

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Самарский государственный технический университет»

На правах рукописи

Шевченко Александр Петрович

**ТЕОРИЯ И МЕТОДЫ КОМПЬЮТЕРНОГО ГЕОМЕТРИКО-
ТОПОЛОГИЧЕСКОГО АНАЛИЗА И ПРОГНОЗИРОВАНИЯ СТРОЕНИЯ
И ФИЗИЧЕСКИХ СВОЙСТВ КООРДИНАЦИОННЫХ СОЕДИНЕНИЙ**

1.4.4. Физическая химия

Автореферат

диссертации на соискание ученой степени

доктора химических наук

Самара - 2025

Работа выполнена в Международном научно-исследовательском центре по теоретическому материаловедению (МНИЦТМ) федерального государственного бюджетного образовательного учреждения высшего профессионального образования "Самарский государственный технический университет"

Научный консультант: **Блатов Владислав Анатольевич**
доктор химических наук, профессор

Официальные оппоненты: **Еремин Николай Николаевич**
доктор химических наук, член-корреспондент РАН, декан геологического факультета, зав. кафедрой кристаллографии и кристаллохимии, Московский государственный университет им. М. В. Ломоносова

Киселева Надежда Николаевна
доктор химических наук, зав. лабораторией полупроводниковых материалов, Институт металлургии и материаловедения им. А.А. Байкова РАН

Адонин Сергей Александрович
доктор химических наук, профессор РАН, зам. директора по научной работе, Иркутский институт химии им. А.Е. Фаворского Сибирского отделения РАН

Ведущая организация: **ФГБУН Институт Элементоорганических соединений им. А. Н. Несмеянова РАН, г. Москва**

Защита состоится 25 марта 2025 г. в 14:00 на заседании диссертационного совета 24.2.377.03 при ФГБОУ ВО «Самарский государственный технический университет» по адресу: 443100, г. Самара, ул. Молодогвардейская, 244, ауд. 200.

Отзывы по данной работе в двух экземплярах, заверенные гербовой печатью, просим направлять по адресу: Россия, 443100, г. Самара, ул. Молодогвардейская, 244, Главный корпус, на имя ученого секретаря диссертационного совета 24.2.377.03; тел./факс (846) 3322122; e-mail: orgchem@samgtu.ru. В отзыве просим указывать почтовый адрес, номер телефона, электронную почту, наименование организации, должность, шифр и наименование научной специальности.

С диссертацией можно ознакомиться в библиотеке Самарского государственного технического университета (ул. Первомайская, 18) и на сайте диссертационного совета 24.2.377.03 <https://d24237703.samgtu.ru>

Автореферат разослан «___» _____ 2025 г.

Ученый секретарь
диссертационного совета 24.2.377.03,
кандидат химических наук, доцент

Ивлева Е.А.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования и степень ее разработанности. Возможности современной экспериментальной техники и программного обеспечения позволяют быстро и с высокой точностью определять кристаллическую структуру практически любого химического соединения. Экспериментальная структурная информация аккумулируется в электронных хранилищах, наиболее известные из которых – Кембриджская структурная база данных (CSD), База данных неорганических структур (ICSD), Файл порошковых данных (PDF4), Банк белковых данных (PDB), База кристаллических данных Пирсона (PCD). На начало 2024 года они содержат сведения о строении более 2 млн. кристаллических веществ.

Несмотря на уже длительную историю своего существования (первая из указанных баз данных, CSD, была основана в 1965 г.) электронные хранилища в течение длительного времени использовались только как справочники. Ситуация качественно изменилась, когда в 2011 г. в США была принята программа «Геном материалов», основной задачей которой является ускорение разработки новых материалов за счет внедрения компьютерных методов и инструментов прогнозирования. В результате ее выполнения планируется сократить срок разработки новых материалов с 10-15 лет до 2-3 лет и существенно уменьшить стоимость этого процесса. Для осуществления указанных задач необходимо создание новых методов моделирования строения и свойств химических веществ, аккумулирование накопленной химической и структурной информации в единых электронных базах данных, использование современных методов машинной обработки и машинного обучения для анализа собранных данных в целях извлечения из них новых полезных знаний («вторичной» информации). Эта информация, собранная в базы знаний, станет основанием будущих систем искусственного интеллекта (ИИ), способных с высокой точностью прогнозировать структуру и свойства веществ и материалов.

Именно разработка баз знаний, соответствующих методов и программного инструментария становится общемировым трендом, в русле которого лежит данная работа. В настоящее время существует лишь несколько примеров баз знаний по кристаллохимии и материаловедению, а системы ИИ все еще находятся в стадии разработки. Одной из таких баз знаний является Mogul, которая интегрирована в систему CSD. Однако разработка баз знаний существенно ограничивается необходимостью подготовки надежных исходных кристаллографических данных, создания новых дескрипторов для характеристики особенностей кристаллической структуры и ее свойств, а также разработки специальных методов и инструментов для создания баз знаний из исходных наборов

кристаллографических данных. CSD является хорошим примером тщательной проработки этих проблем в течение длительного времени, однако даже информация из CSD требует дополнительной подготовки для машинного анализа. База знаний Mogul содержит параметры только локального окружения молекул в кристаллах, но не рассматривает свойства структуры как целого, в частности, способ связанности структурных строительных блоков. Кроме того, Mogul ориентирована только на молекулярные кристаллы и не содержит дескрипторов для полимерных структур, как неорганических, так и координационных. Разработка новых дескрипторов, в свою очередь, требует развития теоретического базиса описания и моделирования кристаллических веществ.

Целью работы являлась разработка теоретических моделей строения кристаллических веществ, алгоритмов и методов автоматизированной обработки объемных выборок кристаллографических данных для поиска закономерностей «состав-структура-свойство» кристаллов, а также прогнозирования их строения и физических свойств на примере координационных соединений (КС) и их важнейшего подкласса, металл-органических координационных полимеров (МОКП).

Для достижения указанной цели были поставлены следующие **задачи**:

- разработка теоретических основ геометрического и топологического моделирования атомных ансамблей и свободного пространства в кристаллических структурах с помощью их представления в виде периодических графов и разбиения Вороного;
- создание системы геометрико-топологических дескрипторов кристаллической структуры и ее структурных строительных блоков (ССБ) для прогнозирования свойств и поиска закономерностей строения простых веществ, твердых электролитов, КС и МОКП;
- разработка универсальных алгоритмов и программного обеспечения для автоматизированного расчета геометрико-топологических дескрипторов в КС и МОКП любого химического состава и строения;
- разработка методов машинного поиска корреляций между геометрико-топологическими дескрипторами, химическим составом КС и его физическими свойствами, а также создание системы баз знаний, содержащих найденные корреляции;
- разработка на основе корреляций, содержащихся в созданных базах знаний, схем прогнозирования для дизайна КС, обладающих заданной структурой и физическими свойствами.

Положения, выносимые на защиту:

- универсальный геометрико-топологический подход к моделированию и анализу кристаллической структуры, основанный на представлении ее атомов в виде полиэдров Вороного или Ван-дер-Ваальсовых полиэдров, связанных в трехмерную периодическую сетку, и позволяющий проводить классификацию систем химических связей, поиск взаимосвязей «химический состав – структура – свойство» и прогнозирование новых структур в соответствии со строгими алгоритмами;
- совокупность геометрико-топологических дескрипторов и методов их расчета для описания структурной роли атомов и ССБ, а также строения и физических свойств кристаллических структур;
- правила, принципы и закономерности, сформулированные на основании геометрико-топологического подхода, в том числе: классификация лигандов по форме их координационных фигур, схема машинного анализа структурных данных на основе геометрико-топологических дескрипторов, численный критерий оценки жесткости катионов в структуре КС и МОКП, условия миграции катиона в структуре твердого электролита, взаимосвязи радиуса сферического домена атомов в структуре простых веществ с длинами волн де Бройля валентных электронов, рассчитанными из потенциалов ионизации изолированных атомов;
- топологические алгоритмы, программное обеспечение и методы анализа кристаллических веществ, основанные на понятии о множественности топологических представлений кристаллической структуры и способов ее топологического упрощения;
- базы знаний химических, геометрических и топологических дескрипторов простых веществ, а также известных на настоящий момент структур КС и МОКП, библиотеки и методы сборки их ССБ, рассчитанные на основе геометрико-топологического подхода, реализованного в программном комплексе ToposPro и интернет-сервисах TopCryst;
- прогностическая модель для оценки степени окисления атомов-комплексобразователей в кристаллической структуре;
- схема прогнозирования и дизайна структуры МОКП, опирающаяся на корреляции между его локальными и глобальными топологическими дескрипторами;
- метод анализа свободного пространства в кристаллической структуре, основанный на представлении атомов в виде полиэдров Вороного и

Ван-дер-Ваальсовых полиэдров и позволяющий прогнозировать сорбционные свойства и ионную проводимость кристалла.

Научная новизна. Создано новое междисциплинарное направление, объединяющее методы и модели кристаллохимии и материаловедения с информационными технологиями и ориентированное на создание универсальных систем ИИ, пригодных для прогнозирования структуры и свойств кристаллических веществ любой природы. Впервые объединены топологический и геометрический подходы при анализе и моделировании структур КС и МОКП. Развита теория топологических представлений кристаллических структур, в которой формализованы методы выделения ССБ в зависимости от их химической природы и топологических свойств для изучения и контроля механизмов их сборки. Предложена схема прогнозирования ряда топологических свойств структур КС и МОКП, в том числе размерности и топологии их базовой сетки на основании разработанных количественных методов определения формы координационных фигур ССБ. С использованием современных алгоритмов машинного обучения разработана прогностическая модель для оценки степени окисления атомов-комплексобразователей в кристаллической структуре, исходя из геометрических дескрипторов их полиэдров Вороного, а также положения в Периодической системе Д. И. Менделеева. Впервые на примере кристаллов простых веществ исследованы и проанализированы взаимосвязи радиуса сферического домена атомов с длинами волн де Бройля валентных электронов, рассчитанными из потенциалов ионизации изолированных атомов. Найденные корреляции позволяют объяснить геометрическую модель атома в структуре кристаллов, как упаковку деформируемых, но несжимаемых сфер, с позиции электронного строения атома.

Теоретическая и практическая значимость работы. Разработанный геометрико-топологический подход к анализу и прогнозированию кристаллических структур и их свойств активно используется исследователями во всем мире. Созданные базы данных топологических типов и ССБ, а также разработанные на их основе сервисы, востребованы при топологическом и кристаллохимическом описании новых синтезируемых КС и МОКП, при поиске закономерностей строения кристаллических веществ, анализе свободного пространства и скрининге структурной информации для отбора потенциальных твердых электролитов. На конец 2023 г. программный комплекс ToposPro, в котором реализован указанный подход, используют более 7800 исследователей из 103 стран мира; ежегодно публикуются около 500 научных статей, в которых ToposPro применяется для анализа кристаллических веществ. Созданные топологические базы данных и

использующие их сервисы, доступны на сайте <https://topcryst.com>; на конец 2023 г. указанные сервисы были применены для анализа более 33 тыс. кристаллических структур пользователями из 74 стран мира.

Методология и методы исследования. В качестве экспериментальных методов исследования в диссертационной работе использовались методы геометрико-топологического анализа кристаллоструктурных данных большого объема. Математическая обработка рассчитанных данных осуществлялась методами машинного обучения, такими как случайный лес, метод опорных векторов, дерево решений, K-ближайших соседей. Для обработки экспериментальных данных и построения моделей интеллектуального анализа использовалась среда разработки Python's Integrated Development Environment, а также программные пакеты ToposPro, Excel, PostgreeSQL.

Личный вклад автора. Выбор темы, постановка задач и целей исследования, обсуждение и обобщение всех полученных результатов, формулировка научных положений и выводов, которые выносятся на защиту, принадлежат лично автору. Все вычислительные эксперименты, описанные в диссертации, выполнены автором в сотрудничестве с коллегами МНИЦТМ.

Степень достоверности и апробация результатов. Достоверность результатов работы обеспечивается строгостью примененных методов математического анализа и машинного обучения, а также большими объемами выборок экспериментальных структурных данных, используемых для машинного обучения прогностических моделей. Достоверность представленных научных результатов подтверждается воспроизводимостью экспериментальных и расчетных данных, а также их соответствием результатам работ других авторов.

Основные положения работы и результаты докладывались на Национальных кристаллохимических конференциях (Черноголовка 2003, Суздаль 2016, Суздаль 2018, Чегет 2021, Нальчик 2024), II Международной научной конференции «Наука будущего» (Казань 2016), IUCr 2017 (Hyderabad, India), Intermetallics 2017 (Bad Staffelstein, Germany), IV Международной конференции по информационным технологиям и нанотехнологиям (Самара 2018), 9-й Международной научной конференции «Numerical Geometry, Grid Generation and Scientific Computing» (Москва 2018), XXI Менделеевском съезде по общей и прикладной химии (Санкт-Петербург 2019), 1-м Международном симпозиуме по компьютерному прогнозированию структуры и перспективным материалам (Xi'an 2019, China), XII Международной научной конференции «Кинетика и механизм кристаллизации. Кристаллизация и материалы нового поколения» (Иваново

2023), 7-й Международной конференции “Глубокое обучение в вычислительной физике” (Санкт-Петербург 2023), XXV Международной конференции по анализу и управлению данными в областях с интенсивным использованием данных (Москва 2023).

Публикации. По материалам диссертации опубликованы 2 главы в монографиях и 35 статей в рецензируемых журналах. Получены 18 свидетельств о регистрации программ для ЭВМ и баз данных.

Объем и структура работы. Диссертационная работа состоит из введения, обзора литературы, экспериментальной части, обсуждения результатов, выводов, заключения, списка литературы (284 источника) и приложения (изложенного на 15 страницах и содержащего 12 таблиц, 2 рисунка). Текст диссертационной работы изложен на 204 страницах, включает 36 таблиц и 91 рисунок.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **Введении** обоснована актуальность темы исследования, сформулирована цель работы, определены задачи, которые необходимо решить для достижения поставленной цели, описаны научная новизна и области практического применения полученных результатов.

В **Первой главе** проведен анализ литературных данных, относящихся к теме диссертации. Рассмотрены всемирные кристаллоструктурные базы данных: Кембриджская структурная база данных (CSD), База данных неорганических структур (ICSD), а также специфика информации, которая собирается в них. Начиная с 2011 года, в CSD и ICSD ежегодно добавляется 50-65 тыс. записей, а общее количество записей в них на декабрь 2022 года составило 1.4 млн, из которых около 700 тыс. - координационные соединения. Огромное количество накопленной структурной информации усложняет ее обработку, но приводит к появлению новых возможностей для поиска правил и закономерностей в данных.

Кристаллические структуры являются сложными объектами, которые могут содержать сотни атомов в независимой части (примитивной ячейке). В частности, структуры сольвата бис(гексафторфосфата) бис(2,9-диметил-1,10-фенантролин) - (пиразино(2,3-f) (1,10) фенантролин)рутения - $[C_{42}H_{32}N_8Ru]^{2+}(PF_6^-)_2 \cdot 0.11(C_4H_{10}O) \cdot 1.22(C_3H_6O)$ (UKEBUJ¹) и полиоксометаллата $Li_{74}K_{56}H_{14}[(P_8W_{48}O_{184})((P_2W_{14}Mn_4O_{60})(P_2W_{15}Mn_3O_{58})_2)_4](H_2O)_{680}$ (422376) являются рекордсменами по числу симметрично независимых атомов – соответственно 998 и 1561. Для описания и анализа таких структур необходимо использовать упрощенные модели, преимущества и недостатки

¹ Здесь и далее приводятся справочные коды определений структуры кристаллов в CSD или ICSD.

которых рассмотрены в диссертации. Особое внимание уделено геометрическим и топологическим моделям, которые затем использованы для расчета дескрипторов и при поиске закономерностей в конкретных классах кристаллических структур.

В наиболее известных *геометрических моделях* атомы представляют в виде жестких или мягких (деформируемых) сфер. В рамках первой модели атом аппроксимируют сферой фиксированного радиуса, что отражает близость расстояний А–В между заданной парой атомов А и В в разных соединениях. В рамках модели мягких сфер допускается возможность деформации атомов, следствием которой является непостоянство расстояний А–В. При постоянном давлении и незначительном изменении температуры максимальная деформация атома ограничена некоторым несжимаемым объемом. Если химический сорт окружающих атомов фиксирован, то независимо от их числа, объем, приходящийся на центральный атом, будет постоянен. Модель мягких сфер называется также полиэдрической моделью, так как она аппроксимирует полиэдрами Вороного атомы в кристаллическом поле (рис. 1 слева).

