

Федеральное государственное автономное образовательное учреждение
высшего образования «Сибирский федеральный университет» (СФУ)

На правах рукописи

Залого Александр Николаевич



АВТОМАТИЗАЦИЯ МЕТОДА ПОЛНОПРОФИЛЬНОГО
АНАЛИЗА ПОЛИКРИСТАЛЛОВ С ИСПОЛЬЗОВАНИЕМ
ГЕНЕТИЧЕСКИХ АЛГОРИТМОВ

Специальность 01.04.01 – приборы и методы экспериментальной физики

Диссертация на соискание ученой степени
кандидата физико-математических наук

Научный руководитель
доктор физико-математических наук
профессор И. С. Якимов

Красноярск – 2017

Содержание

Введение.....	4
Глава 1. Обзор методов порошковой дифракции.....	8
1.1. Общая характеристика современного состояния методов ПД	8
1.2. Этапы определения кристаллической структуры	12
1.3. Межлабораторные испытания по определению кристаллической структуры.....	13
1.4. Определение параметров решетки и пространственной группы.....	15
1.4.1. Процедура индирования дифрактограмм.....	15
1.4.2. Программы индирования.....	17
1.4.3. Определение пространственной группы	21
1.5. Классические методы определения кристаллических структур по данным порошковой дифракции	23
1.5.1. Задача определения кристаллической структуры	23
1.5.2. Метод Паттерсона.....	24
1.5.3. Прямые методы	26
1.6. Методы определения кристаллических структур в прямом пространстве.....	28
1.6.1. Методы поиска по сетке.....	30
1.6.2. Методы Монте-Карло.....	31
1.6.3. Методы имитации отжига.....	34
1.6.4. Методы эволюционных и генетических алгоритмов	38
1.6.5. Гибридные подходы	44
1.6.6. Перспективы развития.....	47
1.7. Генетические алгоритмы как средство решения оптимизационных задач	48
1.7.1. Основы теории генетических алгоритмов.....	48
1.7.2. Получение фенотипа из генотипа	54
1.8. Полнопрофильный анализ методом Ритвельда	59
1.9 Сравнительная оценка вариантов возможных решений и выбор направления развития генетического алгоритма эволюционной оптимизации	61
Глава 2. Мультипопуляционный генетический алгоритм моделирования кристаллической структуры из рентгенодифракционных данных.....	64
2.1 Схема работы индивидуальных процессов ГА в составе МПГА	64
2.1.1 Предварительная установка профильных параметров.....	64
2.1.2 Описание генетических операторов	65
2.1.3 Алгоритмы штрафования генерируемых структурных моделей.....	67
2.1.4 Локальная оптимизация индивидов в популяциях МПГА	69
2.2 Схема взаимодействия индивидуальных ГА в МПГА	70
2.3 Программное обеспечение, реализующее алгоритм МПГА	71
2.3.1 Задание профильных и структурных параметров	72
2.3.2 Задание параметров процесса МПГА	73
2.3.3 Работа с молекулами, заданными в виде жестких фрагментов.....	76
2.3.4 Контроль за процессом работы программы МПГА	76
2.3.5 Сохранение результатов работы МПГА.....	78
2.4 Результаты апробации МПГА на тестовых структурах	79
2.4.1 Описание процесса тестирования МПГА.....	79
2.4.2 Обсуждение результатов тестирования МПГА	82

2.4.3 Выводы по результатам апробации МПГА.....	92
2.5 Сравнение МПГА с современными методами определения кристаллических структур в прямом пространстве	94
2.5.1 Определение структуры Ba_2CrO_4	94
2.5.2 Определение структуры $\text{Pd}(\text{NH}_3)_2(\text{NO}_2)_2$	97
2.5.3 Определение структуры $\text{Er}_{10}\text{W}_2\text{O}_{21}$	101
2.5.4 Выводы по результатам сравнения	104
Глава 3. Определение неизвестных кристаллических структур при помощи МПГА.....	105
3.1 Описание объектов исследования.....	105
3.2 Описание методики исследования.....	105
3.3 Определение структуры соединения $[\text{Pt}(\text{NH}_3)_5\text{Cl}]\text{Br}_3$	106
3.4 Определение структуры соединения $[\text{Pd}(\text{CH}_3\text{NH}_2)_4][\text{PdBr}_4]$	114
Выводы и основные результаты работы	118
Список источников	119
Приложение А. МПГА с использованием синхротронных дифракционных данных.....	130
Приложение Б. Карты распределения атомов в популяциях в процессе сходимости ГА .	137

Введение

Актуальность темы.

Информация об атомной кристаллической структуре веществ накапливается в структурных базах данных и используется для объяснения и прогнозирования физических и химических свойств веществ и материалов. Для веществ, получаемых в поликристаллической форме, основным методом изучения структуры служит рентгеновская порошковая дифракция. Исследование структуры включает определение приближенной модели структуры и ее оптимизацию. Оптимизация выполняется с помощью метода полнопрофильного анализа Ритвельда, основанного на расчете дифрактограмм из структурных моделей и минимизации относительной невязки между расчетной и экспериментальной дифрактограммами (профильного R-фактора) по нелинейному МНК. Поиск же моделей кристаллической структуры новых порошковых веществ остается значительной проблемой.

В последние годы значительная часть структур порошковых веществ определяется с помощью вычислительно-стохастических методов глобальной оптимизации в прямом пространстве (Монте-Карло, имитации отжига, генетические алгоритмы), что обусловлено быстрым увеличением производительности компьютеров. Основным из них служит метод имитации отжига, реализованный в ряде известных программ (FOX, DASH, TOPAS, EXPO). В меньшей степени применяют генетические алгоритмы (программы EAGER, MAUD, GEST). Генетические алгоритмы имитируют эволюционные процессы естественного биологического отбора и включают формализованные операции парного скрещивания, мутаций и селекции над множеством случайно сгенерированных структурных моделей. Оба метода используют для минимизации целевые функции, основанные, как и в методе Ритвельда, на разности между расчетной и экспериментальной дифрактограммами. Проблемой этих методов является ухудшение сходимости при увеличении сложности определяемых структур, связанное с нелинейным повышением вероятности стагнации в локальных минимумах гиперповерхности R-фактора. Известны попытки улучшения сходимости за счет параллельных вычислений на суперкомпьютерных кластерах (варианты программ FOXgrid, DASHgrid). Преимуществом генетических алгоритмов является то, что их конструкция является наиболее подходящей для реализации на суперкомпьютерных кластерах и дает потенциальную возможность использования для структурного анализа всей мощи параллельных вычислений. В целом же вопрос, может ли рациональное применение параллельных вычислений обеспечить синергический эффект и кардинально повысить эффективность структурного анализа в прямом пространстве, остается практически неизученным.

Отмеченные проблемы препятствуют более широкому распространению порошкового дифракционного анализа в прямом пространстве. Данная работа направлена на развитие подхода для их решения, основанного на разработке новых автоматизированных методов математической обработки экспериментальных дифракционных данных, интегрирующих полнопрофильный анализ по Ритвельду, генетические алгоритмы и параллельные вычисления.

Работа выполнялась при поддержке грантов ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009-2013 гг. №П2430 и №П2263 по направлению «Физические методы исследования химических соединений»; гранта ГФ-3 в рамках государственного задания министерства образования и науки РФ Сибирскому федеральному университету на 2014 г.

Цель и задачи.

