположения о способности модели качественно оценивать селективность лигандов, соответственно ее можно использовать далее для создания генеративной модели.

## 6. Генеративная модель

На основе разработанных моделей для констант устойчивости был проведен генеративный поиск перспективных комплексообразователей с использованием генетического алгоритма. В качестве целевой функции выступало отношение предсказанных констант для пар Am/Eu и Am/Cm, представляющих наибольший практический интерес. В качестве генеративной модели мы применили

модифицированный вариант известного в литературе подхода JANUS [16]. Сам подход заключается в использовании двух параллельных генетических процессов, один из которых отвечает за широкий поиск по химическому пространству, а второй за оптимизацию удачных молекул. Также два эти процессе периодически обмениваются подходящими молекулами для улучшения качества, схема работы алгоритма представлена на рисунке 8.

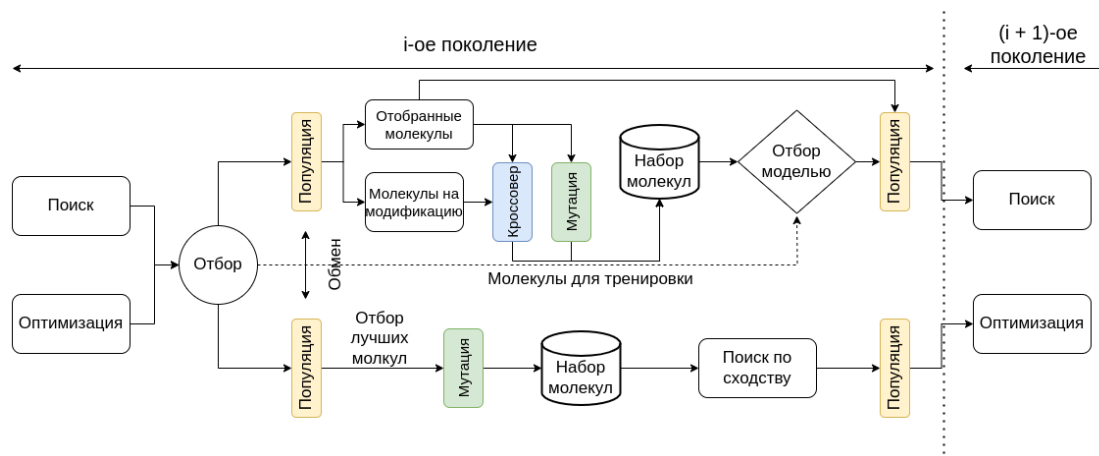
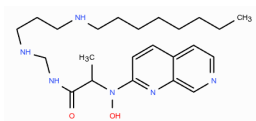
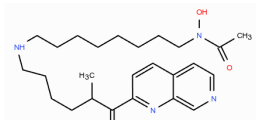
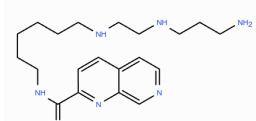
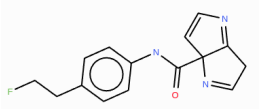
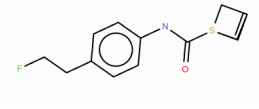
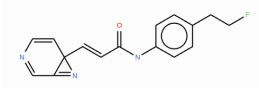


Рисунок 8 — Схема работы генеративной модели

Первичная проверка на случайной начальной выборке показала ожидаемую динамику: по мере смены популяций целевой параметр (селективность) закономерно смещался в сторону увеличения, особенно в «локальной» выборке, отвечающей за тонкую настройку структуры. Генерация выявила как достоинства подхода, так и его ограничения. Алгоритм успешно предложил разнообразные структуры с высокими целевыми показателями, однако, будучи ориентированным исключительно на селективность, он генерировал соединения с сомнительной синтетической доступностью (например, длинные алифатические фрагменты или напряженные гетероциклы, табл. 7).

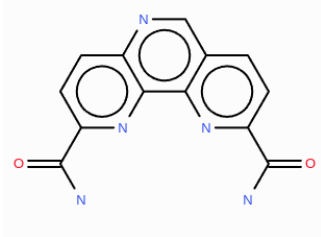
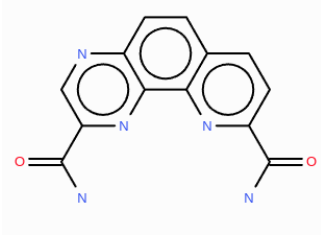
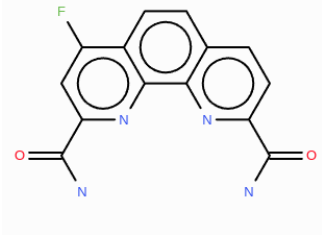
Таблица 7 — Результаты генераций на основе случайных начальных выборок