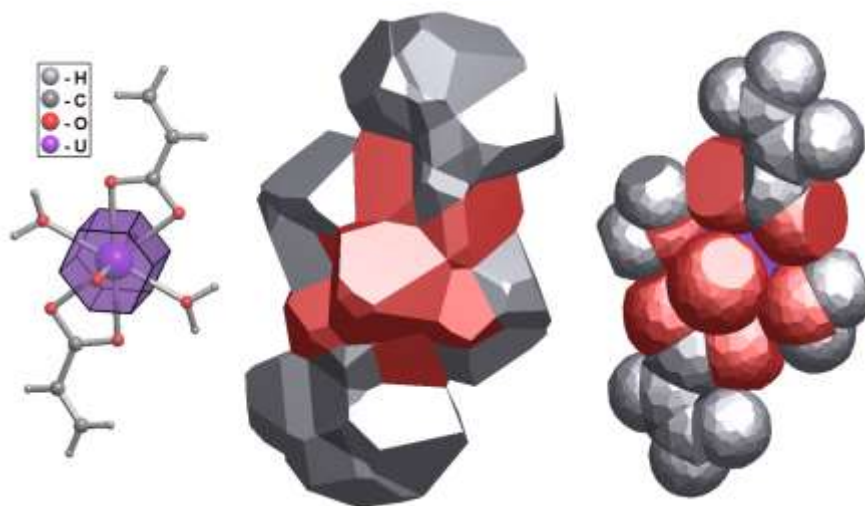


Рис. 1. Молекулярный фрагмент структуры $\text{UO}_2(\text{C}_3\text{H}_3\text{O}_2)_2(\text{H}_2\text{O})_2$ (NAFRIY², слева), где $\text{C}_3\text{H}_3\text{O}_2^-$ - акрилат анион, с полиэдром Вороного атома урана и представления этого фрагмента в виде полиэдров Вороного (в центре) или Ван-дер-Ваальсовых полиэдров (справа).

Полиэдром Вороного атома называют область (домен) кристаллического пространства, которая ограничена плоскостями, проведенными через середины отрезков, соединяющих центральный атом с атомами ближайшего окружения, перпендикулярно отрезкам. Все внутренние точки полиэдра Вороного находятся ближе к центру данного атома, чем к любому другому атому структуры. Важной особенностью

² Klepov V.V., Vologzhanina A.V., Alekseev E.V. et al. CrystEngComm 2016, 18, 1723-1731.

полиэдра Вороного является то, что его вершины наиболее удалены от ближайших атомов структуры и, таким образом, могут рассматриваться как центры пустот структуры, а ребра соответствуют каналам между пустотами.

Топологическим представлением (моделью) кристаллической структуры является периодический граф, вершинам которого отвечают структурные строительные блоки (ССБ), а ребра отождествлены с различными видами межатомных связей (Рис. 2). В качестве ССБ могут быть рассмотрены атомы, молекулы или молекулярные фрагменты. Если каждый атом кристаллической структуры представляет собой ССБ, то это представление отвечает *полной структурной сетке*. При упрощении структурной сетки посредством замены и/или удаления некоторой части вершин и/или ребер получают различные *базовые сетки*, которые характеризуют связность ССБ разного типа в кристалле и описывают глобальную топологию кристалла в соответствующем представлении. Так, например, комплекс бис(5-хлороникотинато)-меди(II) (Рис. 3 слева) может быть охарактеризован двумя различными представлениями: *стандартным*, в котором ССБ являются атомы меди и лиганды 5-хлороникотината (Рис. 3 вверху справа), и *кластерным*, где в качестве ССБ рассматриваются биметаллические кластеры - «китайские фонарики» (Рис. 3 внизу справа).

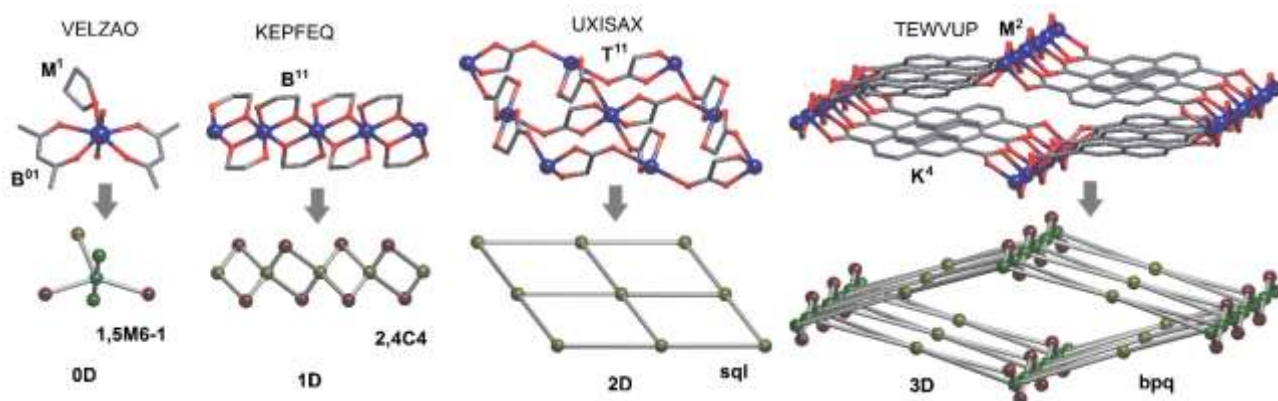


Рис.2. Фрагменты координационных сеток с размерностью 0D, 1D, 2D и 3D (вверху) и соответствующих им базовых сеток с названиями топологических типов (внизу) в кристаллических структурах (слева направо): (диоксо)-бис(пентане-2,4-дионато)-(тетрагидрофуран-О)-уран(VI) $[\text{UO}_2(\text{C}_5\text{H}_7\text{O}_2)_2(\text{C}_4\text{H}_8\text{O})]$ (VELZAO), бис(этилендиолато)-титан(IV) $[\text{Ti}(\text{C}_2\text{H}_4\text{O}_2)_2]$ (KEPFEQ), бис(гликолято)-меди(II) $[\text{Cu}(\text{C}_2\text{H}_3\text{O}_3)_2]$ (UXISAX), (гидроксо)-(нафталин-2,6-дикарбоксилато)-ванадий(III) $\text{V}(\text{OH})(\text{C}_{12}\text{H}_6\text{O}_4)$ (TEWVUP). Координация атомов металла в этих соединениях описывается координационными формулами $\text{AB}^{01}_2\text{M}^1_3$, AB^{11}_2 , AT^{11}_2 и AK^4M^2 , соответственно. Атомы водорода не показаны для наглядности. Референтные коды структур в базе данных CSD указаны в верхней части рисунка.

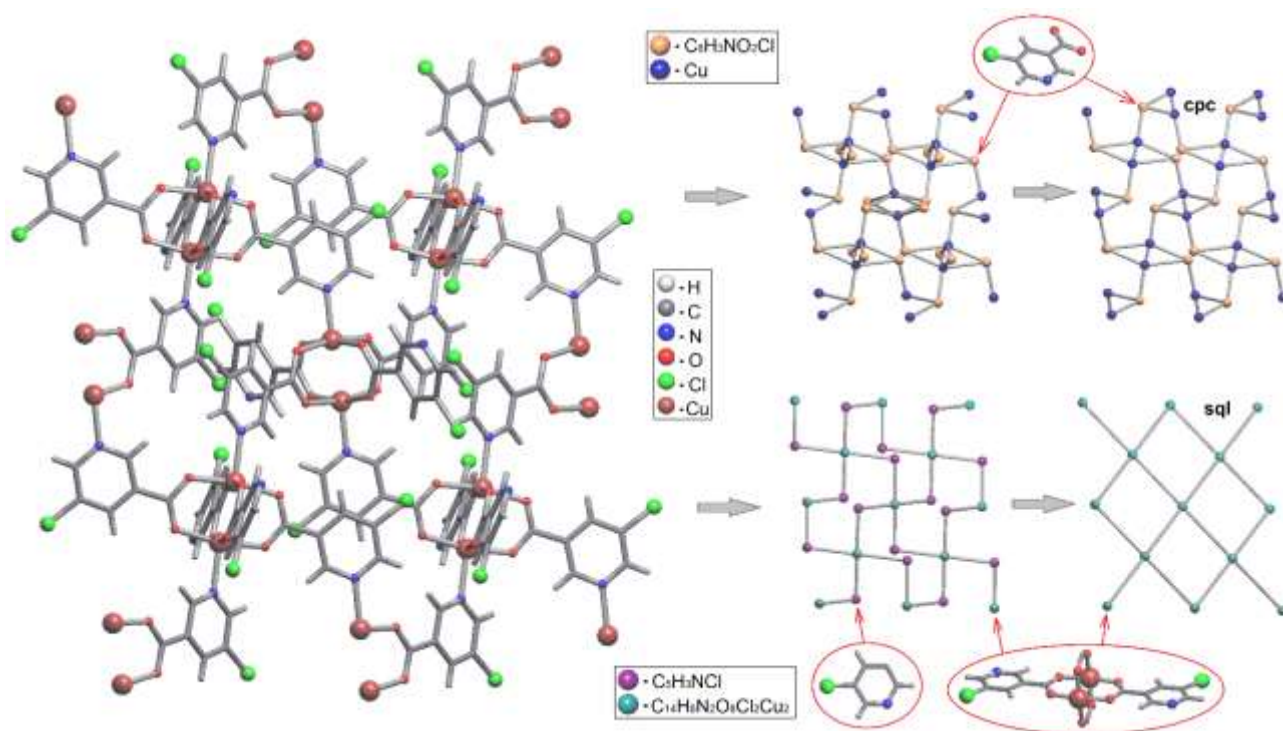


Рис. 3. Представления кристаллической структуры бис(5-хлороникотинато)-медь(II) $[\text{Cu}(\text{C}_5\text{H}_3\text{NO}_2\text{Cl})_2]$ (слева) (АСАКАТ) и соответствующих базовых сеток: стандартное представление (вверху справа), кластерное представление (внизу справа) с топологическими типами **cpc** и **sql**, соответственно. Базовые сетки до удаления 2-координированных узлов показаны в центре рисунка. В легенде окраски узлов указаны формулы ССБ, используемых при упрощении.

Базовую сетку можно отнести к определённому *топологическому типу*, определяемому набором топологических индексов (инвариантов). В главе рассмотрены известные инварианты сеток и топологические библиотеки, используемые для классификации топологии кристалла по топологическим типам. Для обозначения топологии сеток обычно используется номенклатура RCSR³ (Reticular Chemistry Structure Resource) с трехбуквенными символами. Если сетка отсутствует в RCSR, то она обозначается по номенклатуре ToposPro в виде символа ND_n, где N — последовательность координационных чисел всех неэквивалентных узлов сетки, D — периодичность сетки (D=M, C, L, T для 0-, 1-, 2-, 3-периодических сеток), а n — порядковый номер сетки в множестве всех топологически разных сеток с заданной последовательностью ND.

Далее описываются известные геометрико-топологические характеристики (*дескрипторы*) атомов, ССБ и кристаллической структуры в целом (Табл. 1), а также методы их расчета в рамках вышеупомянутых моделей на примере конкретных кристаллических структур. Локальная топология сетки характеризуется координационной фигурой АВ_{CN}, которую образуют узлы В_i, смежные с данным узлом А; если сетка представлена

³ <http://rcsr.anu.edu.au/>

атомами, то количество атомов B_i равно КЧ атома А. Тип координации лиганда, согласно номенклатуре⁴, задается в виде символа $L^{mbtkpghond}$, в котором лиганд L обозначается одной из латинских букв M, B, T, K, P, G, H, O, N или D в зависимости от количества донорных атомов, связанных с металлом (1, 2, 3, 4, 5, 6, 7, 8, 9 или 10, соответственно). Общее число атомов металла, координируемых лигандом (v), задается строкой mbtkpghond в виде верхнего индекса, где каждое целое число m, b, t, k, p, g, h, o, n, d равно количеству атомов металла, координируемых одним, двумя, тремя и так далее донорными атомами; $v = m + b + t + k + p + g + h + o + n + d$. Локальная топология комплексной группы записывается в виде координационной (или кристаллохимической) формулы $A_n L_{1i_1} L_{2i_2} L_{3i_3} \dots$, где А обозначает атом металла, n - число атомов металла в полиядерном звене, L1, L2, L3... обозначают типы координации лигандов, а $i_1, i_2, i_3 \dots$ количество таких лигандов.

Топологические методы используются не только для описания атомной структуры, но и для характеристики свободного пространства, не занятого атомами. Наиболее известной является модель натурального тайлинга⁵, элементы которого, натуральные тайлы, представляют собой обобщенные многогранники, образованные узлами и ребрами атомной сетки и заполняющие кристаллическое пространство без пропусков и пересечений (рис. 4). Натуральные тайлы характеризуют размер и форму минимальных клеток (полостей) каркаса, а тайлинг в целом кодирует способ соединения клеток через общие грани тайлов, соответствующие каналам между полостями⁶.

⁴ Serezhkin V.N., Vologzhanina A.V., Serezhkina L.B. et al. Acta Cryst. 2009, B65, 45–53.

⁵ Blatova O.A., Golov A.A., Blatov V.A. Z. Kristallogr. Cryst. Mater., 2019, 234 (7-8), 421–436.

⁶ Delgado-Friedrichs O., O’Keeffe M. Acta Cryst. 2005, A61, 358-362.

Табл. 1. Список использованных дескрипторов межатомного контакта, атома, многоатомного ССБ и кристаллической структуры

Обозначение [#]	Дескриптор
<i>Межатомные контакты</i>	
Rank	наименьшее число валентных связей, соединяющих атомы, которыми образована грань ПВ
<i>Атом</i>	
AtNum	порядковый номер металла в Периодической системе
Group	номер группы металла в Периодической системе
Mult	кратность кристаллографической позиции
CN	координационное число
NZ	взвешенное координационное число
V _{CP}	объем координационного полиэдра (Å ³)
G ₃	второй момент инерции полиэдра Вороного
Δ _{CM}	смещение атома металла из геометрического центра тяжести его полиэдра Вороного (Å)
R _{sd}	радиус сферического домена, объем которого равен объему полиэдра Вороного (Å)
V _{VP}	объем полиэдра Вороного (Å ³)
S _{VP}	площадь поверхности полиэдра Вороного (Å ²)
ΣΩ(X)	доля телесных углов граней, которые образованы соседними атомами X определенного химического сорта в полиэдре Вороного этого атома
OxSt	степень окисления
T _{VP}	топология реберной сетки упрощенного полиэдра Вороного
<i>Многоатомный ССБ</i>	
TF	химический состав в виде валовой формулы (Total Formula)
CT	тип координации лиганда (Coordination Type)
ξ	ионный заряд
V _{VP} ^M	молекулярный объем
S _{VP} ^M	площадь внешней поверхности
CN _{UN}	координационное число в базовой сетке
CF _{UN}	координационная фигура в базовой сетке
<i>Кристаллическая структура</i>	
CCF	кристаллохимическая формула соединения
χ	соотношение первичных ССБ в соединении
T _{UN} ^{stand}	топология базовой сетки в стандартном представлении
T _{UN} ^{cluster}	топология базовой сетки в кластерном представлении при заданной величине максимального размера колец (η) в ССБ
Z	количество взаимопроникающих структурных сеток
P	доля свободного пространства, рассчитанная по объему, занятому Ван-дер-Ваальсовыми полиэдрами атомов
D _{chan}	размерность системы каналов при заданном радиусе пробной частицы
Σξ _i	дисбаланс положительных и отрицательных зарядов
TAM	топологический механизм сборки из ССБ (Topological Assembly Mechanism)

[#]Жирным шрифтом выделены дескрипторы, которые впервые введены в этой работе.

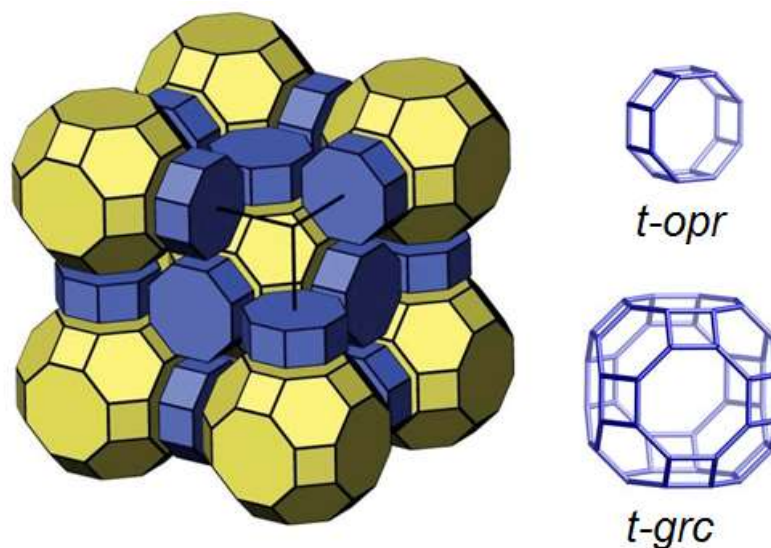


Рис. 4. Натуральный тайлинг в цеолите RHO (слева), который состоит из двух типов натуральных тайлов с именами *t-opr* и *t-grc* (справа).