Целью работы является разработка и исследование эффективности автоматизированных методов полнопрофильного рентгеноструктурного анализа для определения структуры в прямом пространстве, основанных на интеграции метода Ритвельда, генетических алгоритмов и параллельных вычислений на базе многоядерных ПК и суперкомпьютерных кластеров.

Для реализации этой цели потребовалось решение следующих задач.

1. Разработка мультипопуляционного параллельного генетического алгоритма и эволюционного метода МПГА определения кристаллической структуры в прямом пространстве, включающего случайную генерацию множеств (популяций) структурных моделей и их циклическую эволюцию путем парных скрещиваний, случайных мутаций, селекции по R-фактору, обменов между популяциями и регулярного полнопрофильного уточнения лучших моделей.

2. Разработка программного графического интерфейса для исследований по автоматизации и интеграции с МПГА полнопрофильного анализа поликристаллических материалов, реализованного в двух вариантах: на базе оптимизированного кода метода Ритвельда и на базе консольной программы метода минимизации производной разности (МПР).

3. Разработка математического обеспечения, реализующего новый метод МПГА на многоядерных ПК и суперкомпьютерном кластере и обеспечивающего параллельную коэволюцию различных популяций структурных моделей на разных вычислительных узлах.

4. Исследование сходимости МПГА и эффекта от использования параллельных вычислений при поиске известных (тестовых) кристаллических структур разной сложности.

5. Сравнение эффективности поиска тестовых структур по МПГА и по некоторым известным современным программам структурного анализа в прямом пространстве.

6. Апробация МПГА при определении ранее неизвестных кристаллических структур.

Научная новизна заключается в следующем.

1. Предложен автоматизированный комплексный метод поиска и полнопрофильного уточнения моделей кристаллической структуры веществ в прямом пространстве по данным порошковой дифракции, основанный на мультипопуляционном параллельном генетическом алгоритме. Разработано его математическое обеспечение для многоядерных ПК и суперкомпьютерных кластеров. Экспериментально установлено, что увеличение количества параллельных эволюционных вычислительных процессов значительно увеличивает вероятность сходимости поиска и позволяет определять более сложные структуры.

2. С помощью мультипопуляционного параллельного генетического алгоритма определены новые кристаллические структуры комплексных соединений $[\text{Pt}(\text{NH}_3)_5\text{Cl}]\text{Br}_3$ и $[\text{Pd}(\text{CH}_3\text{NH}_2)_4][\text{PdBr}_4]$.

Практическая ценность результатов работы:

1. Программы мультипопуляционного параллельного генетического алгоритма для определения кристаллических структур неорганических соединений по данным порошковой дифракции на многоядерных ПК и суперкомпьютерных кластерах внедрены в лаборатории рентгеновских методов ЦКП СФУ.

2. Программный графический интерфейс для автоматизированного полнопрофильного анализа многофазных поликристаллических материалов с интегрированной в него программой метода минимизации производной разности выставлен для общего пользования на сайте разработчика данного метода (<http://sites.google.com/site/ddmsuite>). Программа структурного анализа МПГА выставлена для общего пользования на сайте ЦКП СФУ.

Автор выносит на защиту:

1. Мультипопуляционный параллельный генетический алгоритм и математическое обеспечение МПГА для поиска и полнопрофильного уточнения моделей кристаллической структуры веществ в прямом пространстве по данным порошковой дифракции.

2. Экспериментальные результаты исследования сходимости МПГА на известных тестовых кристаллических структурах разной сложности и оценки повышения эффективности структурного анализа за счет параллельных вычислений на многоядерных ПК и суперкомпьютерном кластере СФУ.

3. Результаты определения с помощью МПГА неизвестных структур комплексных соединений $[\text{Pt}(\text{NH}_3)_5\text{Cl}]\text{Br}_3$ и $[\text{Pd}(\text{CH}_3\text{NH}_2)_4][\text{PdBr}_4]$.

Публикации. По теме диссертации имеется 24 публикации (13 – в Российских журналах, рекомендованных ВАК, и реферируемых иностранных журналах, и 11 – в сборниках научных трудов и тезисов докладов на российских и международных конференциях).

Апробация результатов работы. Основные результаты работы доложены на международных конгрессах «Цветные металлы», Красноярск, 2009 и 2013 г.г.; на школе-конференции «Дифракционные методы исследования вещества: от молекул к кристаллам и наноматериалам», Черноголовка, 2010 г.; на ЕСМ-26, Darmstadt, Germany, 2010 г.; на VI Национальной кристаллохимической конференции, Суздаль, 2011 г.; на IX конференции «Аналитика Сибири и Дальнего Востока», Красноярск, 2012 г.; на конференции EMRS 2014 Fall Meeting, Варшава, 2014 г; на конференции EPDIC-2016, Бари (Италия), 2016 г.

Глава 1. Обзор методов порошковой дифракции

1.1. Общая характеристика современного состояния методов ПД

Знание кристаллической структуры веществ (т.е. геометрического расположения атомов в ячейке, коэффициентов заполнения позиций и тепловых колебаний атомов, а также производных характеристик - таких как межатомные расстояния и т.п.) является предпосылкой к знанию и предсказанию свойств материалов. Наиболее часто используемая техника для определения кристаллических структур – исследование монокристаллов (или отдельных больших кристаллитов). Структурный анализ методами порошкового дифракционного анализа необходим в тех случаях, когда получение вещества в форме монокристалла, подходящего для анализа развитыми монокристалльными дифракционными методами, затруднительно. Поскольку большинство веществ и, тем более, многофазных материалов находится в поликристаллической форме, потенциальная область применения методов порошковой дифракции исключительно обширна, а их развитие - актуально.

Структура поликристаллических материалов характеризуется не только усредненной периодической геометрической структурой расположения атомов в идеализированной кристаллической ячейке, но и микроструктурой, обусловленной несовершенствами, дислокациями и всеми видами беспорядка, которые часто ответственны за интересные свойства материалов при изучении. Исследование макроскопических напряжений в компонентах, оставшихся после обработки и, в зависимости от ситуации, под нагрузкой, проводится при помощи порошковой дифракции; например, определение текстуры в поликристаллических образцах. На сегодняшний день порошковая дифракция представляет собой решающий инструмент в установлении характеристик материалов, и с развитием инструментария, методов, анализа данных и моделирования, становится все более мощным и количественным. Порошковые дифрактограммы объединяют все богатство информации об истинной кристаллической структуре.

Хотя порошковые методы были разработаны ещё в 1916 г. Дебаем и Шерером, их использование было, в основном, ограничено качественным и полуколичественным фазовым анализом, и измерением макроскопических напряжений. Основная причина этого может быть в том, что принципиальная проблема порошковой дифракции: случайное и систематическое перекрытие пиков, обусловленное проекцией трехмерного обратного пространства на одномерную ось 2θ (θ – брэгговский угол дифракции), приводит к сильному уменьшению информации, целиком содержащейся в наборе данных от отдельного кристалла. Тем не менее, не смотря на потери угловой информации, в одномерный набор данных передается достаточная информация для реконструкции 3D структуры. В целом, количественный анализ дифрактограмм с использованием

современных порошковых дифрактометров, компьютеров и программ приносит богатую суммарную информацию о структуре образца, что иллюстрируется на рисунке 1.