№	Лиганд	$lgK_{Eu(III)}$	$lgK_{Am(III)}$	$lgK_{Cm(III)}$	$\frac{K_{Am(III)}}{K_{Eu(III)}}$	$\frac{K_{Am(III)}}{K_{Cm(III)}}$
1		9.81	5.89	9.59	0.0001	0.0002
2		6.80	5.27	6.89	0.03	0.0004
3		7.34	6.43	11.20	0.12	$1.70 * 10^{-5}$

4		7.27	10.1	7.04	671.33	1134.58
5		5.27	8.67	5.24	2484.40	2687.19
6		6.81	8.44	5.85	42.62	382.31

Для повышения практической значимости в начальную выборку были добавлены каркасы известных экстрагентов. В этом режиме модель продемонстрировала способность к «переоткрытию» тонких химических эффектов: увеличивая селективности лиганда, она независимо предлагала введение электроноакцепторных групп в фенантролиновое ядро, что совпадает с известными литературными данными. Таким образом, сочетание генетического алгоритма с предсказательной моделью позволяет генерировать перспективные структуры, но обязательным условием практической применимости является учет синтетической доступности на этапе дизайна (Рис. 8).

Таблица 8 — Примеры генераций на основе фенантролинов

№	Лиганд	$\lg K_{Eu(III)}$	$\lg K_{Am(III)}$	$\lg K_{Cm(III)}$	$\frac{K_{Am(III)}}{K_{Eu(III)}}$	$\frac{K_{Am(III)}}{K_{Cm(III)}}$
1		3.48	6.62	6.61	1372.40	1.01
2		3.47	6.63	6.61	1455.12	1.05
3		3.68	6.93	7.31	1775.67	0.42

## 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  $0.44 \lg(\text{SF})$ . 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 не изученными 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  $0,44 [\lg(\text{SF})]$ .
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 актининов со средней ошибкой  $0,4 [\lg K]$ , 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)/Sm(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 публикаций по теме диссертации

Статьи, опубликованные в рецензируемых научных изданиях, рекомендованных для защиты в МГУ имени М.В. Ломоносова по специальности и отрасли наук:

1. Solov'ev V.P., Predictive Models for HOMO and LUMO Energies of N-Donor Heterocycles as Ligands for Lanthanides Separation / Solov'ev V.P., Ustynyuk Yu.A., Zhokhova N.I., Karpov K.V. // *Molecular Informatics*. - 2018. - Vol. 37 - № 11. - 1800025. EDN: EMNYCD (Импакт-фактор 3.1(JIF), 0,75 п.л./ 50%)
2. Karpov K.V., Size Doesn't Matter: Predicting Physico- or Biochemical Properties Based on Dozens of Molecules / Karpov K.V., Mitrofanov A.A., Korolev V., Tkachenko V. // *The Journal of Physical Chemistry Letters*. - 2021. - Vol. 12 - № 38. - pp. 9213–9219. EDN: XXIZCO (Импакт-фактор 4.7(JIF), 0.7 п.л./ 70%)
3. Andreadi N.G., Tree Parzen estimator for global geometry optimization: A benchmark and database of experimental gas-phase structures of organic molecules / Andreadi N.G., Zankov D., Karpov K.V., Mitrofanov A.A. // *Journal of Computational Chemistry*. - 2022. - Vol. 43 - № 21. - pp. 1434–1441. EDN: QHHK LX (Импакт-фактор 4.8(JIF), 0.6 п.л. / 40%)
4. Karpov K.V., On F-Element Extractant Design Using HSAB Theory / Karpov K.V., Mitrofanov A.A., Kalmykov S.N. // *Inorganic Chemistry*. - 2025. - Vol. 64 - № 24. - pp. 11925–11930. EDN: BWTOWU (Импакт-фактор 4.7(JIF), 0.8 п.л. 70%)

## Список цитируемой в автореферате литературы

1. *Agency I. A. E.* Energy, Electricity and Nuclear Power Estimates for the Period up to 2050 : tech. rep. / International Atomic Energy Agency. — 2022. — Pp. 1–137.
2. Selective Extraction of Minor Actinides from Acidic Media Using Symmetric and Asymmetric Dithiophosphinic Acids / D. R. Peterman [et al.] // *Separation Science and Technology*. — 2010. — July. — Vol. 45, no. 12/13. — Pp. 1711–1717. — ISSN 0149-6395.