Далее рассмотрены методы поиска закономерностей «состав-структура-свойство», в том числе с использованием интеллектуальных методов анализа данных (методов машинного обучения), которые только начинают использоваться в кристаллохимии. В качестве примера выбраны две хорошо известные в кристаллохимии проблемы: определение формы координационного полиэдра атомов в кристаллической структуре⁷ и расчет степени окисления атомов металлов в структуре координационных соединений⁸. Для каждого варианта решения этих задач рассмотрены преимущества и недостатки. Так, авторы структурных исследований, как правило, определяют форму координационной фигуры визуальным способом, что вносит субъективность и неоднозначность в качество оценки. При автоматизированном определении степени окисления для всех видов атомов металлов, как в неорганических, так и в металлоорганических кристаллических структурах до сих пор не разработана универсальная схема расчета.

Вторая глава содержит описание разработанного автором универсального геометрико-топологического подхода к моделированию и анализу кристаллической структуры любого химического состава и строения.

В основе этого подхода лежит представление всех атомов структуры в виде их полиэдров Вороного или Ван-дер-Ваальсовых полиэдров. Основная роль полиэдра Вороного заключается в переходе от геометрического («физического») описания структуры кристалла, непосредственно получаемого из эксперимента, в котором рассматриваются только атомы без

⁷ Klein H.-J., Mennerich C. Int. J. Adv. Sys. Measur. 2009, 2–3, 192–203.

⁸ Jablonka K.M., Ongari D., Moosavi S.M., Smit B. Nat. Chem. 2021, 13, 771–777.

учета взаимодействия между ними, к топологическому («химическому») описанию, в котором восстанавливается связность кристаллического пространства с учетом межатомных взаимодействий. При этом постулируется, что межатомное взаимодействие имеет место только в случае наличия общей грани полиэдров Вороного у данной пары атомов, то есть при непосредственном контакте атомных доменов. В том случае, если линия, соединяющая атомы, не пересекает общую грань их полиэдров Вороного, контакт называется *непрямым* и считается невалентным, так как обменные взаимодействия между атомами в этом случае затруднены. Размер грани, оцениваемый ее площадью или телесным углом, с учетом радиусов взаимодействующих атомов и размера других граней полиэдра Вороного атома, позволяет с высокой точностью определять тип межатомного контакта (валентный или невалентный), а использование дополнительных геометрических параметров – характеризовать невалентные контакты как водородные, специфические (галоген-галоген) или Ван-дер-Ваальсовы.

Данный подход алгоритмизирован и реализован автором в виде процедур программного пакета *TorosPro*. Для сохранения информации о связности (топологии) структуры кристалла разработан уникальный формат, в котором учитывается пространственная симметрия кристалла, включая трансляционную составляющую. В настоящее время *TorosPro* является единственным программным пакетом, способным описывать топологию всего кристалла, а не только его отдельных молекулярных компонент.

Понятие Ван-дер-Ваальсова полиэдра (рис. 1 справа) введено автором впервые как обобщение понятия полиэдра Вороного. При построении Ван-дер-Ваальсова полиэдра половинное деление межатомных контактов сохраняется лишь в том случае, если межатомное расстояние меньше суммы Ван-дер-Ваальсовых радиусов взаимодействующих атомов. Когда межатомное расстояние превышает эту величину, контакт делится плоскостью на удалении, равном Ван-дер-Ваальсовому радиусу центрального атома полиэдра. В плотноупакованных ионных или интерметаллических соединениях межатомные расстояния, как правило, не превосходят суммы Ван-дер-Ваальсовых радиусов атомов, и предложенный алгоритм сводится к построению разбиения Вороного. Напротив, при наличии в кристаллах свободного пространства, как в случае МОКП или молекулярных кристаллов, поверхности атомов будут усечены Ван-дер-Ваальсовыми радиусами на границе с пустотой, что дает возможность количественно с высокой точностью оценить, как объем пустоты, так и характеристики ее внутренней поверхности.

Координационные соединения содержат в качестве ССБ атомы металлов-комплексобразователей, лиганды, сольваты и гостевые молекулы;

каждый тип ССБ имеет определенный набор дескрипторов. В TorosPro нами впервые реализована комплексная оценка структурных свойств ССБ с помощью полиэдров Вороного и Ван-дер-Ваальса. В дополнение к классическим кристаллохимическим дескрипторам (межатомные расстояния, валентные углы, плотность кристаллической структуры), нами были предложены объемы и площади поверхности атомных и молекулярных полиэдров Вороного и Ван-дер-Ваальса, а также дескрипторы координационных фигур в базовой сетке. Их совокупность, в отличие от классических дескрипторов, характеризует ССБ и связи между ними как трехмерные объекты.

Особое внимание в геометрико-топологическом подходе уделено координационной фигуре ССБ в упрощенной базовой сетке, так как топологический тип сетки зависит от формы и числа вершин, образующих ее координационных фигур, а также от их соотношения в конкретной структуре. Так как координационные фигуры в базовой сетке отличаются большим разнообразием, впервые предложена специальная система обозначения координационных фигур в виде NAME- $n\{\alpha_1, \alpha_2, \dots, \alpha_m\}$, где NAME - имя класса фигур, n - число вершин у фигуры (координационное число атома или лиганда), а $\{\alpha_i\}$ - набор всех возможных углов (в градусах) между прямыми, проходящими через центр фигуры и пару ее вершин, упорядоченных по возрастанию и кратных 15° ; $m = n(n-1)/2$. Поскольку углы координационной фигуры в конкретной кристаллической структуре могут отличаться от эталонных значений α_i , мы относим ее к определенному типу, если выполняется условие $|\alpha_i - \alpha_i| < 7.5^\circ$. Например, координационные фигуры лиганда и атома металла в кристаллической структуре бис(этилендиолато)-титан(IV) $[\text{Ti}(\text{C}_2\text{H}_4\text{O}_2)_2]$ относятся к уголкового A-2{75} и тетраэдрической T-4{90,105,135}, соответственно, хотя их углы ({69.13} и {88.84, 104.66, 104.66, 110.87, 110.87, 129.50}) отличаются от значений эталона (Рис. 5).

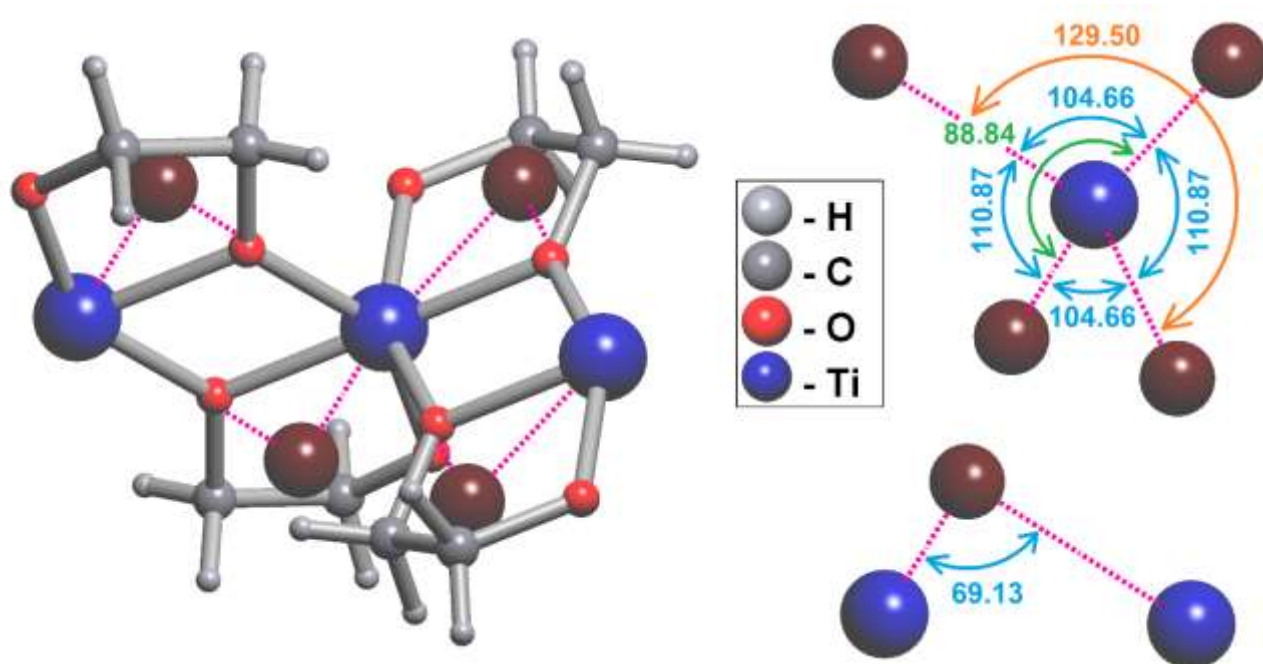


Рис. 5. Фрагмент структурной и базовой сеток (слева), а также координационные фигуры металла T-4{90,105⁴,135} и лиганда A-2{75} (справа) с их плоскими углами (в град.) для цепочечной кристаллической структуры бис(этилендиолата)-титан(IV) [Ti(C₂H₄O₂)₂] (KPFQEQ). Коричневые шары соответствуют центру тяжести лигандов.

Геометрико-топологический подход является универсальным, позволяя анализировать не только атомную структуру, но и свободное пространство кристалла с использованием тех же моделей и программных средств. В связи с этим нами введен термин «дуальная кристаллохимия», то есть кристаллохимия не материальных объектов (атомов и связей между ними), а пустот и каналов между пустотами.

Для пористых структур важно не только наличие свободного пространства, но и его доступность для миграции молекул-адсорбентов через каналы. Характеристики каналов кристаллической структуры можно определить, исходя из сетки ребер атомных полиэдров Вороного (*сетки Вороного*). Анализ сетки Вороного позволяет рассчитать периодичность системы каналов, максимальный радиус сферы, размещаемой внутри пустоты (R_i), а также радиус самого большого сферического зонда (R_f), который может свободно мигрировать в этой системе каналов. Например, анализ структуры [Zn(QDA)]·0.5H₂O·0.7DMF (FIZGUT), где DMF – диметилформамид, H₂QDA – хиолин-2,6-дикарбоновая кислота, показал наличие одномерных каналов в направлении [0,0,1], характеризующихся $R_i = 2.38 \text{ \AA}$ и $R_f = 1.52 \text{ \AA}$ (Рис. 6).

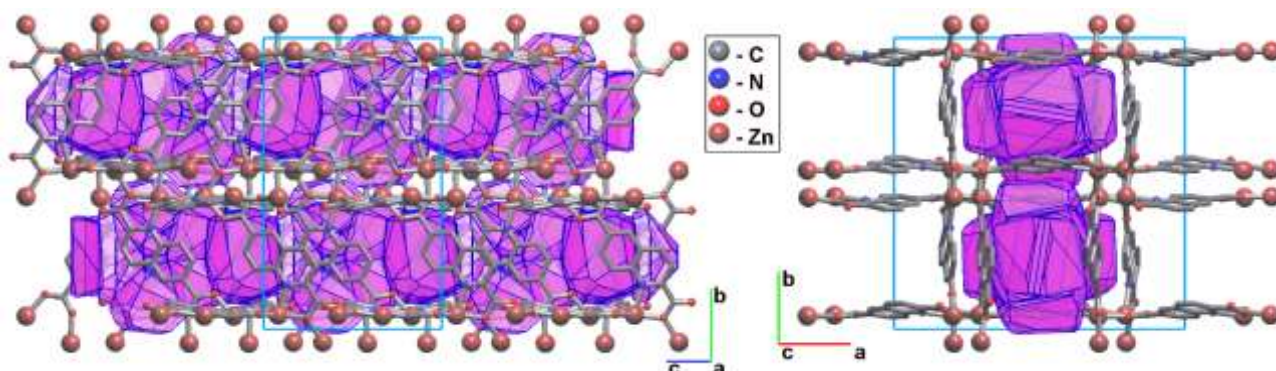


Рис. 6. Полиэдрическое представление каналов в структуре $\text{Zn(QDA)} \cdot 0.5\text{H}_2\text{O} \cdot \text{DMF}$ (FIZGUT) в проекциях 0YZ (слева) и XY0 (справа). Атомы водорода не показаны на рисунке для наглядности. Границы элементарной ячейки изображены голубым цветом.

Значения предложенных геометрико-топологических дескрипторов были рассчитаны для 67 087 неорганических, 557 900 координационных и 244 832 органических кристаллических структур из баз данных CSD и ICSD, а также 604 273 гипотетических цеолита и 502 аллотропа углерода. В результате была создана система баз знаний (электронных библиотек, Табл. 2). Обработка данных и анализ этих библиотек осуществляется с помощью разработанной нами программы Samara Assistant of Crystal Design (SACryDes), которая позволяет формировать сложные запросы, как к отдельным, так и к нескольким библиотекам. Удаленный доступ к этим данным осуществляется через разработанные нами веб-интерфейсы, размещенные на сайте topcryst.com.

Табл. 2. Топологические электронные библиотеки

Название библиотеки	Сокращение	Содержимое библиотеки	Количество записей [#]
Topological Types Database	TTD	топологические типы сеток	808 445
Topological Types Observed	TTO	представители топологических типов кристаллических структур	2 675 446
Topological Types of metal Atoms	TTA	геометрико-топологические дескрипторы атомов металла	1 928 816
Topological Types of Molecules	TTM	геометрико-топологические дескрипторы органических молекул	309 353
Topological Types of Ligands	TTL	геометрико-топологические дескрипторы лигандов	269 232
Topological Types of Nanoclusters	TTN	топологические типы нанокластеров	1 234
Topological Types of Tiles	TTT	топологические типы тайлов	415
Topological Types of Polyhedra	TTP	топологические типы полиэдров	4 241

[#]Количество записей для всех библиотек указано по состоянию на первый квартал 2024 г.

Указанные программные инструменты, совместно с программным комплексом ToposPro, были использованы при подготовке выборок для поиска закономерностей и правил, а сам поиск осуществлялся средствами программной среды [Python](#). Для стандартизации анализа кристаллографических данных машинными методами автором предложена Схема I, которая содержит 9 этапов:

1. Указываем соответствующую задачу, *целевой дескриптор*, прогноз которого будет выполнять модель, и выбираем класс кристаллических структур для ее решения. Класс должен содержать достаточное количество данных, чтобы искать устойчивые корреляции, а задача должна быть строго и однозначно сформулирована.
2. Формулируем условия отбора данных, необходимых и достаточных для решения задачи. Их правильный выбор позволит сократить количество влияющих факторов и, тем самым, уменьшить количество обрабатываемых дескрипторов, что упрощает задачу и улучшает качество ее решения.
3. Выбираем структурные данные из первоисточников по заданным условиям.
4. Подбираем набор дескрипторов, подходящих для решения задачи. Реализуем алгоритмы их вычисления в программном коде, если это необходимо.
5. Готовим данные для анализа. На этом этапе мы исключаем исходные данные с ошибками и низким качеством, помечаем дублирующиеся записи и выбираем представителя для каждого набора идентичных записей.
6. Выполняем статистический анализ значений дескрипторов, чтобы проверить качество данных и исключить записи, отвечающие статистическим выбросам.
7. Используя численные критерии оценки качества результата, выбираем метод машинного обучения, который лучше всего подходит к данной задаче.
8. Выполняем машинное обучение для прогнозирования целевого дескриптора и выбираем значимые дескрипторы из обучающего набора.
9. Прогнозируем целевой дескриптор для кристаллических структур, которые не содержались в обучающей выборке и, сравнивая результаты прогноза с ожидаемыми значениями дескриптора, проверяем качество созданной нами модели.