Рисунок 1 – Общее информационное содержание порошковой дифрактограммы

Современный инструментарий позволяет получать данные беспрецедентного качества, а развитие современных методов анализа все увеличивают нашу возможность вычлнять полезную информацию из порошковых данных, необходимую для исследования поликристаллических материалов в разнообразных и важных приложениях. Здесь представлены основные сведения о современном состоянии методов порошкового анализа для исследования кристаллической структуры, имея ввиду, что базовое введение в кристаллографию и дифракцию доступно в учебниках и монографиях, таких, как «Основы порошковой дифракции» Pecharsky and Zavalij [1].

Главным прорывом в порошковых методах, как в количественных инструментах, было развитие метода Ритвельда в 1969 г. – метода для уточнения кристаллической структуры, который впервые использовал полный профиль дифрактограммы вместо избирательного анализа индивидуальных, неперекрываемых Брэгговских рефлексов. Этот подход минимизировал влияние перекрытых и вырожденных пиков посредством расчета полного профиля дифрактограммы кристаллической модели, включая различные зависимости параметров пиков от условий проведения эксперимента и состояния образца. Параметры модели, такие, как атомные позиции, параметры ячейки и экспериментальные факторы, которые влияют на форму пика и фон, изменя-

ются посредством использования метода наименьших квадратов (МНК), пока совпадение между рассчитанным и измеренным дифракционным профилем не оптимизируется. Это уточняющий метод: при хороших начальных приближениях, или непосредственном знании, модель уточняется итерационно, малыми поправками. Подход зарекомендовал себя как чрезвычайно успешный, судя по быстро возрастающему количеству уточнений структур по Ритвельду, представляемых в литературе. Этот метод был быстро воспринят из обработки нейтронограмм, обладающих хорошей гауссовой формой линий и отсутствием атомного форм-фактора, в область рентгеновской порошковой дифракции, синхротронной порошковой дифракции и в данные времяпролетной нейтронографии. На данной стадии раскрыты источники богатого информационного содержания формы линий, дающие средства для современного моделирования и анализа формы профиля всей дифрактограммы. Форма профиля рассчитывается из первооснов, учитывающих состояние образца, например, распределение частиц по размерам, неоднородные напряжения и текстуры, в качестве экспериментальных установок и аберраций. Положительный обратный эффект в том, что в результате лучшего описания профиля и получения более точных интенсивностей, уточнение структуры становится более детальным.

Подобно истории с линией профиля, в данное время начаты исследования по раскрытию источников информационного содержания фона. Аккуратная физически обоснованная коррекция предпочтительнее вычитания параметрической аппроксимации фона, как это делалось в общепринятом уточнении по Ритвельду, так как она может быть проведена с учетом таких экспериментальных эффектов, как флуоресценция, комптоновское и диффузное рассеяние от образца. Скорректированный фон под и между брегговскими рефлексами дает богатую информацию о рассеянии на образце, в котором зашифрованы данные о локальной структуре, и о том, как она отличается от «средней» структуры образца на предмет дефектов и связанных колебаний решетки (фононов). Метод общего рассеяния только сейчас был полностью признан для количественного анализа в обратном пространстве – с использованием функции парного распределения атомов (PDF метод), и в прямом пространстве – с использованием метода моделирования отжига Монте-Карло.

Вначале по Ритвельду исследовались простые структуры. Сегодня, с высоким качеством данных, полученных от синхротронных источников, вкупе с превосходными алгоритмами прямых методов или методов глобальной оптимизации в прямом пространстве, определение достаточно сложной кристаллической структуры из данных порошковой дифракции становится обычным методом почти во всех областях естественных наук и инженерии. Этот уровень успеха зависит, в основном, от трех параметров: высокого качества измерительной аппаратуры, качества описания профиля дифрактограммы и алгоритма определения структуры. Становится очевидным, что использование высокомонокроматического параллельного пучка синхротрон-

ного излучения должно сильно улучшить точность в определении атомных параметров. В некоторых случаях такие детали, как вращательное разупорядочение могут быть получены из данных порошковой дифракции, если метод максимальной энтропии объединить с синхротронными данными высокого разрешения.

Измерения в порошковой дифракции стремятся к тому, чтобы быть информативными и быстрыми, особенно при параллельном сборе данных, например, использование 1D и 2D детекторов. Параллельный сбор данных делает возможным исследования в условиях, отличных от окружающих, где такие параметры, как температура, давление, электрическое и магнитное поле или состав газовой фазы, изменяются и состояние образца (включая фазовый состав, структуру, локальную структуру, распределение частиц по размерам и напряжения, и т.д.) количественно отслеживается вместе с изменением условий. Новые двумерные детекторы, особенно, в комбинации с синхротронным излучением позволяют проводить определение и уточнение структур во времена порядка времени химической реакции.

Развитие высокочувствительных скоростных детекторов открывает новые области для порошковых методов, в частности, для фазового и структурного анализа многофазных материалов. Например, количественная регистрация и анализ малых содержаний полиморфных фаз представляет большой интерес для фармацевтических исследований и отдельных областей бизнеса. Перспективы порошковой дифракции превосходны. Простота и широкая применяемость структурного определения и уточнения с использованием порошков начинают составлять конкуренцию некоторым традиционным областям анализа монокристаллов. Рисунок 2 иллюстрирует связь между анализом отдельных монокристаллов и порошковой дифракцией в области решения структур.