3. Separation of Am(III) from Eu(III) by extraction based on in situ extractant formation of dithiocarbamate derivatives / S. Miyashita [et al.] // *Progress in Nuclear Energy*. — 2008. — Mar. — Vol. 50, no. 2. — Pp. 499–503. — ISSN 0149-1970.
4. Extraction of Am(III) and lanthanides(III) with organo dithiophosphinic acids / Q. Xu [et al.] // *Radiochimica Acta*. — 2008. — Dec. — Vol. 96, no. 12. — Pp. 771–779. — ISSN 2193-3405.
5. Sulfur donating extractants for the separation of trivalent actinides and lanthanides / N. P. Bessen [et al.] // *Coordination Chemistry Reviews*. — 2020. — Oct. — Vol. 421. — P. 213446. — ISSN 0010-8545.
6. Advancing the Am Extractant Design through the Interplay among Planarity, Preorganization, and Substitution Effects / X. Zhang [et al.] // *Inorganic Chemistry*. — 2022. — Aug. — Vol. 61, no. 30. — Pp. 11556–11570. — ISSN 0020-1669.
7. 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 [et al.] // *Inorganic Chemistry*. — 2022. — Nov. — Vol. 61, no. 44. — Pp. 17911–17923. — ISSN 0020-1669.
8. Mechanism of trivalent actinide/lanthanide separation using synergistic mixtures of di(chlorophenyl)dithiophosphinic acid and neutral O-bearing co-extractants / G. Ionova [et al.] // *New Journal of Chemistry*. — 2001. — Jan. — Vol. 25, no. 3. — Pp. 491–501. — ISSN 1369-9261.
9. An(III)/Ln(III) solvent extraction: Theoretical and experimental investigation of the role of ligand conformational mobility / A. Mitrofanov [et al.] // *Journal of Molecular Liquids*. — 2021. — Mar. — Vol. 325. — P. 115098. — ISSN 0167-7322.
10. CatBoost: unbiased boosting with categorical features / L. Prokhorenkova [et al.]. — 01/2019. — arXiv:1706.09516 [cs].
11. Probing the difference in covalence by enthalpy measurements: a new heterocyclic N-donor ligand for actinide/lanthanide separation / Y. Yang [et al.] // *Dalton Transactions*. — 2015. — Vol. 44, no. 19. — Pp. 8959–8970.
12. Deep Learning Insights into Lanthanides Complexation Chemistry / A. Mitrofanov [et al.] // *Molecules*. — 2021. — Vol. 26, no. 11. — P. 3237.
13. BTBPs versus BTPPhens: some reasons for their differences in properties concerning the partitioning of minor actinides and the advantages of BTPPhens / F. Lewis [et al.] // *Inorganic Chemistry*. — 2013. — Vol. 52, no. 9. — Pp. 4993–5005.
14. 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 [et al.] // *Radiochimica Acta*. — 2016. — Vol. 104, no. 8. — Pp. 555–565.
15. *Antunes L., Grau-Crespo R., Butler K.* Distributed representations of atoms and materials for machine learning // *npj Computational Materials*. — 2022. — Vol. 8, no. 1. — Pp. 1–9.

16. *Nigam A., Pollice R., Aspuru-Guzik A.* Parallel tempered genetic algorithm guided by deep neural networks for inverse molecular design // *Digital Discovery*. — 2022. — Vol. 1, no. 4. — Pp. 390–404.

## Благодарности

Автор выражает глубокую признательность своему научному руководителю - Митрофанову Артему Александровичу за постоянное внимание к работе, профессиональное наставничество, ценные знания и заразительную увлечённость наукой.

За неоценимую помощь в проведении экспериментальных исследований и обработке полученных результатов автор выражает благодарность Евсюниной Марии Валерьевне и Герасимову Михаилу Алексеевичу.

За высокую вовлечённость, интеллектуальную поддержку и содержательные научные дискуссии автор искренне благодарит коллектив лаборатории интеллектуального химического дизайна. Особую благодарность выражает Смирновой А.А., Елисееву А.А. и Королеву В.В.

Автор благодарит Борисову Н.Е., Орлову М.А. и Гольцова А.Н. за внимательное рецензирование работы и ценные замечания, способствовавшие её улучшению.

Автор благодарит коллег кафедры радиохимии за энергичную атмосферу и возможность осуществлять научную работу, а также, в частности, Петрова Владимира Геннадиевича и Матвеева Петра Игоревича за интерес к работе и ценные замечания.

Отдельная и самая тёплая благодарность — семье за неизменную веру, терпение и поддержку на всех этапах выполнения работы.

Работа выполнена при поддержке Российского научного фонда (грант № 23-73-30006)