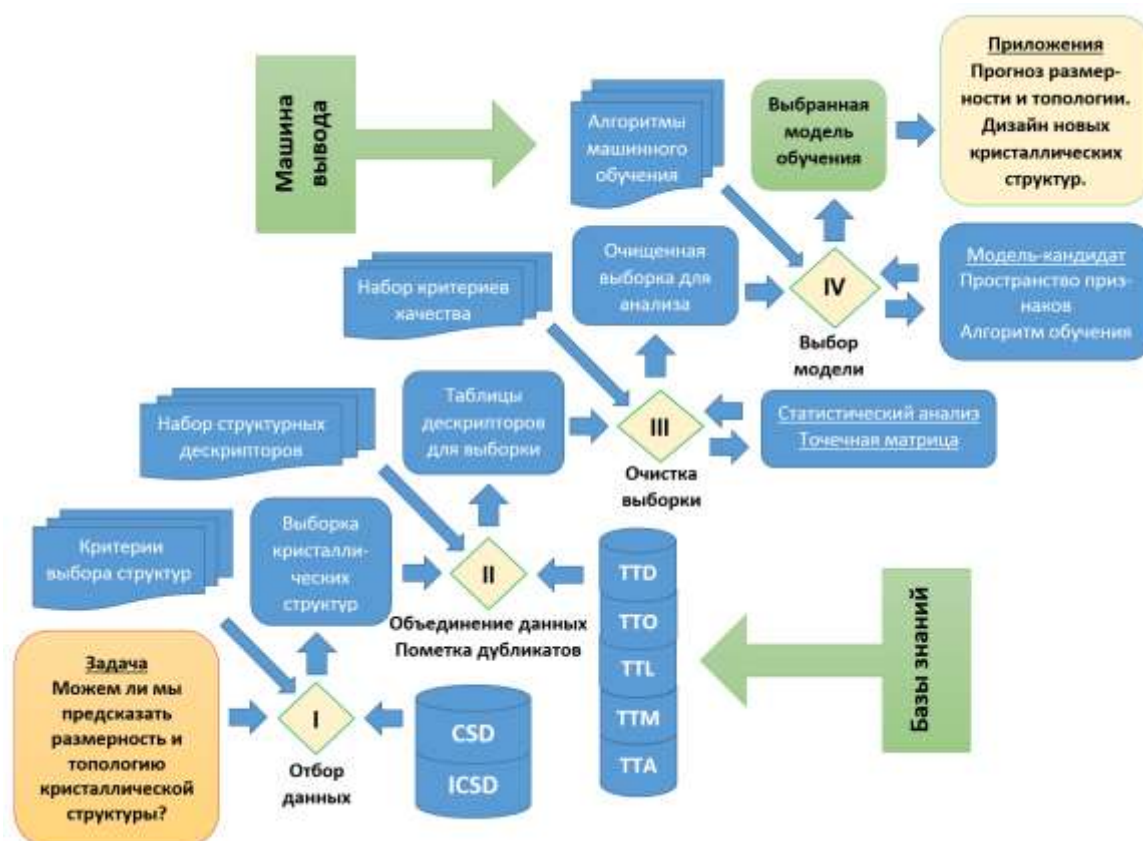


Схема I. Общая схема подготовки и анализа кристаллографических данных, а также итоговые компоненты системы ИИ.

Для проверки работоспособности и универсальности описанной схемы, нами были сформулированы и решены несколько типичных физико-химических задач (Табл. 3), которые детально рассматриваются в следующих главах.

В **Третьей главе** рассмотрено применение предложенного автором подхода для решения задач 1–7 (Табл. 3). При решении первой задачи были изучены структуры координационных соединений Cu и Zn, атомы которых имеют КЧ=5 и σ -связаны с атомами кислорода или азота, однако могут иметь две основные геометрические формы координационной фигуры – тригональную бипирамиду и квадратную пирамиду, а также формы, промежуточные между указанными и сложные для визуальной оценки. Нами предложены три метода классификации формы координационной фигуры, которые взаимно дополняют друг друга.

Табл. 3. Физико-химические задачи и выборки кристаллоструктурных данных, использованные для их решения

Физико-химическая задача	Объем выборки	Раздел
1. Определение формы координационной фигуры	9 642	3.1
2. Прогноз размерности и топологического типа в КС и МОКП	7 690	3.2
3. Анализ свободного пространства в структуре МОКП	33 790	3.3
4. Прогноз степени окисления атомов металла в КС и МОКП	33 253	3.4
5. Анализ размеров атомов в структуре простых веществ	1 249	4.1, 4.2
6. Анализ каналов проводимости в Na-содержащих структурах	402	4.3
7. Классификация кристаллических структур интерметаллидов	38 000	4.4
8. Прогноз плотности кристаллов насыщенных углеводородов	147	4.5

В *первом методе* для классификации используется реберная сетка упрощенного полиэдра Вороного центрального атома координационной фигуры (рис. 7а), так как геометрически правильная тригональная бипирамида и квадратная пирамида имеют четко различающиеся формы полиэдров Вороного: тригональную призму и квадратную пирамиду, соответственно.

Второй метод основан на итеративном наложении координационного и эталонного многогранников. При этом положение эталонного многогранника не изменяется, а все операции выполняются с координационным многогранником, который сначала масштабируется до объема эталонного многогранника, а затем совмещается с ним путем сдвига и поворота (Рис. 7б).

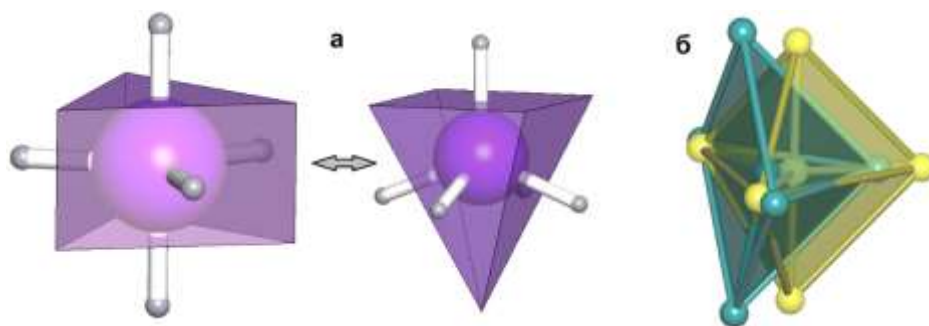


Рис. 7. а - Координационные фигуры и полиэдры Вороного для тригональной бипирамиды (слева, топологический тип $\{5/9/6-1\}$) и квадратной пирамиды (справа, топологический тип $\{5/8/5-1\}$); б - результат совмещения треугольной бипирамиды и квадратной пирамиды с максимальным объемом их пересечения 75.4%. Для обозначения топологического типа использована номенклатура $\{Г/Р/В-Н\}$, где буквами обозначено количество Граней, Ребер, Вершин в упрощенном полиэдре Вороного, а также уникальный Номер топологии при совпадении первых трех значений.

В *третьем методе* для определения формы координационной фигуры используется алгоритм сравнения *угловых отпечатков* (Рис. 8), то есть векторов, описывающих распределение углов в координационной фигуре и соответствующих одному из ее обозначений вида $T-4s\{105^4, 120^2\}$ (см. Главу 2).

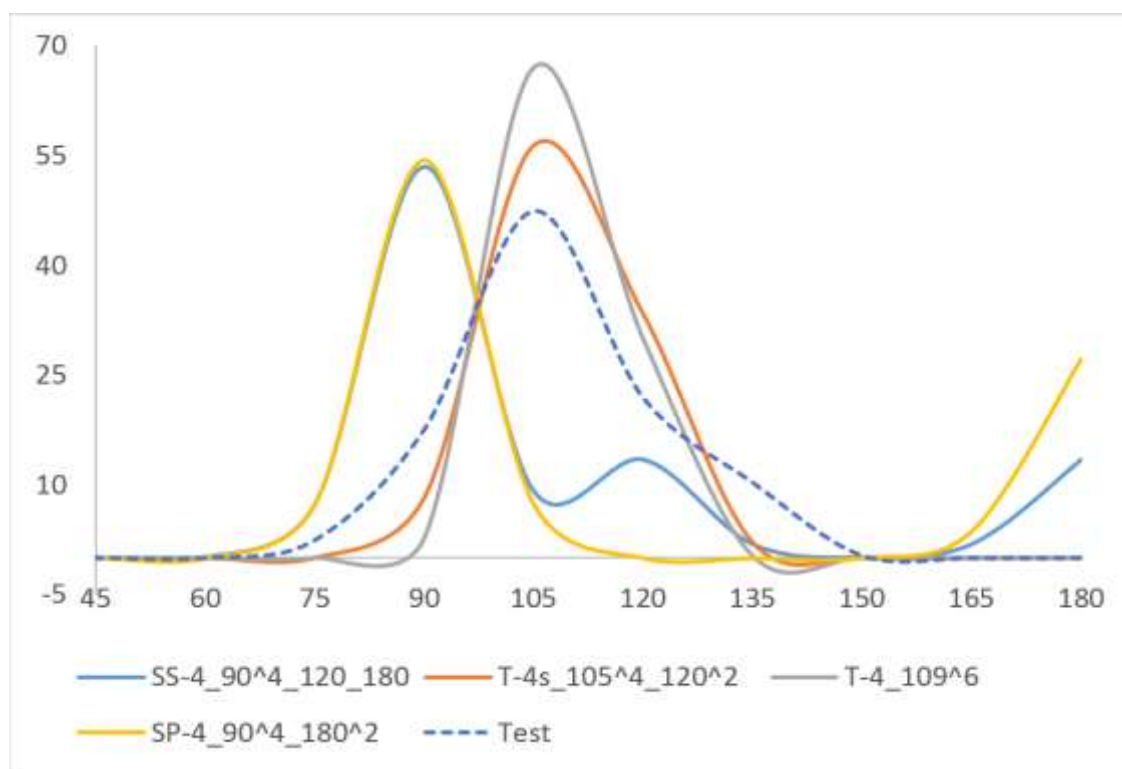


Рис. 8. Модельные отпечатки правильного $T-4\{109^6\}$ (серая линия) и искаженного $T-4s\{105^4, 120^2\}$ (оранжевая линия) тетраэдра, квадрата $SP-4\{90^4, 180^2\}$ (желтая линия), бисфеноида $SS-4\{90^4, 120, 180\}$ (синяя линия), а также отпечаток искаженного тетраэдра $T-4\{90, 105^4, 135\}$ в базовой сетке структуры бис(этилендиолата)-титан(IV) $[Ti(C_2H_4O_2)_2]$ (КЕРФЕК) (синяя штриховая линия), для которого наименьшее расстояние 9.6% соответствует модельному отпечатку $T-4s\{105^4, 120^2\}$.

Расчета углового отпечатка состоит из следующих этапов:

- диапазон возможных значений углов от 0° до 180° делится на равные интервалы (например, если число интервалов равно 10, то длина интервала $\delta=18^\circ$);
- для каждого интервала рассчитывается количество попадающих в него значений углов;
- найденное для каждого интервала число попаданий аппроксимируется гауссовой функцией (1) со стандартным отклонением σ и средним значением μ , равным середине интервала (например, $\sigma=9^\circ$ для $\delta=18^\circ$).

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (1)$$

С помощью функции (2) рассчитываются компоненты F_i отпечатка-вектора, имеющего размерность, равную числу интервалов.

$$F_i = 100 \frac{\sum_{k=1}^{180/\delta} f_k(\mu_i)}{\sum_{m=1}^{180/\delta} \sum_{k=1}^{180/\delta} f_k(\mu_m)} \quad (2)$$

Таким образом, компонент F_i отвечает сумме вкладов всех интервальных гауссовых функций в i -ый интервал и выражен в процентах к полной их сумме. Чтобы определить степень сходства двух координационных фигур достаточно найти расстояние между их отпечатка-векторами F и F' по формуле (3).

$$r = \sqrt{\sum_{i=1}^{180/\delta} \left(\frac{F_i - F'_i}{2} \right)^2} \quad (3)$$

Каждый из предложенных трех методов имеет свои преимущества и недостатки. Так первый метод очень быстрый, но не позволяет оценить степень искажения координационной фигуры. Второй метод самый затратный, не учитывает положение центрального атома, но позволяет сравнивать координационные фигуры произвольной формы и с любым числом вершин. Метод отпечатков является таким же быстрым, как метод Вороного, хотя и применим только для координационных фигур с одинаковым числом вершин. Чтобы компенсировать недостатки каждого из алгоритмов, мы использовали комбинацию этих трех методов, которая позволила выявить следующие правила координации в рассматриваемых классах соединений (рис. 9):

- Если атомы азота доминируют в координационном полиэдре N_nO_m ($n>m$), то тригонально-бипирамидальная координация более вероятна (62/38), в противном случае с той же вероятностью 0.62 преобладает квадратно-пирамидальная координация.

- Общая топология сетки неоднозначно зависит от формы координационной фигуры: некоторые топологии, как 3,5L2, **fet**, **fes**, предпочтительны для тригонально-бипирамидного типа, топологии **xah**, 4,4L1 допустимы только для квадратно-пирамидального типа, в то время как топологии, расположенные близко к линии равных вероятностей (**sql**, **hcb**, **srs**), нечувствительны к типу координационной фигуры.

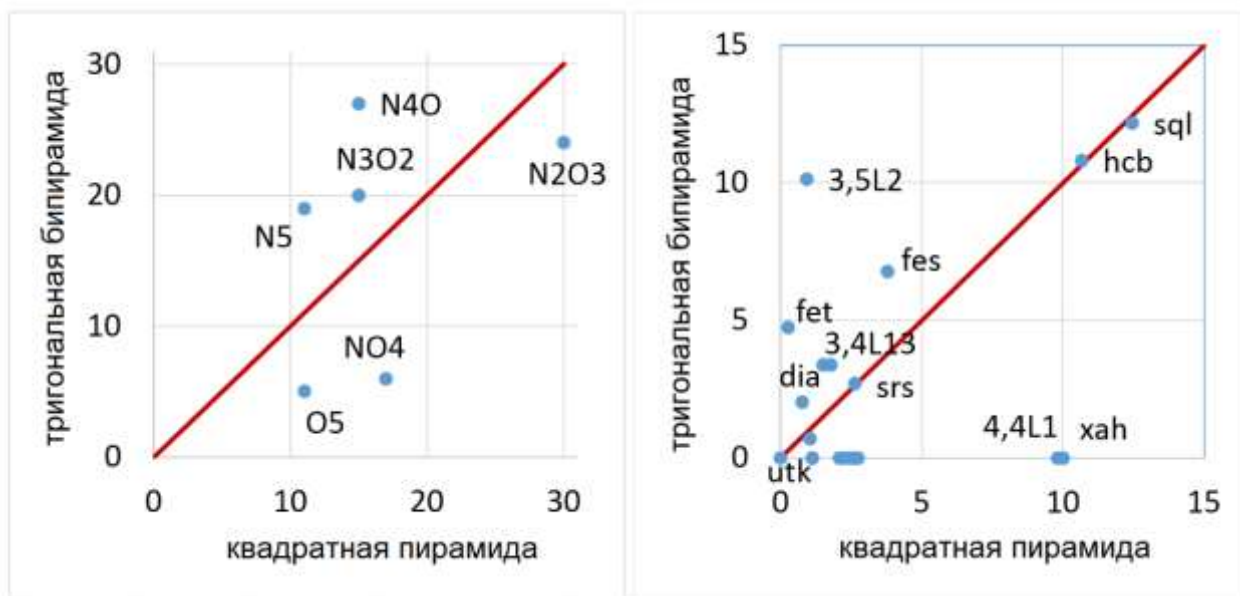


Рис. 9. Диаграммы рассеяния «квадратная пирамида – тригональная бипирамида» для встречаемости координационных фигур (в %) по химическому составу координационной фигуры (слева) и по топологии базовой сетки (справа).

Найденные правила, представленные в виде дерева решений (Рис. 10), позволяют прогнозировать вероятное строение кристаллов на основании знаний о химической природе лигандов. Например, если мы хотим получить МОКП с топологией **xah**, то нужны лиганды с преобладанием координирования атомами кислорода, которые склонны к образованию структурного блока «китайский фонарик», допускающие локальную координацию $A_2K^4_2B^2$. Напротив, для получения МОКП с топологией **fet** следует выбирать лиганды с преобладанием координации по атому азота, для которых не характерно образование структурного блока «китайский фонарик», и которые допускают локальную координацию AT^3B^2 (Рис. 11). Вероятность получить топологию **fet** при соблюдении указанных выше условиях составит 23% (Рис. 10).

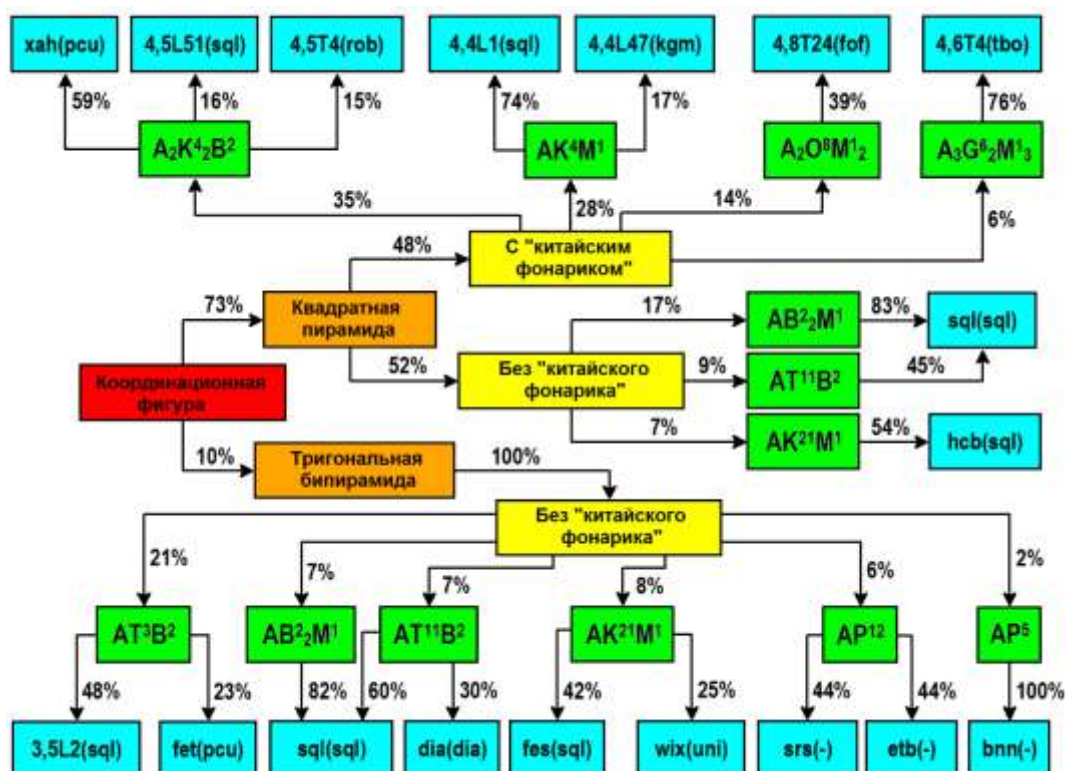


Рис. 10. Множественные корреляции «Координационная фигура (оранжевый) - вторичная строительная единица (желтый) - локальная координация (зеленый) - общая топология (голубой)» для 2D и 3D кристаллических структур. Топологический тип указан в стандартном и кластерном представлениях (в скобках).