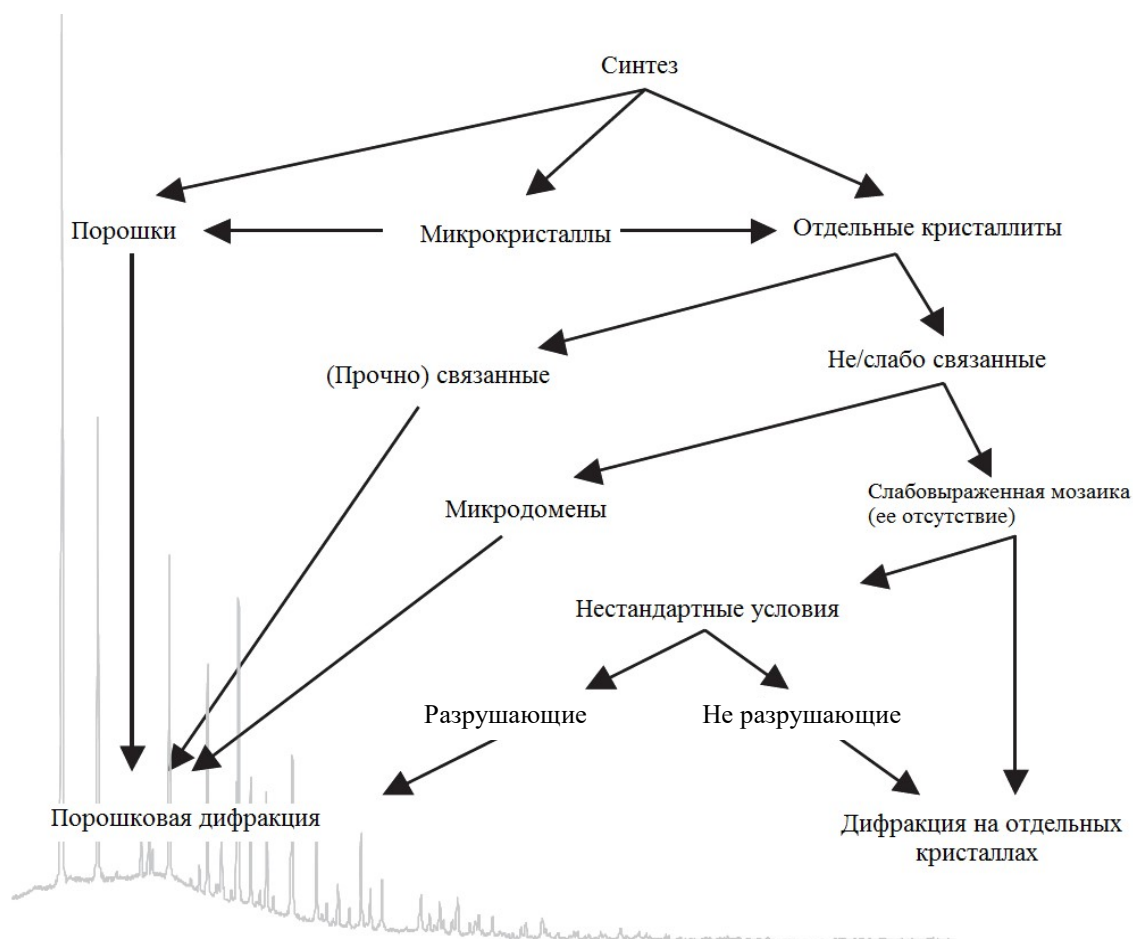


Рисунок 2 – Иллюстрация связи между анализом отдельных кристаллов и порошковой дифракцией в области решения структур

Эти две техники хорошо дополняют друг друга, обладая при этом собственными достоинствами и недостатками, областями применения. Однако, области, где порошковая дифракция имеет большое влияние, несомненно развиваются и становятся разнообразнее.

1.2. Этапы определения кристаллической структуры

Рассмотрим основные этапы определения кристаллической структуры из данных порошковой дифракции. На рисунке 3 представлена обобщенная схема определения кристаллической структуры (заимствованная из monograph 13 of IUCr [2]). После установления формулы соединения и регистрации дифрактограммы структурный анализ начинается с определения (индексирование) и уточнения параметров решетки, и далее, пространственной группы симметрии. Данная группа методов достаточно хорошо развита, за исключением случая многофазных материалов, когда часть дифракционных рефлексов новой фазы перекрыта рефлексами от других фаз. Если нет возможности экстракции новой фазы в чистом виде, то необходима тщательная фазовая идентификация линий примесных фаз по многофазной дифрактограмме. Для индексирования и уточнения параметров решетки существует ряд эффективных программ.

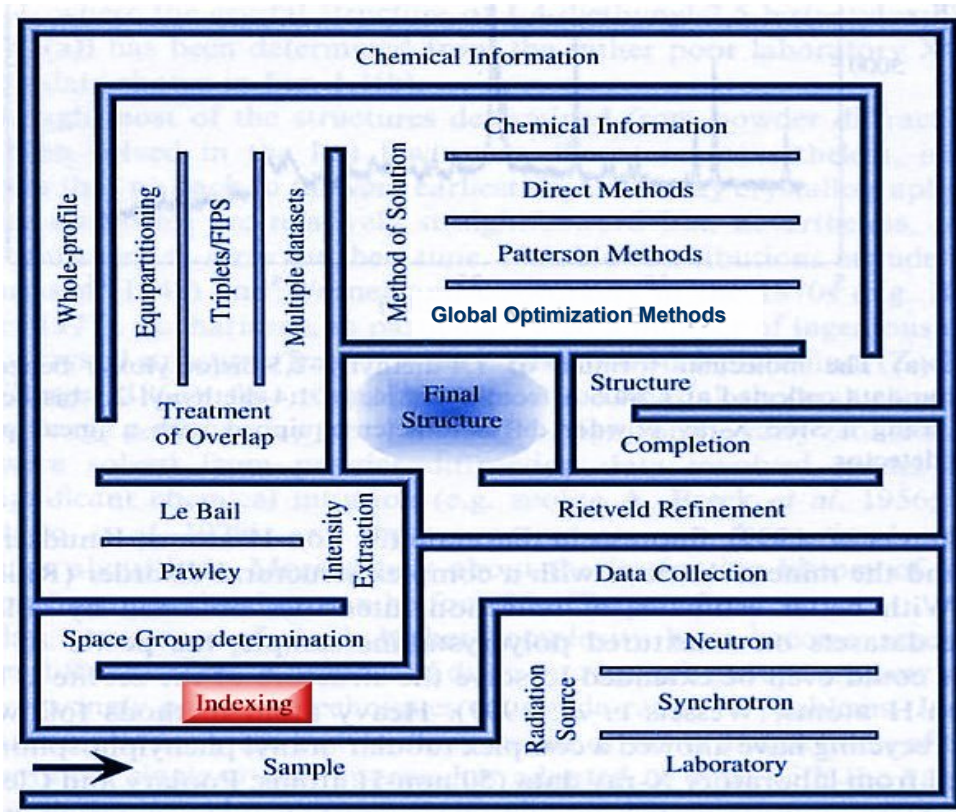


Рисунок 3 – Обобщенная схема определения кристаллической структуры из данных порошковой дифракции

Далее, выполняется полнопрофильная декомпозиция интенсивности перекрытых рефлексов дифрактограммы с использованием методов Le Bail или Pawley. Следующим этапом структурного анализа является поиск модели неизвестной кристаллической структуры. Здесь существует ряд методов для поиска моделей в обратном (прямые и паттерсоновские методы) или в прямом пространствах (методы глобальной оптимизации). Третьим этапом структурного анализа является уточнение модели кристаллической структуры. Основным методом для этой цели служит вышеупомянутый метод полнопрофильного анализа по Ритвельду. Современное развитие метода Ритвельда позволило применить его также для нестандартного количественного фазового анализа и определения величин кристаллитов, микродеформаций решетки и текстур. Основными проблемами метода являются *проблема корректного отделения фона* дифрактограммы при наличии аморфной составляющей и *проблема хороших исходных приближений* профильных и структурных параметров, без чего метод не сходится в глобальный минимум, соответствующий структурному решению.

1.3. Межлабораторные испытания по определению кристаллической структуры

По данным А. Le Bail [3], в последние несколько лет с помощью указанных методов ежегодно, во всем мире производится определение около двухсот кристаллических структур новых

веществ. На практике сейчас принимаются за решение структур с 40-50 степенями свободы [4]. Об эффективности современных методов определения кристаллической структуры из данных порошковой дифракции можно судить по результатам регулярно проводимых межлабораторных испытаний по определению кристаллической структуры (Structure Determination by Powder Diffractometry Round Robin - SDPDRR). Три таких испытания проводились в 1998, 2002 и 2008 г.г. группой Le Bail под эгидой комиссии по порошковым данным международного союза кристаллографов (CPD IUCr). Каждый раз в Интернет на срок до трех месяцев выставлялись дифрактограммы двух соединений с неизвестной кристаллической структурой, которые участники должны были определить. Сообщалась, также, вероятная химическая формула и данные индексирования. Целью этих межлабораторных испытаний являлась оценка эффективности методов и программного обеспечения порошкового структурного анализа. Подразумевалось, что наиболее эффективный инструментарий должен позволять квалифицированному специалисту определение одной структуры примерно за день. В таблице 1 представлено сравнение основных результатов трех SDPD Round Robin [3].

Таблица 1 - Сравнение результатов SDPD Round Robins

Supposing the space group and structure known, minimal number of degrees of freedom at the structure solution (sDoF) stage from X-ray data for various methods, and total number of refinable (rDoF) atomic coordinates (non-H atoms).

	SDPDRR-1		SDPDRR-2		SDPDRR-3	
	sample 1	sample 2	sample 1	sample 2	sample 1	sample 2
Chemical formula	[Co(NH ₃) ₅ CO ₃]NO ₃ ·H ₂ O	C ₂₂ H ₂₄ N ₂ O ₈ HCl	Al ₂ F ₁₀ [C ₆ N ₄ H ₂₀]	Sr ₅ V ₃ (F/O/OH/H ₂ O) ₂₂	CaC ₄ H ₄ O ₆ ·4H ₂ O	La ₁₈ W ₁₀ O ₅₇
Space Group	P2 ₁	P2 ₁ 2 ₁ 2 ₁	P2/c	P2 ₁ /c	P-1	P-62c
Z	2	4	2	4	2	2
Total independent non-H	15	33	12	30	15	20
Sites to find by DM/Patt	~10	~20	~10	~15	~10-15	6W+4La
sDoF DS : mol+atoms	6x3+3=21	6+3=9	6x3 = 18	6x8=48	6+5x3=21	10x3=30
rDoF	44	99	30	90	45	43
Data quality	low	medium	low	high	low	high

DM : Direct Methods, Patt : Patterson method, DS : Direct Space method, Ref : SDPDRR-1, samples 1 (Zhu et al., 1999), 2 (Clegg & Teat 2000); SDPDRR-2 samples 1 (Adil et al., 2007), 2 (Le Bail, 2003); SDPDRR-3 sample 1 (Le Bail et al., 2008), 2 (Chambrier et al., 2008)

В таблице 1 приведены анализируемые кристаллические соединения и их пространственные группы. В строке «Total...» указано количество не водородных атомов в независимой части ячейки. В строке sDoF указано минимальное число степеней свободы при поиске по порошковым данным структурной модели: координат атомов и целых молекул, представляющих априори известные структурные фрагменты, позиции которых задаются 3-мя координатами, и, иногда, торсионными углами. Например, при поиске модели структуры соединения CaC₄H₄O₆·4H₂O (тартрата кальция) sDoF=5x3+6=21 означает, что независимо определены только координаты атомов кальция и кислорода воды (5x3), и координаты, и торсионные углы

(6) группы тартрата $C_4H_4O_6^{2-}$. В строке rDoF указано общее число степеней свободы уточненных координат не водородных атомов модели (т.к. водородные атомы плохо поддаются определению порошковыми методами). Т.е. при уточнении найденной структуры тартрата кальция уточняются уже rDoF=45 координат всех не водородных атомов. Показательно то, что во всех межлабораторных испытаниях правильные результаты представляло лишь небольшое число участников. Например, в Round Robin SDPD–3 2008г. результаты были получены всего от 9-ти участников из 170-ти скачавших дифрактограммы.

1.4. Определение параметров решетки и пространственной группы

1.4.1. Процедура индирования дифрактограмм

Основная цель процесса индирования порошковой дифракционной картины – геометрическое восстановление трехмерного обратного пространства из одномерного распределения наблюдаемых значений межплоскостных расстояний d . Наименование «индирование» относится к тому факту, что определение элементарной ячейки эквивалентно присвоению соответствующих троек индексов Миллера каждому наблюдаемому межплоскостному расстоянию [5].

Индирование порошковой дифракционной картины является первым шагом в *ab initio* процессе определения структуры. Первый важный опытный подход к индированию был предложен Рунге [6]. Несмотря на его долгую историю и последние большие достижения в сферах экспериментального оборудования, математических методов и скорости вычислений, индирование порошковой дифракционной картины по-прежнему остается сложной задачей. Базовое уравнение индирования, связывающее параметры обратной ячейки и индексы, является следующей квадратичной формой:

$$Q_{hkl} = h^2 A_{11} + k^2 A_{22} + l^2 A_{33} + hk A_{12} + hl A_{13} + kl A_{23}, \quad (1)$$

где

$$Q_{hkl} = \frac{10^4}{d_{hkl}^2}, \quad d_{hkl} = \frac{\lambda}{2 \sin \theta_{hkl}},$$

$$A_{11} = 10^4 a^{*2}, \quad A_{22} = 10^4 b^{*2}, \quad A_{33} = 10^4 c^{*2}, \quad A_{12} = 10^4 \cdot 2a^* b^* \cos \gamma^*,$$

$$A_{13} = 10^4 \cdot 2a^* c^* \cos \beta^*, \quad A_{23} = 10^4 \cdot 2b^* c^* \cos \alpha^*.$$

Корректная элементарная ячейка может быть идентифицирована n межплоскостными расстояниями, где n зависит от симметрии решетки. Минимальными значениями n являются: $n = 1$ для кубической системы, $n = 2$ для тетрагональных и гексагональных кристаллов, $n = 3, 4, 6$ для ромбической, моноклинной и триклинной систем, соответственно.

Как только параметры обратной ячейки оказываются вычисленными, легко определить прямую ячейку. Качество результатов индирования зависит от качества $\{Q_{hkl}\}$ и, следовательно, от точности определения положений пиков.

Особое внимание требуется, когда имеются численные соотношения между осями ячейки (геометрические неоднозначности, см. Раздел 1.3). В дополнение ко всему, наличие посторонних фаз может сделать процесс индирования еще более сложным. Ошибки в Q_{hkl} модифицируют задачу к нахождению $\{A_{ij}\}$, удовлетворяющего следующему отношению:

$$Q_{hkl} - \Delta < h^2 A_{11} + k^2 A_{22} + l^2 A_{33} + hkA_{12} + hlA_{13} + klA_{23} < Q_{hkl} + \Delta, \quad (2)$$

где Δ есть подходящий параметр допуска.

Критерии достоверности (Figures of Merit).

Вне зависимости от метода индирования, в качестве возможной может быть предложено большое количество правдоподобных ячеек. Таким образом, полезно применять критерий достоверности (Figure of Merit, *FOM*) для определения наиболее вероятных из них, а также оценки их качества. Наиболее распространенными *FoM* являются M_{20} и F_N , предложенные де Вольфом [7], Смитом и Снайдером [8], соответственно. M_{20} определяется как:

$$M_{20} = \frac{Q_{20}}{2\langle \varepsilon \rangle N_{20}},$$

где Q_{20} есть значение Q для случая двадцати наблюдаемых пиков, $\langle \varepsilon \rangle$ есть среднее отклонение между наблюдаемыми и вычисленными значениями Q для 20 индированных пиков, N_{20} есть число вычисленных рефлексов до значения d , соответствующего Q_{20} . M_{20} зависит от (а) различия между наблюдаемыми и вычисленными линиями (через $\langle \varepsilon \rangle$); (б) объема элементарной ячейки (через N_{20}). Чем меньше среднее отклонение $\langle \varepsilon \rangle$ и размер ячейки, тем больше M_{20} и выше уверенность в предложенной элементарной ячейке. Не существует порогового значения M_{20} , которое бы гарантировало правильность решения. Де Вольф предложил, что если число неиндированных пиков среди первых двадцати линий не превышает 2 и $M_{20} > 10$, результаты индирования должны оказаться, в основном, верными [7]. В дополнение к этому, если лучшее решение не найдено из-за наличия примесных линий, желательно провести проверку ячеек, имеющих несколько неиндированных линий и более высокий критерий достоверности де Вольфа [8].