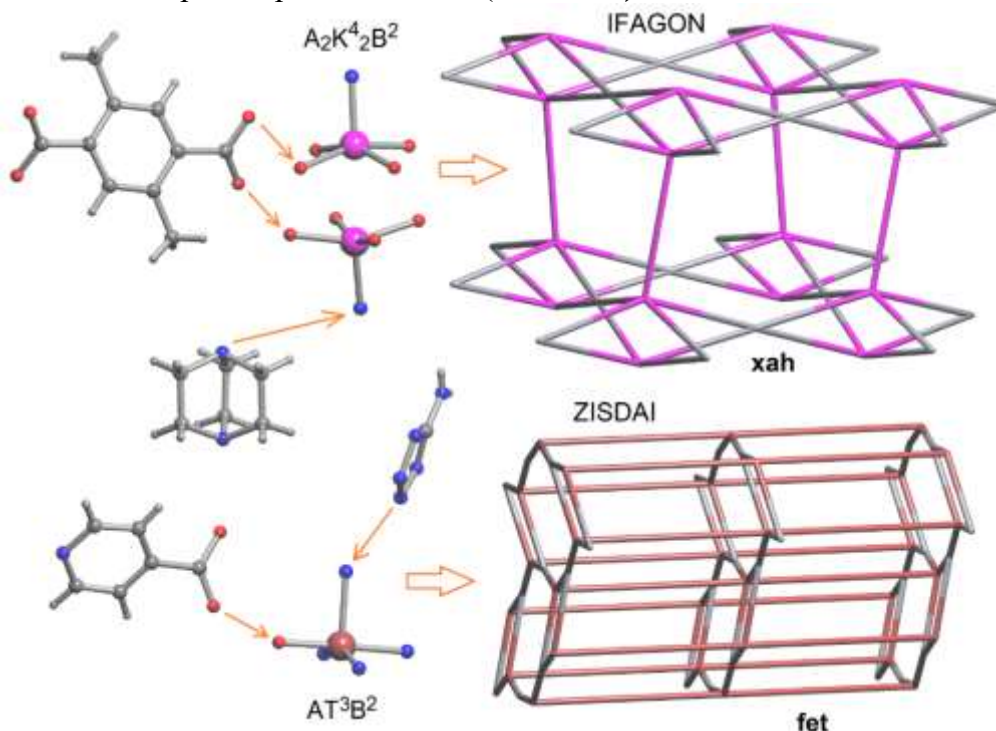


Рис. 11. Влияние природы донорных атомов лигандов на выбор топологического типа базовых сеток **fet** (внизу) или **xah** (вверху), реализованных в кристаллических структурах (μ_3 -5-аминотетразолил) - (μ_2 -изоникотинато) - меди(II) (IFAGON) и бис(μ_4 -2,5-диметилтерефталато)-(μ_2 -1,4-дизаобицикло[2.2.2]октан)-дицинка DMF сольвата (ZISDAI), соответственно.

При решении *второй задачи* для поиска корреляций между структурными дескрипторами, размерностью и топологическим типом базовой сетки в соответствии со Схемой I было протестировано несколько алгоритмов машинного обучения, которые реализованы в библиотеке Scikit-learn⁹ программной среды Python. Оптимальной для решения данной задачи оказалась модель случайного леса. Так, кривые обучения модели для прогноза размерности кристаллической структуры (Рис. 12) свидетельствуют о том, что качество модели улучшается при росте объема тренировочной выборки, так как разрыв между кривыми сокращается, а доля правильных решений для отложенной тестовой выборки (*точность модели*) по завершению обучения составляет 90%.

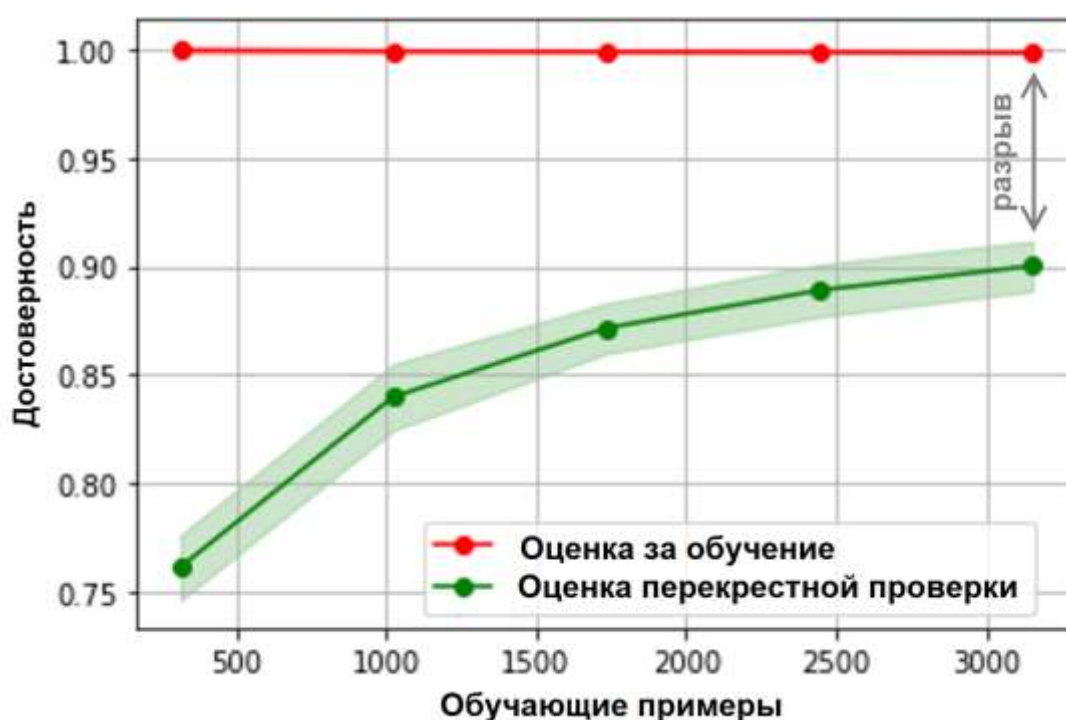


Рис. 12. Кривые обучения для тренировочной (красная линия) и тестовой (зеленая линия) выборок, полученные с использованием классификатора случайного леса для взвешенной схемы набора дескрипторов $CN_{UN}+CT+CF_{UN}$ при оценке размерности кристаллической структуры.

В результате выполненного исследования была создана база знаний и разработаны модели, позволяющие прогнозировать размерность и топологию КС и МОКП, включающие полный набор дескрипторов $CN_{UN}+CT+CF_{UN}$, и обученные с помощью классификатора случайный лес. Точность их прогноза для размерности кристаллической структуры и ее топологического типа составила 90% и 75%, соответственно. Разработанные в результате этого

⁹ Pedregosa F., Varoquaux G., Gramfort et al. J. Mach. Learn. Res. 2011, 12, 2825–2830.

исследования база знаний и машинные модели позволяют ответить на следующие вопросы структурной координационной химии.

1. *Какова координация металла и лиганда?* Важный вопрос при планировании синтеза координационного полимера касается локальной координации структурных групп. Например, правильные тетраэдр и октаэдр отвечают координационным фигурам в идеальных сетках **dia** (алмаз) и **pcu** (α -Po или NaCl), соответственно, в то время как сетка **pts** (PtS) построена из координационных фигур тетраэдрической и квадратной формы. Поэтому если необходимо получить координационный полимер с топологией базовой сетки **dia**, то было бы разумно выбрать металл, для которого характерна тетраэдрическая координация и мостиковые лиганды с координационными фигурами A-2 (изогнутые) или L-2 (линейные) (рис. 13 вверху слева).

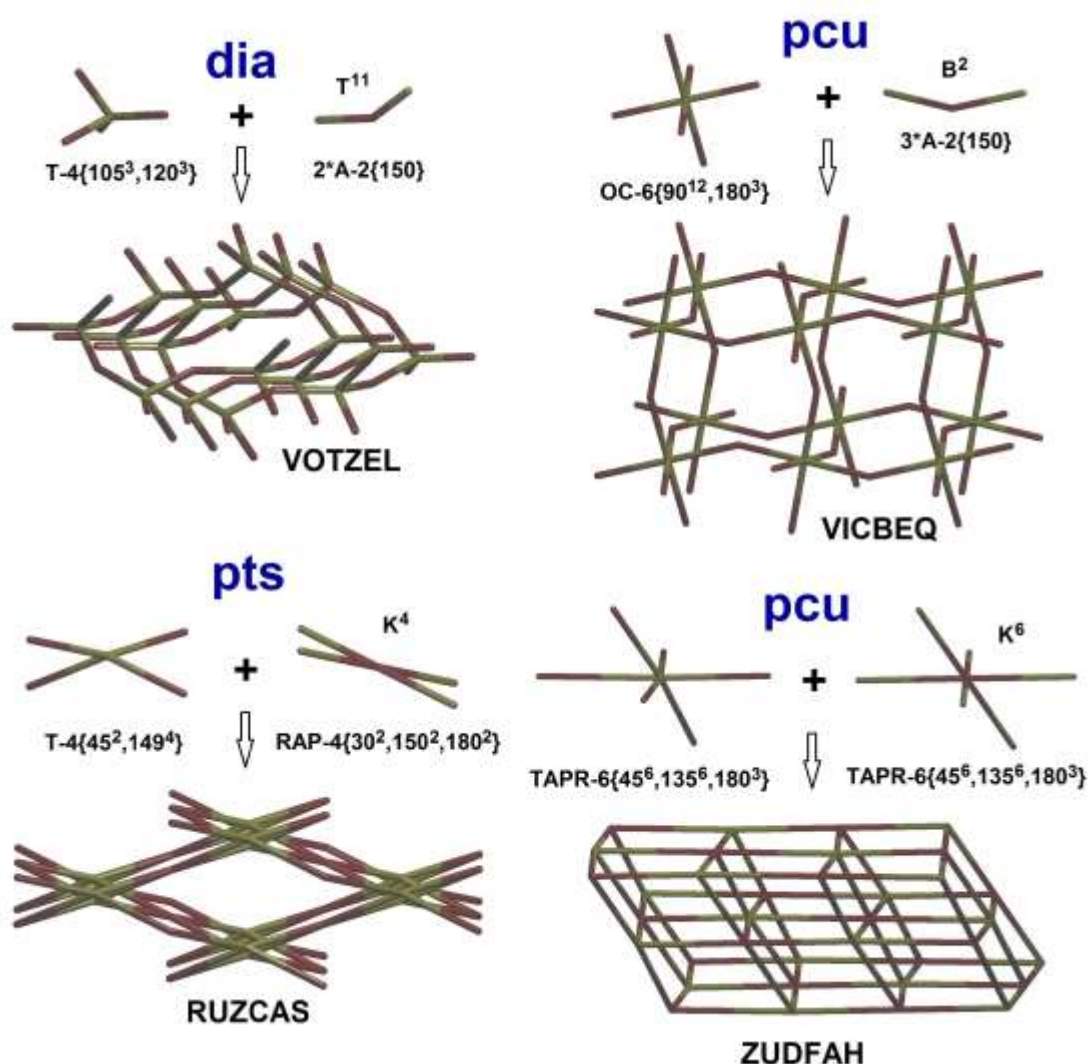


Рис. 13. Сборка координационных полимеров $[(\mu_2\text{-лапахолато})_2\text{Cd}]$ (VOTZEL), $[(\mu_2\text{-формiato})_3\text{Fe}]$ (VICBEQ), $[(\mu_6\text{-нафтагене-2,6- дикарбоксилато})\text{Pb}]$ (ZUDFAH) и $[(\mu_4\text{-4,4'- диметоксибифенил-3,3'- дикарбоксилато})\text{Mn}]$ (RUZCAS) с **dia** (вверху слева), **pcu** (вверху справа и внизу справа) и **pts** (внизу слева) топологией базовой сетки, соответственно. Координационные типы лигандов, а также типы координационных фигур приведены для всех структурных групп.

Подобные лиганды могут быть использованы и для получения координационного полимера с топологией базовой сетки **psc**, но в этом случае металлические центры должны иметь октаэдрическую координацию (рис. 13 вверху справа). Лиганды также могут играть ту же топологическую роль в структуре, что и атомы металла (рис. 13 внизу справа). Чтобы получить полимер с топологией базовой сетки **pts**, нужно подобрать металл и лиганд, для которых характерны тетраэдрическая и квадратная или прямоугольная координационные фигуры (рис. 13 внизу слева).

2. *Сходно ли поведение различных ССБ при сборке кристаллических структур?* Важным шагом в дизайне кристаллов является поиск ССБ, которые могут обеспечить определенный топологический мотив. ССБ могут быть химически совершенно разными, но при этом обладают схожими характеристиками с точки зрения формирования архитектуры МОКП и КС. Если в качестве фингерпринта ССБ взять набор вероятностей реализации конкретных значений заданного дескриптора D в различных структурах, то по половине суммы модулей разности компонент двух таких фингерпринтов Δ_D можно определить степень функциональной схожести ССБ. При этом величина Δ_D меняется в диапазоне от 0 до 1. Когда значение Δ_D близко к 0, ССБ могут считаться обладающими наиболее сходными свойствами по отношению к дескриптору D . Напротив, при приближении значений Δ_D к единице свойства ССБ можно считать абсолютно различными. Например, при поиске похожих лигандов можно сравнить значения Δ_D их координационных типов и/или координационных фигур. Согласно Рис. 14, пивалат подобен бензоату ($\Delta_{КТ}=0.131$, $\Delta_{КФ}=0.311$) и изобутирату ($\Delta_{КТ}=0.168$, $\Delta_{КФ}=0.377$), а трополан ($\Delta_{КТ}=0.096$, $\Delta_{КФ}=0.095$) и дибензоилметан ($\Delta_{КТ}=0.059$, $\Delta_{КФ}=0.110$) ведут себя в кристаллических структурах подобно ацетилацетонату. Используя информацию о сходстве, можно варьировать ССБ, чтобы приспособить их к требуемым дескрипторам КС и МОКП (цена, стабильность, токсичность и т. д.).

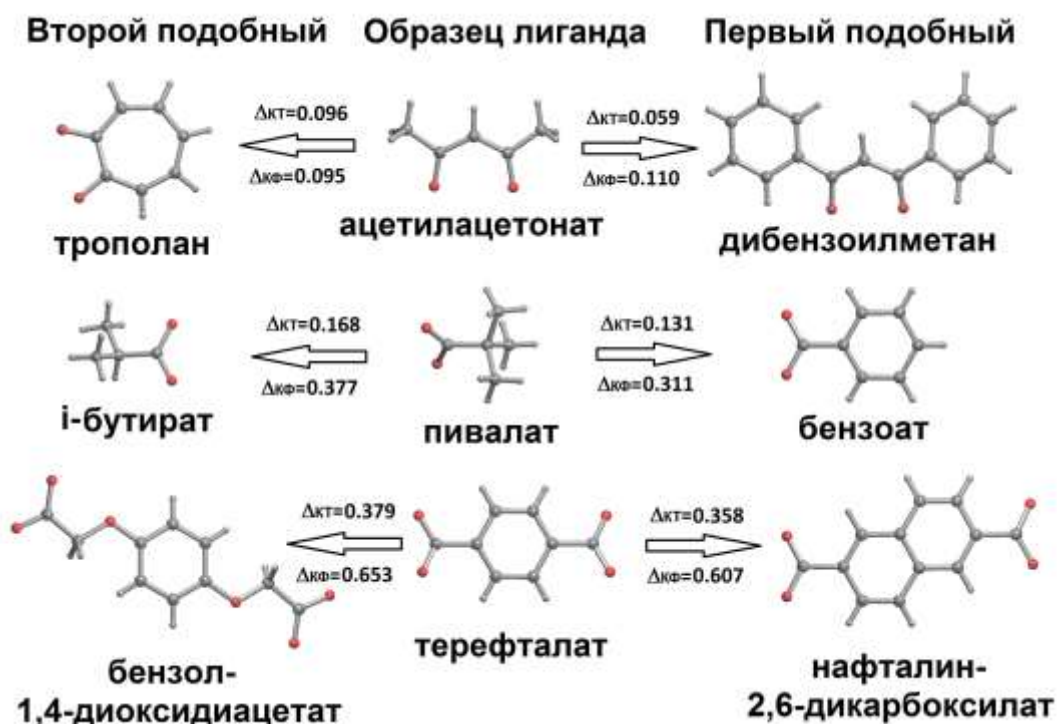


Рис. 14. Три группы лигандов наиболее схожих по их координационному типу (КТ) или координационным фигурам (КФ).

3. Склонен ли лиганд к образованию комплексов с данным металлом?

Одни и те же лиганды могут использоваться для дизайна разных топологий, но в таких случаях они должны взаимодействовать с разными металлическими центрами. База знаний может ответить, какие металлы и лиганды наиболее часто образуют соединения между собой.

Взаимосвязи между топологическими свойствами МОКП и геометрическими дескрипторами свободного пространства в их структуре рассмотрены в качестве *третьей задачи*. Для ее решения использованы топологические модели периодической сетки и натурального тайлинга, а также методы их анализа, которые реализованы в программном пакете ToposPro. Эти инструменты применены к 33 790 кристаллическим структурам трехмерных координационных полимеров, которые отобраны из базы данных CSD. Дескрипторы пористости были рассчитаны для 17 857 полностью определенных структур из этой выборки. Главный вопрос, на который нужно было ответить: *может ли топология каркаса предопределять пористость кристаллической структуры, и в какой степени?*

Пористость (P) является геометрическим свойством и определяется отношением пространства элементарной ячейки, не занятого атомами, к ее общему объему, и, на первый взгляд, не должна зависеть от топологических

свойств координационной сетки. Некоторые авторы¹⁰ пытались связать геометрические и топологические свойства МОКП и вводили термины «деформируемая» и «жесткая» топология, но не было предложено никаких строгих критериев для определения деформируемости или жесткости. Для ответа на поставленный вопрос нами были исследованы корреляции между топологией и пористостью (P) в четырех группах (квартилях) координационных соединений Q1, Q2, Q3 и Q4 с «низкой» ($P < 0.25$), «нормальной» ($P = 0.25-0.50$), «высокой» ($P = 0.50-0.75$) и «очень высокой» пористостью ($P > 0.75$), соответственно. Анализ данных позволил сделать следующие выводы:

- (i) Существуют топологии, которые обеспечивают только низкопористые или только высокопористые каркасы, в то время как многие другие топологии могут описывать каркасы любой пористости.
- (ii) Большинству топологий соответствует средняя пористость около 0.40–0.45 и большое ее рассеяние в кристаллических структурах ($3\sigma > 0.30$, где σ - среднеквадратичное отклонение пористости для выборки структур с заданной топологией), которое охватывает весь диапазон пористости. Такие топологии можно рассматривать как «деформируемые», поскольку они могут сохраняться при сильном геометрическом искажении сетки, обеспечивая тем самым разнообразие свободного пространства. Топологии с малым рассеянием пористости ($3\sigma < 0.06$) можно считать «жесткими»; все они встречаются редко (не более 17 структур данной топологии), что отражает их неспособность к геометрическим деформациям.
- (iii) Детальный анализ показывает, что все группы структур с малым рассеянием пористости представляют собой серии изоструктурных соединений с одним и тем же лигандом (обычно, это монолигандные комплексы) или структурно схожими лигандами и атомами металлов.

Если базовая сетка в наиболее симметричном вложении принимается за эталон, то все остальные вложения можно рассматривать как искажения эталона; то же самое касается и натуральных тайлов сеток. Мы обнаружили, что некоторые тайлы могут сильно деформироваться, тогда как другие в большинстве случаев остаются близкими к эталону (Рис. 15).

Как численный метод оценки гибкости каркаса предложено использовать критерий искажения тайлов Σ , который рассчитывается как усредненный квадрат отклонения вершин эталона x и образца x' :

$$\Sigma = \frac{1}{n} \sum_{i=1}^n \|x'_i - x_i\|^2, \quad (4)$$

¹⁰ Amombo Noa F. M., Grape E.S., Brülls S.M. et al. J. Am. Chem. Soc. 2020, 142(20), 9471–9481.

где n - количество вершин в тайле. При совмещении эталона и образца нами использован второй метод сравнения их формы (см. *первую задачу*).

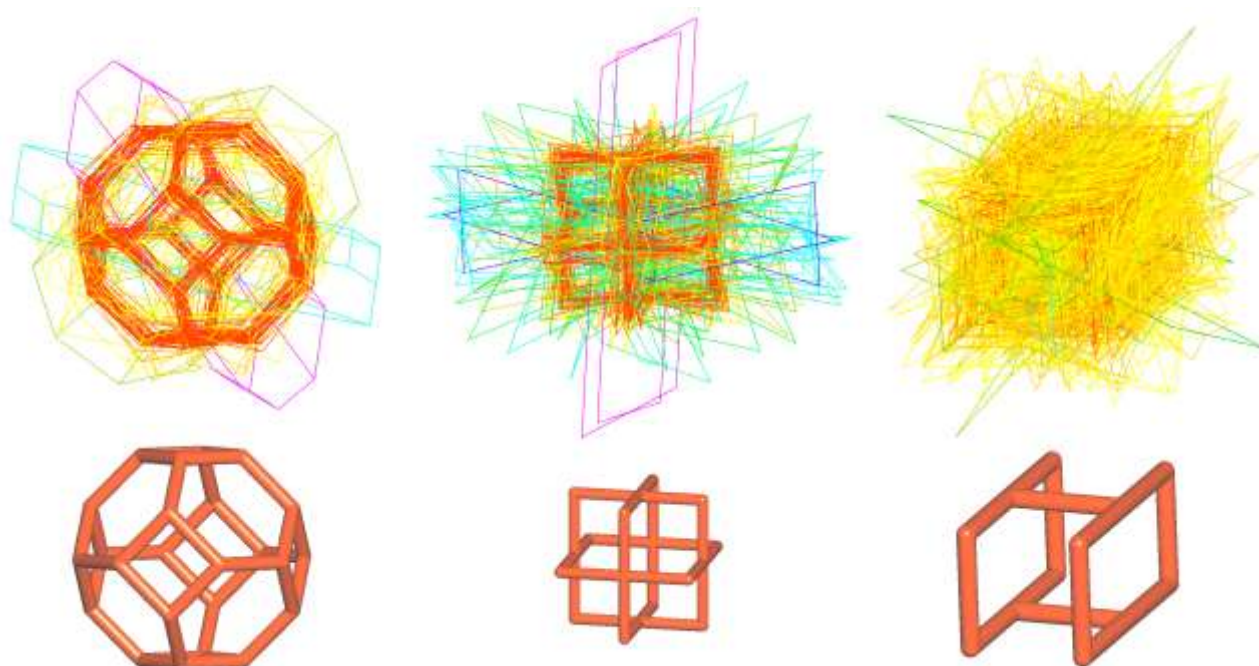


Рис. 15. Диаграммы перекрытия для геометрически искаженных тайлов в 165 **sod** (170 тайлов; $\Sigma = 0.49 \text{ \AA}^2$; слева), 86 **nbo** (137 тайлов; $\Sigma = 2.33 \text{ \AA}^2$; в центре) и 226 **cds** (324 тайла; $\Sigma = 2.64 \text{ \AA}^2$; справа), базовых сеток координационных полимеров, в сравнении с наиболее симметричными формами тайлов (в центре) и соответствующими натуральными тайлами (внизу). Окраска тайлов зависит от их геометрического искажения (Σ), от красного (небольшое искажение) до синего (большое искажение).

Для решения *четвертой задачи*, посвященной прогнозу степени окисления атомов металла, в качестве обучающих выборок были использованы набор координационных соединений (см. *вторую задачу*) и выборка неорганических структур из ICSD. Во всех рассмотренных кристаллических структурах сохранялся баланс зарядов ССБ (лигандов, атомов металлов и некоординированных частиц), что однозначно определяло степень окисления атомов металла в них. Кроме того, нами сформированы тестовые выборки I и II, состоящие из 2359 и 532 кристаллических структур координационных соединений, депонированных в базы данных CSD (в 2020-2021 гг.) и ICSD, соответственно. Данные о 228 атомах металлов, характеризующихся дробно-усредненными степенями окисления, были собраны в тестовую выборку III. Для независимой оценки качества модели данные из обучающих выборок не были включены в тестовые выборки.

При создании модели, прогнозирующей степень окисления, мы использовали пространство из 12 дескрипторов, характеризующих атомы и их окружение: AtNum, Group, Mult, CN, NZ, V_{CP} , G_3 , Δ_{CM} , R_{sd} , V_{VP} , S_{VP} , $\Sigma\Omega(O)$ (табл. 1). Оптимальные оценки критериев качества получены с помощью

метода случайного леса для 6 признаков (не учитывались дескрипторы Mult, NZ, V_{CP} , Δ_{CM} , V_{VP} and S_{VP}). Согласно результатам тестирования (Рис. 16), итоговая модель с точностью не менее 95% оценивает степень окисления атомов металла, а расхождения результатов прогноза с данными первоисточников по тестовым выборкам I и II незначительные (63 атомов в 46 структурах или 1.1% и 1.6%, соответственно).

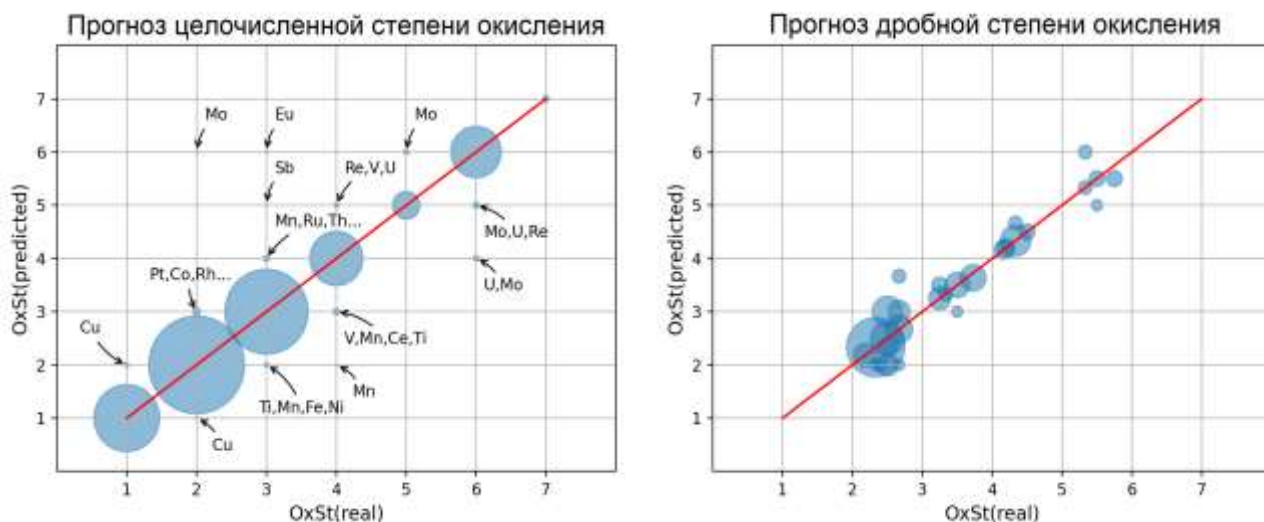


Рис. 16. Расхождения между реальными и предсказанными значениями OxSt для структур из тестовых выборок I и II (слева) и III (справа). Размеры окружностей пропорциональны возникновению конкретной комбинации двух значений OxSt; правильный прогноз соответствует красной диагонали.

Как и следовало ожидать, таким случаям отвечают переходные металлы, лантаноиды и актиниды (Рис. 16). Основными причинами найденных несоответствий оказались ошибочные значения степени окисления в первичном источнике и структурной базе данных, отсутствие нужного сочетания «металл – степень окисления» в обучающей выборке, а также наличие структурного разупорядочения атомов. Например, в структуре $\beta\text{-U}_3\text{O}_8$ (ICSD код 24906) модель прогнозирует у атомов U(1), U(2) и U(3) степени окисления 5, 6 и 5, как наиболее вероятные (68.6%, 98.0% и 77.5%), соответственно, что полностью согласуется с данными автора первоисточника¹¹.

Описанная прогностическая модель реализована нами в виде бесплатного веб-сервиса <https://oxidationstate.topcryst.com/>, который способен предсказывать целочисленные значения степени окисления атомов металла по стандартным кристаллографическим данным (формат CIF). Процесс расчета полностью автоматизирован. В результатах выводится набор возможных степеней окисления с вероятностями их реализации в данной кристаллической структуре.

¹¹ Loopstra B.O. Acta Cryst. Part B, 1970, 26, 656-657.

В Четвертой главе рассмотрено применение разработанных автором методов к химическим веществам, не являющимся КС и МОКП, для проверки универсальности предлагаемого геометрико-топологического подхода. Были исследованы кристаллические структуры простых веществ, твердых электролитов и интерметаллидов.

Оценка размера атомов в структуре простых веществ выполнена на основании модели атомов, представленных в виде мягких сфер. Для отобранных из ICSD 2020.2 данных по 1249 структурным исследованиям кристаллов простых веществ, определенных при атмосферном давлении, изучены зависимости геометрических и топологических дескрипторов атомов от порядкового номера соответствующего элемента в Периодической системе. В качестве дескрипторов рассмотрены координационное число, радиус сферического домена, Ван-дер-Ваальсовый радиус атомов, а также топологический тип кристаллической структуры. Целевым дескриптором являлся размер атомов, оцениваемый радиусом сферического домена, значения которого нами впервые были определены для всех элементов в кристаллах простых веществ при нормальных условиях. Выполнено сравнение полученных величин R_{SD} с различными системами атомных радиусов, что позволило найти совпадения значений радиусов атомов в моделях жестких или мягких сфер.

Дополнительно, было изучено изменения размера атомов инертных газов при фазовых переходах, который оценивался по величине R_{SD} (в Å), вычисляемой из плотности вещества ρ (в г/см³) и атомной массы M (в г/моль) по формуле (5).

$$R_{SD} = \sqrt[3]{\frac{1.245 \cdot M}{\rho \cdot \pi}} \quad (5)$$

Были учтены данные о плотностях инертных газов при различных температурах в твердой, жидкой и газообразной фазах, а также в критической точке. Собранная информация позволила уточнить методом наименьших квадратов коэффициенты C и B уравнения (6) зависимости размера их атомов в жидкой фазе от температуры¹², а также найти взаимосвязи между твердым и жидким фазовыми состояниями. В частности, было установлено (Рис. 17), что для всех атомов инертных газов, кроме гелия, экстраполяция указанных зависимостей к абсолютному нулю дает величину, которая близка к значениям R_{SD} тех же атомов в кристалле при $T = 0^\circ \text{ K}$.

$$\frac{\rho}{\rho_c} - 1 = C \left(1 - \frac{T}{T_c}\right) + B \left(1 - \frac{T}{T_c}\right)^{1/3} \quad (6)$$

¹² Филиппов Л.П. Методы расчета и прогнозирования свойств веществ. М.: Изд-во МГУ, 1988, 252 с.