Критерий F_N определяется следующим образом:

$$F_N = \frac{1}{\langle |\Delta 2\theta| \rangle} \cdot \frac{N}{N_{\text{poss}}}$$

где $\langle |\Delta 2\theta| \rangle$ есть среднее абсолютное расхождение между наблюдаемыми и вычисленными значениями 2θ , а N_{poss} есть число возможных дифракционных линий до N -ой наблюдаемой. Обычно, значение F_N связывается с парой ($\langle |\Delta 2\theta| \rangle$, N_{poss}).

Смит и Снайдер [8] сравнили характеристики M_{20} и F_N , проанализировав набор соединений, относящихся к триклинной, ромбической и кубической системам. Они подчеркнули превосходство F_N по отношению к M_{20} , поскольку последний (а) определяется строго для 20 линий,

(b) сильно зависит от кристаллографической системы и пространственной группы. Но Вернер [2, с. 118] отметил, что увеличение значения M_{20} с симметрией не является недостатком, поскольку индцирование порошковой дифракционной картины для кубической системы более вероятно, нежели для триклинной.

Геометрические неоднозначности.

Порошковая дифракционная картина может быть индцирована различными решетками [7, 9, 10]. Систематические неоднозначности могут проявляться, когда «две или более различные решетки, характеризующие различные приведенные формы, могут давать порошковую дифракционную картину с идентичными положениями 2θ » [11]. Если наблюдаются геометрические неоднозначности, требуется дополнительная предварительная информация (например, монокристалльное исследование) для выбора одной из двух возможных решеток.

1.4.2. Программы индцирования

Наиболее широко используемыми программами индцирования являются ИТО [10], TREOR [12] и DICVOL91 [13]. Все три классические программы индцирования представлены в программном пакете Crysfire [14]. Их подходы к проблеме индцирования различны и будут вкратце описаны.

ИТО основан на методе, изначально предложенным Рунге [6] и развитым Ито [15, 16] и де Вольфом [17, 18]. Кристаллографическая зона в обратном пространстве (т.е. плоскость, проходящая через начало координат) может быть описана через точку начала координат и две точки внутри кристаллической решетки. Если Q' и Q'' есть возведенные в квадрат расстояния к двум точкам из начала координат, значение Q любой точки решетки может быть определено как:

$$Q_{mn} = m^2 Q' + n^2 Q'' + mnR, \quad (3)$$

где m и n целые, $R = 2\sqrt{Q'Q''} \cos \varphi$, а φ есть угол между радиус-векторами двух точек.

Тогда:

$$R = (Q_{mn} - m^2 Q' - n^2 Q'')/mn. \quad (4)$$

В соответствии с алгоритмом, все наблюдаемые значения Q_{hkl} и несколько положительных целых m и n подставляются в уравнение (4): получается и сохраняется большое число значений, некоторые из которых равны в пределах приемлемых ошибок. Из них легко определяются углы между базовыми векторами зоны. Процедура повторяется для различных зон (т.е. различных пар Q' и Q''). Наиболее важными зонами будут те, для которых значение R находится много раз. Как только был выполнен поиск зон, следующими шагами ИТО являются: (а) найти пары зон, имеющих общую линию, (б) определить угол между двумя зонами для описания решетки. Метод был обобщен Виссером [10], который подробно рассматривает симметрию обратного про-

странства. Программа ITO очень эффективна для индирования низкосимметричных дифракционных картин.

TREOR стартует с анализа кубической симметрии и шаг за шагом производит тестирование для более низкосимметричных кристаллографических систем. Для каждой исследуемой системы TREOR выбирает «базовые линии»: этот термин используется для обозначения дифракционных линий (обычно относящихся к малоугловой области по 2θ), которым опытным путем (методами проб и ошибок) присваиваются индексы для поиска элементарной ячейки. Пяти наборов базисных линий обычно достаточно для тестирования ромбической системы, тогда как семи наборов может оказаться недостаточно для тестирования моноклинной из-за наличия доминантной зоны (в этом случае более чем пять базисных линий могут быть индированы с общим нулевым индексом). Для определения наличия доминантных зон для моноклинной симметрии выполняется специальный *short-axis* тест [12].

Если \mathbf{M} есть квадратная матрица индексов Миллера, \mathbf{A} есть вектор неизвестных $\{A_{ij}\}$, а \mathbf{L} есть вектор $\{Q_{hkl}\}$ для базисных линий, каждый пробный набор $\{A_{ij}\}$ в уравнении (1) получается решением системы линейных уравнений $\mathbf{MA} = \mathbf{L}$. Корректность пробных параметров $\{A_{ij}\}$ зависит от точности определения низкоугловых линий. Тестирование нескольких различных комбинаций базисных линий может дать возможность найти верное решение даже в случае ошибки в одной или более базисной линии. Правдоподобные решения находятся, если индированная дифракционная картина характеризуется $M_{20} > 10$ с не более чем одной неиндированной линией среди первых двадцати.

Успех TREOR связан со стандартным набором параметров, предложенным авторами [12] на основании накопленного опыта: в случае неудачи пользователь может легко изменить эти параметры через соответствующие ключевые слова во входном файле. Этот метод был классифицирован Ширли [19, с. 361-382] как полуисчерпывающий, так как производится разумное сокращение, чтобы «ограничить размер поля решений для увеличения скорости».

DICVOL91 есть подход индирования, основанный на *методе дихотомии*, введенном Д. Луйе и М. Луйе [20]. Ширли [19, с. 361-382] определил этот метод как «вероятно, оптимальную исчерпывающую стратегию для пространства параметров». Исходная версия программы была написана для ромбической и более высоких сингоний. Позднее, метод был расширен на моноклинную [21] и триклинную симметрии [13]. Метод дихотомии основывается на варьировании прямого пространства, конечным приращением длин граней ячейки и межосевых углов (m -размерный поиск, где m есть число неизвестных параметров элементарной ячейки): вариации уменьшаются, когда они включают возможное решение. DICVOL91 крайне чувствителен к качеству данных.

Развитие программ индицирования.

В более поздние годы проблема индицирования была заново пересмотрена [22]. Некоторые из классических программ индицирования, TREOR и DICVOL91 развились в более мощные и эффективные версии: N-TREOR09 [23] и DICVOL06 [24], соответственно. В дополнение, были разработаны полностью новые программы, вдохновленные доступностью увеличенной скорости вычислений. Среди них отметим: основанную на генетическом алгоритме программу, разработанную Кариуки и др. [25], SVD-Index [26], X-Cell [27] и McMaille [28]. Все вышеупомянутые программы рассматриваются здесь кратко. Для более подробной информации можно обратиться к исчерпывающим работам Ширли [5] и Бергмана и др. [22].

N-TREOR09 - усовершенствованная версия TREOR. Тем не менее, программа сохранила в неизменности основную стратегию, новые изменения были введены для того, чтобы сделать ее еще более действенной и исчерпывающей. Отметим:

1. если программа не находит удовлетворительных решений с параметрами по умолчанию, она автоматически повторяет поиск с расширенными пределами допуска.
2. введено автоматическое увеличение параметров и объема искомой ячейки при наличии $d > 10 \text{ \AA}$.
3. введена зависимость пределов погрешности от длины волны используемого излучения.
4. введен автоматический поиск решений с более высокой симметрией.
5. введен модифицированный критерий достоверности де Вольфа [6], учитывающий степень наложения пиков в образце.
6. введена автоматическая коррекция сдвига нуля рентгенограммы.
7. улучшена процедура индицирования триклинной системы.