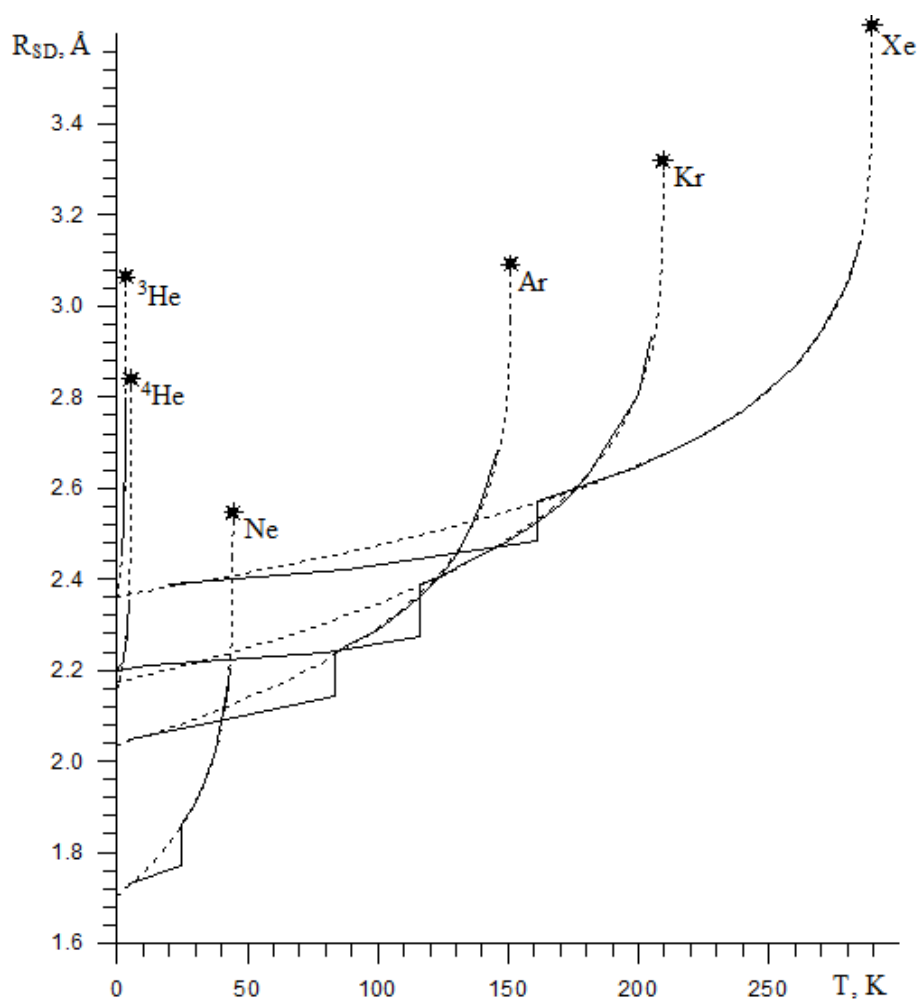


Рис.17. В виде сплошных линий показаны экспериментальные зависимости $R_{SD}(T)$ для жидкого и твердого состояний инертных газов. Штриховыми линиями представлены рассчитанные на основании уравнения (6) методом наименьших квадратов графики функций $R_{SD}(T)$ для жидкого состояния.

Впервые, выполнено сравнение размера атомов в простых веществах, оцененное с позиции модели мягких сфер, с длинами волн де Бройля $\lambda_{дБ}^i$, которые, по аналогии с работой¹³, рассчитывались из i -ых потенциалов ионизации изолированных атомов по формуле (7).

$$\lambda_{дБ}^i = \frac{12.26}{\sqrt{I_i}} \quad (7)$$

Волны де Бройля связаны со стационарными круговыми орбитами, рассчитанными из условия квантования Бора, согласно которому на длине окружности орбиты укладывается целое число длин волн де Бройля. В результате мы имеем стоячую волну на электронной орбите, которая не переносит и не излучает энергию. Нами впервые обнаружены корреляции между длинами волн де Бройля $\lambda_{дБ}^i$ и радиусами сферического домена R_{SD} атомов инертных газов в кристаллическом, жидком состояниях и в

¹³ Bondi A. J. Phys. Chem. 1964. V.68. N.3. P.441-451.

критической точке. Так для атомов Ne, Ar, Kr, Xe, Rn в критической точке (2.55, 3.09, 3.32, 3.62, 3.79 Å) R_{SD} оказалось близко по значению к длине волны де Бройля, рассчитанной из их первого потенциала ионизации I_1 (2.64, 3.09, 3.28, 3.52, 3.74 Å, соответственно). Корреляции с $\lambda_{дБ}^i$ нами обнаружены не только для инертных газов, но и для кристаллов N_2 , O_2 и F_2 . Установлено, что R_{SD} атомов азота, кислорода и фтора коррелирует, как с длиной волны де Бройля, рассчитанной из третьего потенциала ионизации, так и с Ван-дер-Ваальсовыми радиусами этих атомов.

Скрининг потенциальных твердых электролитов. На основании разработанного автором геометрико-топологического подхода впервые предложен метод анализа каналов ионной проводимости. Для их определения использована реберная сетка полиэдров Вороного, рассчитанная для атомов структурной сетки без учета проводящих катионов. Подвижные катионы занимают при миграции пустоты каркаса (или позиции вершин полиэдров Вороного), если их размер соответствует размеру катиона, оцененному по усредненному значению его радиуса сферического домена R_{SD} в подобном окружении.

Метод был успешно применен при определении возможной ионной проводимости в кристаллических структурах состава $Na_xM_yO_z$, каркас которых сформирован ионами металла M^{n+} и кислорода O^{2-} . Величина R_{SD} катионов Na^+ в кислородном окружении составила 1.6 Å. Дополнительным важным критерием отбора было отсутствие катионов M^{n+} на поверхности канала, иначе возникают энергетические барьеры при миграции ионов Na^+ , поэтому отбирали только структуры, поверхность каналов которых образована анионами O^{2-} .

В итоге, среди 402 выбранных из ICSD кристаллических структур были идентифицированы не только все известные Na-ионные проводники (Na-β-глинозем $Na_{1.68}Al_{11}O_{17}$, NaSiCon $Na_{3.05}Zr_2Si_{2.05}P_{0.95}O_{12}$), но и новые потенциальные твердые электролиты, содержащие подвижные катионы натрия (Рис. 18).

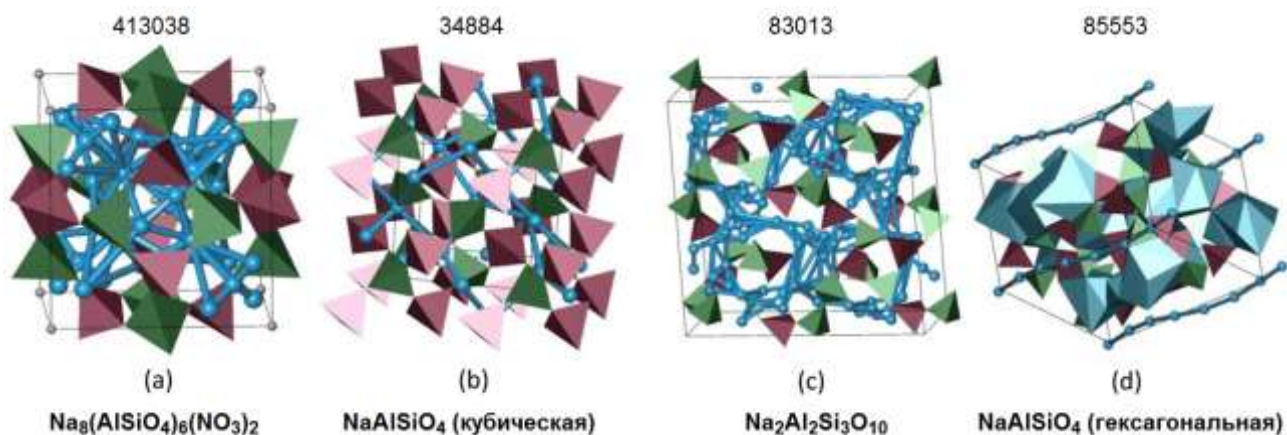


Рис. 18. Четыре наилучших (согласно используемому алгоритму) новых потенциальных Na-ионных проводника. Все они являются алюмосиликатами и содержат AlO_4 (коричневые тетраэдры), SiO_4 (зеленые тетраэдры), NaO_x (синие полиэдры). Атомы кислорода обозначены серыми шарами. Синие шары и цилиндры отвечают выявленным пустотам и каналам.

Топологический анализ структур интерметаллидов. Предлагаемый автором геометрико-топологический подход был использован для классификации более 38 000 интерметаллических соединений по 3708 топологическим типам. При этом интерметаллиды рассматривались на различных уровнях структурной организации: 1) координационные полиэдры; 2) многослойные кластеры; 3) двумерные сетки; 4) трехмерные каркасы.

При анализе координационных полиэдров было показано, что 12-вершинные многогранники являются наиболее частыми (33%) и что почти половина из них имеет форму икосаэдра (46%), за ней следует кубооктаэдр (25%) и, что неожиданно, двухшапочная пятиугольная призма (13%). Для всех 38 000 структур было рассчитано представление интерметаллида в виде сборки из полиоболочечных (луковичных) нанокластеров (Рис. 19), которые были сохранены в TTN коллекции. Результаты анализа данных коллекции показали, что 41% интерметаллидов можно собрать с помощью одного нанокластера и что 22.4% из них упакованы в соответствии с гранецентрированным кубическим мотивом плотнейшей упаковки сфер в трехмерном пространстве. При анализе топологии слоев установлено, что гексагональная решетка, которая соответствует плотнейшей упаковке сфер на плоскости, существует в более чем 11 тыс. кристаллических структур, подтверждая плотноупакованный характер интерметаллидов.

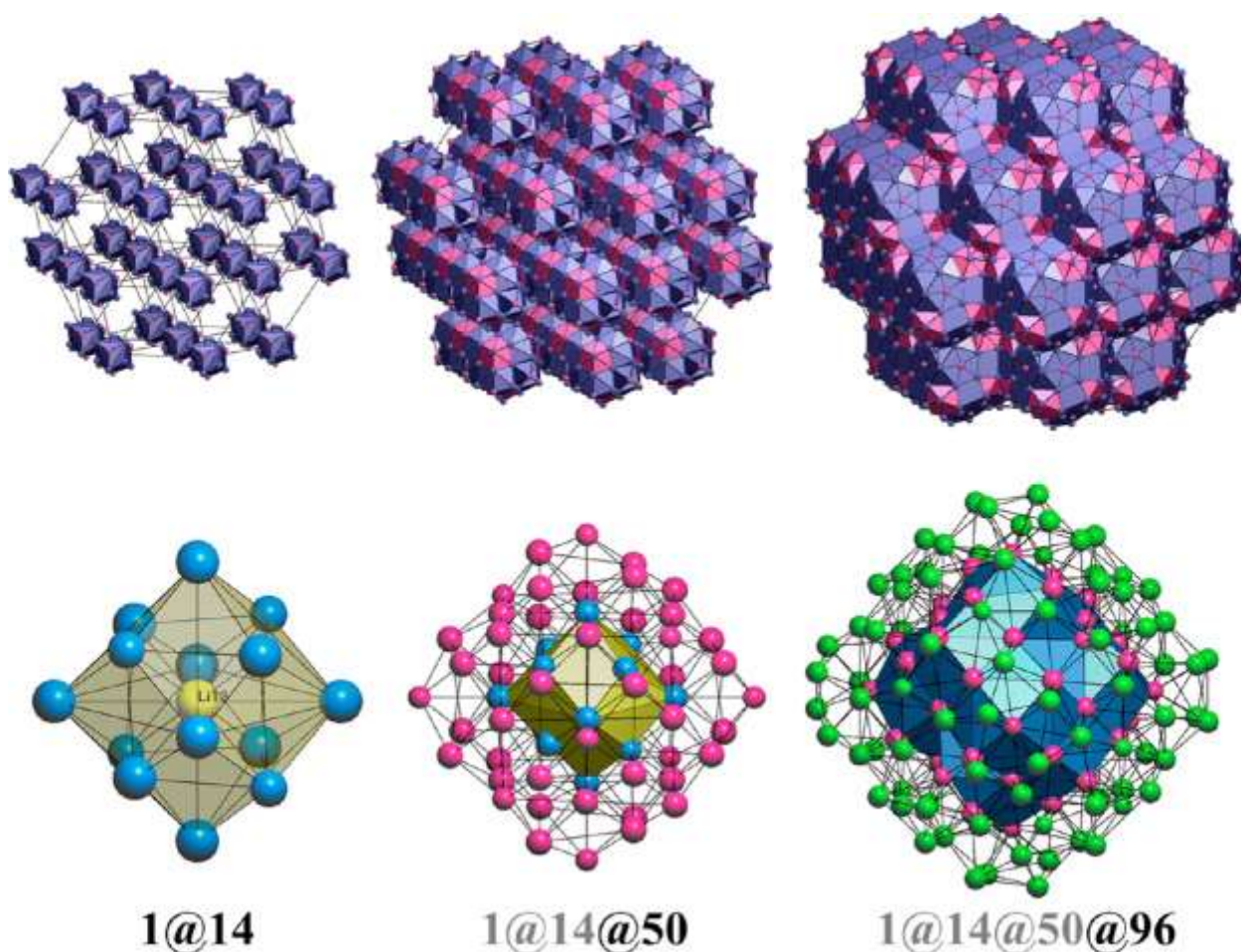


Рис. 19. Послойный рост нанокластеров в кристаллической структуре $\text{Li}_{17}\text{Ge}_4$: (слева) первая 14-атомная оболочка; (в центре) вторая 50-атомная оболочка; (справа) третья 96-атомная оболочка.

Таким образом, проведенные исследования показали универсальность предлагаемого нами геометрико-топологического подхода при анализе структур разной химической природы, а также возможность использования геометрико-топологических дескрипторов для поиска корреляций с различными физическими свойствами кристаллов. Строгость алгоритмов и их реализация в виде компьютерных программ позволяют в автоматическом режиме обрабатывать большие массивы структурной информации, накапливать структурные дескрипторы в электронных хранилищах. В свою очередь, анализ собранных данных с помощью современных интеллектуальных методов и машинного обучения позволяют обнаруживать новые корреляции «состав-структура-свойство» и прогнозировать свойства, что может быть применено для дизайна новых практически значимых материалов.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ

Выполненные диссертантом исследования позволили получить следующие результаты и выводы:

1. На основе разработанного геометрико-топологического подхода предложена универсальная методология машинного анализа кристаллических структур, позволяющая прогнозировать их строение и физические свойства, базируясь на имеющейся структурной информации с использованием современных методов интеллектуального анализа данных, а также представления атомов в виде полиэдров Вороного или Ван-дер-Ваальсовых полиэдров, кристаллических структур в виде периодических сеток. В рамках данной методологии предложены новые дескрипторы структурных строительных блоков, разработаны машинные модели для прогноза степени окисления атомов металлов, а также размерности и топологии периодических сеток, с точностью прогноза 95%, 90% и 75%, соответственно. Разработанная методология, дескрипторы и машинные модели реализованы в программном обеспечении и веб-сервисах, которые используются во всем мире.
2. В рамках предложенной методологии структурного анализа разработаны и реализованы в виде компьютерной программы три новых алгоритма определения формы координационной фигуры атома или структурной группировки, основанные на топологии полиэдров Вороного, максимизации объема пересечения координационных полиэдров и сравнении угловых отпечатков, которые позволили найти ряд правил и закономерностей строения пятикоординационных соединений меди и цинка в N,O-окружении.
3. При помощи разработанных программных средств созданы информационные хранилища структурных строительных блоков и их дескрипторов для 1 928 816 атомов металла, 269 232 лигандов в структуре координационных соединений, 309 353 органических молекул, 1 234 нанокластеров, а также 808 445 топологических типов кристаллических структур и 2 675 446 их представителей. Все обучающие и тестовые выборки для машинного обучения были сформированы на основании этих данных. Доступ к хранилищу через Интернет реализован в виде веб-сервисов на сайте <https://topcryst.com>.
4. Предложен новый метод оценки пористости кристаллических структур с помощью Ван-дер-Ваальсовых полиэдров атомов. Метод использован для изучения взаимосвязей доли свободного пространства в кристаллической структуре с ее топологическим типом в 33 790 каркасных координационных соединениях. Показано, что существуют топологии периодических сеток, которые обеспечивают только низкую или только высокую пористость соответствующих атомных каркасов, в то время как многие другие топологии

могут описывать каркасы любой пористости. На основании разнообразия формы полостей атомного каркаса, представленных в виде натуральных тайлов, предложен численный метод оценки гибкости каркаса.

5. Показано, что геометрико-топологические методы являются универсальными и могут быть применены не только для структур координационных соединений, но и других классов химических веществ. На основании предлагаемого подхода осуществлен отбор 402 потенциальных Na-ионных твердых электролитов, а также подтвержден плотноупакованный характер интерметаллидов по результатам анализа коллекции нанокластеров и топологий слоев кристаллических структур. Обнаружены корреляции радиуса сферического домена атомов в простых веществах с длинами электронных волн де Бройля, рассчитанных из потенциалов ионизации изолированных атомов, что подтвердило правомерность модели «мягких сфер», то есть представления атомов в кристаллах и жидкостях, как деформируемых сфер, полностью заполняющих все пространство.

Основное содержание диссертации изложено в следующих работах:

Главы в монографиях

1. **Shevchenko A.P.**, Alexandrov E.V., Blatova O.A., Yablokov D.E., Blatov V.A. *From topological descriptors to expert systems: a route to predictable materials*. 2019, In book: Materials Informatics: Methods, Tools and Applications. ISBN: 978-3-527-34121-4.
2. Blatov V.A., Alexandrov E.V., **Shevchenko A.P.** 2.21 - *Topology: ToposPro*. In Comprehensive Coordination Chemistry III (Third Edition). Elsevier, 2021, 389-412.