DICVOL06. Среди новых удобных усовершенствований, отметим:

1. уточнение нулевой точки порошковых дифракционных данных.
2. возможность учета наличия примесных линий.
3. автоматический поиск приведенной ячейки для моноклинной и триклинной кристаллографических систем.
4. новые опции для выбора стратегии поиска – для работы с нестабильными случаями (доминантные зоны, случаи с большими сторонами ячейки, и т.д.)

Методы в прямом пространстве

Индицирование с помощью генетического алгоритма (ГА). ГА есть методика оптимизации, использующая идею дарвиновской эволюции: наиболее приспособленный член популяции выживает и порождает улучшенное поколение, случайные мутации позволяют системе эволюционировать [29, 30]. Методы для индицирования порошковых дифракционных данных были предложены Тэмом и Комптоном [31]; Пашковицем [32]. Их подходы, подобно традиционным

методам индирования, используют геометрию дифракции. Кариуки и др. [25] применяют ГА, используя также полнопрофильную подгонку (whole-profile fitting). Целью является нахождение параметров решетки $\{a, b, c, \alpha, \beta, \gamma\}$, обладающих наилучшим соответствием с экспериментальной порошковой дифракционной картиной, т.е. согласование в соответствие с глобальным минимумом гиперповерхности $R_{wp}\{a, b, c, \alpha, \beta, \gamma\}$, где:

$$R_{wp} = \sum_{regions} \left[\frac{\sum_i w_i (y_i - y_{ci})^2}{\sum_i w_i y_i^2} \right]^{\frac{1}{2}},$$

$\sum_{regions}$ обозначает сумму по всем областям 2θ , на которые пользователем была разделена дифракционная картина, i пробегает по всем точкам в каждой области, y_i и y_{ci} есть наблюдаемые и вычисленные дифракционные картины на i -ом шаге эксперимента. Интенсивности извлекаются по алгоритму Ле Бэйла [33]. Процедура более затратна по времени, но менее чувствительна к небольшой примеси (глобальный минимум будет достигнут, когда верно индирована основная фаза).

Индирование с помощью метода Монте-Карло – McMaille. Эта индированная программа использует информацию о полном порошковом профиле, как уже было предложено Кариуки и др. [24], а также методы Монте-Карло для случайного генерирования параметров ячейки. Как только параметры были сгенерированы, вычисляются индексы Миллера и положения пиков, качество ячейки оценивается подходящим критерием достоверности. Использование интенсивностей делает программу относительно нечувствительной к примесям, если их вклад в общую интенсивность не превышает 10–15%. Допуск отклонения нулевой точки составляет $|0.05^\circ|$

Проблема индирования обычно решается за несколько минут, если: (а) симметрия не ниже моноклинной; (б) объем ячейки не превышает 2000 \AA^3 ; (с) параметры ячейки меньше 20 \AA . Больше вычислительное время необходимо для триклинной симметрии: наибольший недостаток программы состоит во времязатратности в случае низкосимметричных кристаллов.

SVD-Index. Это коммерческая программа для индирования [26], являющаяся частью программного комплекса TOPAS [34]. Соотношение для обратной решетки в уравнении (1) решается через разложение по сингулярным числам (Singular Value Decomposition) [35]. Этот метод рекомендуется для случаев, когда уравнений больше, нежели переменных [35].

В отличие от N-TREOR09 и TREOR, *SVD-Index* производит процедуру индирования не только для нескольких линий (базисных линий в N-TREOR09 или TREOR), но для всех наблюдаемых линий. Новый подход обещает быть нечувствительным к $2\theta_0$ -ошибкам, наличию примесей и нехватке низкоугловых дифракционных пиков.

X-CELL. Это коммерческая программа, включенная в программный пакет Material Studio [36]. Она выполняет исчерпывающий поиск через последовательный метод дихотомии, подобно программе DICVOL, совмещая его с (а) поиском сдвига нулевой точки дифракционной картины; (b) уровнями допуска примесей (определяемыми пользователем), задающим допустимое число неиндексированных пиков.

Несколько полезных замечаний: Бергманн и др. [22] сравнили производительность классических и новых программ индексирования. При этом они получили, что некоторые из последних программ более устойчивы к ошибкам сдвига нулевой точки, наличию примесных линий, недостаточной точности данных, но их общее заключение (см. также ссылку [2, с. 118]) было таковым, что вероятность успеха индексирования пропорционально возрастает с числом различных примененных для этого программ.

1.4.3. Определение пространственной группы

Анализ дифракционных интенсивностей дает информацию о группе Лауэ и систематически отсутствующих рефлексах, откуда могут быть определены символ погасания и, в случае удачи, пространственная группа. К сожалению, в случае порошков, дифракционная картина не может быть однозначно интерпретирована: перекрытие пиков (систематическое или случайное), вклад фона, возможное наличие предпочтительной ориентации и/или примесных пиков делает вычисление интенсивностей ненадежным. Соответственно, интенсивности рефлексов, относящихся к малым пикам, не могут быть точно извлечены. Аккуратная визуальная проверка экспериментальной картины является наиболее общей практикой для идентификации пространственной группы: зачастую, с данным экспериментальным распределением пиков совместимы многие пространственные группы. Когда имеется выбор, корректный символ погасания может быть определен при помощи одной из следующих стратегий: (а) рассмотрение предварительной информации о молекулярных свойствах изучаемой структуры, которые могут повлиять на выбор пространственной группы; (b) попытка решить структуру для каждой пространственной группы, а затем выбрать наиболее подходящее решение.

В последние годы было разработано несколько новых процедур, стратегий и программ для *ab initio* автоматического решения порошковых структур. Они включают инновационные подходы, избегающие ручной проверки дифракционной картины, и основываются на статистическом анализе интегральной интенсивности рефлексов, для автоматического определения пространственной группы. Было предложено два альтернативных метода: первый – Марквардсеном и др. [37] (реализован в программе DASH [38]), и второй – Альтомаре и др. [39, 40] (реализован в программе EXPO [41]). Оба они: (а) основываются на статистическом анализе интегральных интенсивностей рефлексов, и, следовательно, избегают проверки дифракционной кар-

тины вручную; (b) дают количественную оценку относительных вероятностей для всех групп погасания, совместимых с кристаллографической системой. Предпочтительной будет группа погасания с большей вероятностью: наилучшими кандидатами на верное решение будут пространственные группы, с ней сочетаемые.

Процедура DASH. Она базируется на вычислении общей вероятности $p(E_{gr}|\mathbf{I}^P)$, где E_{gr} обозначает символ погасания, а $\mathbf{I}^P = (I_1, I_2 \dots I_N)$ есть взаимосвязанные интенсивности рефлексов, полученные из уточнения линейным методом наименьших квадратов, где выбраны наиболее общие группы погасания рассматриваемых кристаллографических систем. Из теоремы Байеса:

$$p(E_{gr}|\mathbf{I}^P) = p(E_{gr}) p(\mathbf{I}^P|E_{gr}) / p(\mathbf{I}^P).$$

Начальное вероятностное распределение $p(E_{gr})$ постоянно, так как все символы погасания рассматриваются как равным образом возможные; $p(\mathbf{I}^P)$ также постоянно, так как данные не изменяются при варьировании групп погасания. Соответственно:

$$p(E_{gr}|\mathbf{I}^P) \sim p(\mathbf{I}^P|E_{gr}),$$

где $p(\mathbf{I}^P|E_{gr})$ дает относительную вероятность каждой группы погасания E_{gr} . Она может быть представлена как:

$$p(\mathbf{I}^P|E_{gr}) = \int p(\mathbf{I}|E_{gr}) p(\mathbf{I}^P|\mathbf{I}) d\mathbf{I},$$