Статьи

3. Блатов В.А., **Шевченко А.П.**, Сережкин В.Н. *Автоматизация кристаллохимического анализа комплекс компьютерных программ TOPOS*. // Координац. химия. 1999. Т.25. №7. С.483-497.
4. **Шевченко А.П.**, Сережкин В.Н., Михайлов Ю.Н. *Координационные полиэдры ThO_n в структуре кристаллов* // Радиохимия 2000, Т. 42, N.1, 32-35.
5. **Шевченко А.П.** *Метрика атомных доменов в кристаллах He, Ne, Ar, Kr, Xe, Rn* // Журн. физ. химии. 2001. Т.75, №9. С.1651-1655.
6. **Шевченко А.П.**, Сережкин В.Н., Михайлов Ю.Н. *Анализ координации атомов Pa, Pu, Am, Cm, Bk и Cf в кислородсодержащих соединениях с помощью полиэдров Вороного-Дирихле* // Радиохимия 2001, Т. 43, N. 1, 32-36.
7. **Шевченко А.П.**, Сережкин В.Н. *Домены атомов актинидов в кристаллах фторсодержащих соединений* // Радиохимия 2002, Т. 44, N. 1, 5-9.
8. **Шевченко А.П.** *Домены атомов в кристаллах элементов второго периода* // Журн. физ. химии. 2002. Т. 76. №2. С. 291–294.

9. **Шевченко А.П.**, Сerezжкин В.Н. *Стереoaтомная модель и строение кристаллов насыщенных углеводов* // Журн. физ. химии. 2004. Т. 78. № 10. С.1817-1825.
10. **Шевченко А.П.** *Особенности стереохимии sp^2 -гибридных атомов углерода в кристаллах ненасыщенных углеводов*. // Вестник Самарского государственного университета, 2005, №3. 184-195.
11. Сerezжкин В.Н., Сerezжкина Л.Б., **Шевченко А.П.**, Пушкин Д.В. *Новый метод анализа межмолекулярных взаимодействий в структуре кристаллов: насыщенные углеводороды* // Журн. физич. химии. 2005. Т. 79. №6. С. 1058-1069.
12. Сerezжкин В.Н., **Шевченко А.П.**, Сerezжкина Л.Б., Прокаева М.А. *Новый метод анализа межмолекулярных взаимодействий в структуре кристаллов: ненасыщенные углеводороды* // Журн. физич. химии. 2005. Т. 79. № 6. С. 1070-1080.
13. Биканина Е.А., **Шевченко А.П.**, Сerezжкин В.Н., *Координационные полиэдры KCl_n в структуре кристаллов* // Координац. химия. 2005. Т. 31. №1. С.72-80.
14. Сerezжкин В.Н., **Шевченко А.П.**, Сerezжкина Л.Б. *Новый метод анализа межмолекулярных взаимодействий в структуре кристаллов: π - комплексы* // Координац. химия. 2005. Т. 31. № 7. С.495-505.
15. Вологжанина А.В., Пушкин Д.В., **Шевченко А.П.**, Сerezжкин В.Н. *Особенности стереохимии цетрия в структуре кислородсодержащих соединений* // Журн. неорган. химии. 2005. Т.50. № 4. С.660-666.
16. Сerezжкин В.Н., Сerezжкина Л.Б., **Шевченко А.П.**, Пушкин Д.В. *Взаимосвязь между энтальпией сублимации и структурой кристаллов насыщенных углеводов* // Журн. физич. химии. 2005. Т. 79. № 6. С. 1058-1069.
17. Blatov V.A., **Shevchenko A.P.**, Serezhkin V.N. *TOPOS 3.1. Program package for multipurpose geometrical and topological analysis of crystal structures* // J. Appl. Cryst. 1999, 32, 377.
18. Blatov V.A., **Shevchenko A.P.**, Serezhkin V.N. *TOPOS 3.2 – a new version of the program package for multipurpose crystal-chemical analysis* // J. Appl. Cryst., 2000, 33, 1193.
19. Blatov V.A., **Shevchenko A.P.** *Analysis of voids in crystal structures: the methods of 'dual' crystal chemistry* // Acta Cryst. 2003, A59, 34-44.
20. Virovets, A.V., Blatov V.A., **Shevchenko A.P.** *Methods of crystallochemical analysis of supramolecular complexes by means of Voronoi-Dirichlet polyhedra: A study of cucurbituril host-guest compounds* // Acta Crystallographica Section B: Structural Science 2004, 60(3), 350-357.
21. Blatov V.A., **Shevchenko A.P.**, Proserpio D.M. *Applied Topological Analysis of Crystal Structures with the Program Package ToposPro* // Cryst. Growth Des. 2014, 14(7), 3576-3586. DOI: [10.1021/cg500498k](https://doi.org/10.1021/cg500498k)

22. Alexandrov E.V., **Shevchenko A.P.**, Asiri A.M., Blatov V.A. *New knowledge and tools for crystal design: local coordination versus overall network topology and much more* // Cryst. Eng. Comm. 2015, 17, 2913-2924. DOI: [10.1039/C4CE02418D](https://doi.org/10.1039/C4CE02418D)
23. Meutzner F., Münchgesang W., Leisegang T., Schmid R., Zschornak M., Ureña de Vivanco M., **Shevchenko A.P.**, Blatov V.A., Meyer D.C. *Identification of solid oxygen-containing Na-electrolytes: An assessment based on crystallographic and economic* // Cryst. Res. Technol., 1–12 (2016). DOI: [10.1002/crat.201600223](https://doi.org/10.1002/crat.201600223)
24. **Shevchenko A.P.**, Blatov I.A., Kitaeva E.V., Blatov V.A. *Local Coordination versus Overall Topology in Crystal Structures: Deriving Knowledge from Crystallographic Databases* // Cryst. Growth Des., 2017, 17 (2), pp 774–785. DOI: [10.1021/acs.cgd.6b01630](https://doi.org/10.1021/acs.cgd.6b01630)
25. Akhmetshina T.G., Blatov V.A., Proserpio D.M., **Shevchenko A.P.** *Topology of Intermetallic Structures: From Statistics to Rational Design* // Accounts of Chemical Research, 2018, 51 (1), 21–30. DOI: [10.1021/acs.accounts.7b00466](https://doi.org/10.1021/acs.accounts.7b00466)
26. Alexandrov E.V., **Shevchenko A.P.**, Blatov V.A. *Topological databases: why do we need them for design of coordination polymers?* // Cryst. Growth Des. 2019, 19, 5, 2604–2614. DOI: [10.1021/acs.cgd.8b01721](https://doi.org/10.1021/acs.cgd.8b01721)
27. Gomez G.E., Ridenour J.A., Byrne N.M., **Shevchenko A.P.**, Cahill C.L. *Novel Heterometallic Uranyl-Transition Metal Materials: Structure, Topology, and Solid State Photoluminescence Properties* // Inorg. Chem., 2019, 58, 11, 7243–7254. DOI: [10.1021/acs.inorgchem.9b00255](https://doi.org/10.1021/acs.inorgchem.9b00255)
28. Blatov I.A., Kitaeva E.V., **Shevchenko A.P.**, Blatov V.A. *A universal algorithm for finding the shortest distance between systems of points* // Acta Cryst. 2019, A75, 827-832. DOI: [10.1107/S2053273319011628](https://doi.org/10.1107/S2053273319011628)
29. **Shevchenko A.P.**, Eremin R.A. and Blatov V.A. *The CSD and knowledge databases: from answers to questions* // CrystEngComm, 2020, 22, 7298–7307. DOI: [10.1039/d0ce00265h](https://doi.org/10.1039/d0ce00265h)
30. **Shevchenko A.P.**, Alexandrov E.V., Golov A.A., Blatova O.A., Duyunova A.S., Blatov V.A. *Topology versus porosity: what can reticular chemistry tell us about free space in metal–organic frameworks?* // Chem. Commun., 2020, 56, 9616-9619. DOI: [10.1039/D0CC04004E](https://doi.org/10.1039/D0CC04004E)
31. Blatov V.A., **Shevchenko A.P.** *Simplify to understand: how to elucidate crystal structures?* // Struct. Chem. 2021, 32, 507–519. DOI: [10.1007/s11224-020-01724-4](https://doi.org/10.1007/s11224-020-01724-4)
32. Nicholas T.C., Alexandrov E.V., Blatov V.A., **Shevchenko A.P.**, Proserpio D.M., Goodwin A.L., Deringer V.L. *Visualization and Quantification of Geometric Diversity in Metal–Organic Frameworks* // Chem. Mater. 2021, 33, 8289-8300. DOI: [10.1021/acs.chemmater.1c024398](https://doi.org/10.1021/acs.chemmater.1c024398)
33. Alexandrov E.V., **Shevchenko A.P.**, Nekrasova N.A., Blatov V.A. *Topological methods of analysis and design of coordination polymers* // Russ. Chem. Rev., 2022, 91 (4), RCR5032. DOI: [10.1070/RCR5032](https://doi.org/10.1070/RCR5032)

34. **Shevchenko A.P.**, Smolkov M.I., Wang J., Blatov V.A. *Mining Knowledge from Crystal Structures: Oxidation States of Oxygen-Coordinated Metal Atoms in Ionic and Coordination Compounds* // J. Chem. Inf. Model., 2022, 62, 2332–2340. DOI: [10.1021/acs.jcim.2c00080](https://doi.org/10.1021/acs.jcim.2c00080)
35. **Shevchenko A.P.**, Shabalin A.A., Karpukhin I.Yu., Blatov V.A. *Topological representations of crystal structures: generation, analysis and implementation in the TopCryst system* // Science and Technology of Advanced Materials: Methods, 2022, 2:1, 250–265. DOI: [10.1080/27660400.2022.2088041](https://doi.org/10.1080/27660400.2022.2088041)
36. **Shevchenko A.P.**, Blatov V.A. *Stability of inorganic ionic structures: the uniformity approach* // Acta Cryst., 2024, A80, 446–456. DOI: [10.1107/S2053273324009008](https://doi.org/10.1107/S2053273324009008)
37. Smolkov M.I., Barabanova E.D., Plotnik U.S., **Shevchenko A.P.** *Mining Knowledge from Crystal Structures of Homoleptic Complexes: Oxidation States of Metal Atoms* // Pattern Recognition and Image Analysis, 2024, 34, 3, 877–891. DOI: [10.1134/S1054661824700779](https://doi.org/10.1134/S1054661824700779)

Зарегистрированные ПРОГРАММЫ ДЛЯ ЭВМ и БАЗЫ ДАННЫХ

1. Блатов В.А., Сережкин В.Н., **Шевченко А.П.** ПРОГРАММА ДЛЯ ЭВМ “*Topos*”, 1997, № 970374.
2. Блатов В.А., **Шевченко А.П.** ПРОГРАММА ДЛЯ ЭВМ “*ToposPro*”, 2013, № 2013619016.
3. Блатов В.А., **Шевченко А.П.** БАЗА ДАННЫХ “*Topological Types Collections (TTCollections)*”, 2014, № 2014621100.
4. **Шевченко А.П.**, Блатов В.А. ПРОГРАММА ДЛЯ ЭВМ “*StatPack*”, 2015, № 2015661864.
5. Блатов В.А., **Шевченко А.П.**, Александров Е.В., Яблоков Д.Е. БАЗА ДАННЫХ “*Topological Types Collections Online (TTCollections Online)*”, 2016, № 2016620235.
6. Блатов В.А., **Шевченко А.П.** БАЗА ДАННЫХ “*Topological Types of Crystal Structures*”, 2016, № 2016620105.
7. Яблоков Д.Е., **Шевченко А.П.**, Блатов В.А. ПРОГРАММА ДЛЯ ЭВМ “*Topology Online Research Information System (TORIS)*“, 2017, № 2017613992.
8. **Шевченко А.П.**, Блатов В.А. БАЗА ДАННЫХ “*Coordination of Metals (CoMet)*”, 2017, № 2017620380.
9. Александров Е.В., Еремин Р.А., **Шевченко А.П.**, Блатов В.А. БАЗА ДАННЫХ “*TTL (Topological Types of Ligands)*”, 2018, № 2018620268.
10. Голов А.А., Александров Е.В., **Шевченко А.П.**, Блатов В.А. БАЗА ДАННЫХ “*CaCh (Cavities and Channels)*”, 2018, № 2018620301.
11. Блатов И.А., Китаева Е.В., **Шевченко А.П.**, Блатов В.А. ПРОГРАММА ДЛЯ ЭВМ “*CFShape*”, 2018, № 2018614596.

12. **Шевченко А.П.**, Карпухин И.Ю., Некрасова Н.А., Александров Е.В., Блатов В.А. БАЗА ДАННЫХ “*Underlying net topologies*”, 2018, № 2018621903.
13. Еремин Р.А., Золотарев П.Н., Александров Е.В., **Шевченко А.П.** БАЗА ДАННЫХ “*Композиционное/конфигурационное пространство катодного материала $\text{LiNi}_{0.8}\text{Co}_{0.167}\text{Al}_{0.083}\text{O}_2$* ”, 2018, № 2018621053.
14. Яблоков Д.Е., Шабалин А.А., Александров Е.В., **Шевченко А.П.**, Блатов В.А. ПРОГРАММА ДЛЯ ЭВМ (ВЕБ-СЕРВИС) “*Search for underlying topologies and their occurrences in crystal structures*”, 2019, № 2019611876.
15. Голов А.А., Александров Е.В., **Шевченко А.П.**, Блатов В.А. БАЗА ДАННЫХ “*Database of Metal-Organic and Organic Building Blocks*”, 2020, № 2020622463.
16. Голов А.А., **Шевченко А.П.**, Блатов В.А., Карпухин И.Ю., Шабалин А.А., Александров Е.В. ПРОГРАММА ДЛЯ ЭВМ (ВЕБ-СЕРВИС) “*TopCryst.find_topology*”, 2021, № 2021611244.
17. Кабанов А.А., Шабалин А.А., **Шевченко А.П.**, Блатов В.А., Морхова Е.А., Кабанова Н.А., Карпухин И.Ю., БАЗА ДАННЫХ “*Battery Materials Database (batterymaterials.info)*”, 2021, № 2021621557.
18. Смольков М.И., Шабалин А.А., Карпухин И.Ю., **Шевченко А.П.**, Блатов В.А. ПРОГРАММА ДЛЯ ЭВМ (ВЕБ-СЕРВИС) “*OxidationStatePredictor*”, 2023, № 2023611096.

Используемые обозначения и сокращения:

CSD (Cambridge Structural Database) – Кембриджская структурная база данных

ICSD (Inorganic Crystal Structure Database) – База данных неорганических структур

SD (Spherical Domain) – сферический домен

DMF (N,N-dimethylformamide) – диметилформамид

КТ – координационный тип лиганда

КФ – координационная фигура атома или узла базовой сетки

КЧ – координационное число атома или узла базовой сетки

ССБ – структурный строительный блок

КС – координационное соединение

МОКП – металлоорганический координационный полимер

ИИ – искусственный интеллект

Автор выражает глубокую благодарность и признательность профессору Блатову Владиславу Анатольевичу за всестороннюю поддержку, ценные советы и идеи, помощь и консультации на разных этапах выполнения работы. Большая благодарность от автора Сережкину Виктору Николаевичу за знакомство с удивительным симметричным миром кристаллохимии и строения вещества, а также Балаковскому Сергею Львовичу и Антошкиной Светлане Миленитиновне за направление в самые дебри информационных технологий и за помощь в их успешном прохождении.

Автор также выражает благодарность проф. Давиде М. Прозернито (Миланский университет, Италия), Фалку Мейцнеру (Фрайбергская горная академия, Германия), Джорджу Костакису (Университет города Сассекс, Великобритания) за новые идеи и полезные рекомендации при разработке программных продуктов.

За помощь в выполнении диссертационной работы автор отдельно благодарит своих коллег из МНИЦТМ Самарского университета: Е. В. Александрова, А. А. Шабалина, А. А. Голова, И. В. Медриш, М. В. Пескова, Т. Г. Ахметшину, Н. А. Кабанову, П. Н. Золотарева.

Автор выражает сердечную благодарность своей жене Татьяне Александровне Шевченко за понимание и поддержку во время выполнения диссертационной работы.

Работа выполнена при финансовой поддержке Мегагранта №14.В25.31.0005 Министерства образования и науки Российской Федерации, а также грантов Российского фонда фундаментальных исследований №№ 01-07-90092, 17-43-630619, 18-07-00183, 18-29-04010, 20-13-50007 и Российского Научного Фонда №№ 16-13-10158, 19-73-10026, 23-23-00387.

Автореферат отпечатан с разрешения диссертационного Совета 24.2.377.03
ФГБОУ ВО «Самарский государственный технический университет»
(протокол № 10 от 24 декабря 2024 г.
Заказ № ____ . Тираж 100 экз.
Формат 60×84/16. Отпечатано на ризографе.

ФГБОУ ВО «Самарский государственный технический университет»
Отдел типографии и оперативной печати
443100, г. Самара, ул. Молодогвардейская, 244