где $p(\mathbf{I}^P|\mathbf{I})$ – мультивариантная гауссова функция правдоподобия:

$$p(\mathbf{I}^P|\mathbf{I}) = (2\pi)^{-N/2} |\mathbf{C}|^{-1/2} \exp[-1/2(\mathbf{I}^P - \mathbf{I})^T \mathbf{C}^{-1} (\mathbf{I}^P - \mathbf{I})],$$

а \mathbf{C} есть ковариационная матрица Паули. Все интенсивности рассматриваются как статистически независимые и равномерно распределенные; следовательно, $p(\mathbf{I}|E_{gr}) = \prod p(I_i|E_{gr})$. Если, в соответствии с данной E_{gr} , i -ая интенсивность ожидается отсутствующей, тогда $p(I_i|E_{gr})$ есть дельта-функция, в противном случае она соответствует ацентричному распределению Уилсона [42]. В идеальном случае, когда N пиков на экспериментальной дифракционной картине не перекрываются, матрица \mathbf{C} будет диагональной и расчет $p(\mathbf{I}^P|E_{gr})$ потребует вычисления N одномерных интегралов, которые могут быть решены аналитически. В более реальном случае, когда рефлексы перекрываются, \mathbf{C} может быть аппроксимирована блочной диагональной матрицей: соответствующие интегралы (с размерностью больше единицы), могут быть решены численно методами Монте-Карло.

Процедура EXPO может быть разбита на следующие шаги:

1. Экспериментальная порошковая дифракционная картина раскладывается на отдельные дифракционные интенсивности по методу Ле Бэйла [40] в пространственной группе, имеющей наивысшую симметрию Лауэ и не имеющей условий погасания (например, $P2/m$ для моноклинной, $P2/m2/m2/m$ для ромбической, $P4/mmm$ для тетрагональной, и $Pm\bar{3}m$ для кубической системы).

2. Нормированные интенсивности $z_h = |E_h|^2$ подвергаются статистическому анализу для определения пространственной группы симметрии.
3. Алгоритм дает вероятностное значение для каждого символа погасания, сочетаемого с симметрией решетки, предоставленной процедурой индцирования.

Процедура полностью автоматизирована, имеет графический интерфейс, предоставляющий: (а) графическое окно, где возможные группы погасания ранжируются в соответствие с их вероятностями; (б) список совместимых с каждым данным символом погасания пространственных групп; (с) по запросу пользователя, список рефлексов, соответствующих выбранному символу погасания, которые должны оказаться систематически отсутствующими.

1.5. Классические методы определения кристаллических структур по данным порошковой дифракции

1.5.1. Задача определения кристаллической структуры

Исходными данными для задачи определения кристаллической структуры являются экспериментальная дифрактограмма кристаллического вещества и результаты решения задач определения параметров решетки и пространственной группы по дифрактограмме. Типичные рентгеновские дифракционные эксперименты в задаче определения кристаллической структуры обеспечивают модули структурных факторов, тогда как относительные фазы теряются. Восстановление фазовой информации является ключевым для решения кристаллической структуры и известно в кристаллографии как *фазовая проблема*. В монокристалльной дифракции эта задача решается различными способами:

- a. Используя только дифракционные данные исследуемого соединения, когда недоступна информация о его молекулярной геометрии (*ab initio* методы). К этой категории относятся метод Паттерсона и прямые методы.
- b. Используя дифракционные данные исследуемого соединения, когда доступна информация о его молекулярной геометрии (*методы в прямом пространстве*: частным случаем является *метод молекулярного замещения*).
- c. Используя дифракционные данные исследуемого соединения и дифракционные данные одной или более изоморфных структур (*методы изоморфного замещения*).
- d. Используя эффекты аномальной дисперсии. Один или более (при различных длинах волн) набор данных рассматривается как моделированные изоморфные структуры (методики *SAD-MAD*, *SIRAS-MIRAS*).
- e. Используя экспериментальное сопоставление триплета инвариантных фаз (эффект многоволновой дифракции). Для экспериментального решения фазовой задачи может быть рассмотрен и использован эффект трехволновой дифракции.

Решение кристаллической структуры из порошковых данных не столь прямолинейно из-за следующих проблем: неоднозначность в определении фона, возможное присутствие эффекта преимущественной ориентации и, что главное, свертывание трехмерного обратного пространства в одномерную дифракционную картину, что вызывает существенное перекрытие дифракционных пиков. Как следствие, дифракционные модули рефлексов оцениваются лишь приблизительно. Неопределенность данных о модулях делает некоторые из вышеописанных подходов непригодными в порошковой кристаллографии. Например, случай (е) на сегодня экспериментально неприменим к порошковым образцам, а случаи (с) и (d) полагаются на очень малые различия между модулями структурных факторов, которые очень сложно получить с достаточной точностью из порошковой дифракционной картины [более того, $|F_h|$ и $|F_{-h}|$ перекрываются и невозможно оценить разницу ($|F_h| - |F_{-h}|$)]. Соответственно, в данной главе будут, по большей части, рассмотрены варианты (а) и (b) по данным известных публикаций и особенно [43; 44, с. 227-265].

Несмотря на это, порошковая дифрактометрия играет центральную роль в исследовании и технологии, делая возможным анализ материалов, недоступных в форме монокристаллов приемлемого размера и качества. По этой причине многие недавние усилия были направлены на усовершенствование экспериментальных методик (т.е. использование синхротронного излучения, оптики, генераторов, детекторов, и т.д.), а также разработку новых методов для анализа данных. Например, в последнее десятилетие были предложены альтернативные стратегии решения структуры (методы в прямом пространстве, рассмотренные в следующей главе), которые напрямую используют скорее порошковый дифракционный профиль, а не отдельные интенсивности отражений, таким образом, избегая проблемного процесса извлечения интенсивностей.

1.5.2. Метод Паттерсона

Функция Паттерсона представляет собой свертку электронной плотности $\rho(\mathbf{r})$ с самой собой, инвертированной по отношению к началу отсчета:

$$P(\bar{\mathbf{u}}) = \rho(\bar{\mathbf{r}}) * \rho(-\bar{\mathbf{r}}) = \int_V \rho(\bar{\mathbf{r}})\rho(\bar{\mathbf{r}} + \bar{\mathbf{u}})d\bar{\mathbf{u}}, \quad (5)$$

где * обозначает операцию свертки. Она может быть представлена также как:

$$P(\vec{\mathbf{u}}) = T^{-1} \left[|F(\vec{\mathbf{r}}^*)|^2 \right] = \int_{S^*} |F(\vec{\mathbf{r}}^*)|^2 \exp(-2\pi i \vec{\mathbf{r}}^* \cdot \vec{\mathbf{u}}) d\vec{\mathbf{r}}^*,$$

где оператор T обозначает Фурье-преобразование, а S^* - все обратное пространство.

Так как $|F(\vec{\mathbf{r}}^*)|$ выбирается только в узлах обратной ячейки (где $\mathbf{r}^* = \mathbf{h}$), мы имеем:

$$P(\bar{\mathbf{u}}) = \frac{1}{V} \sum_{\mathbf{hkl}} |F_{\mathbf{hkl}}|^2 \exp(-2\pi i \bar{\mathbf{h}} \cdot \bar{\mathbf{u}}), \quad (6)$$

где V есть объем элементарной ячейки. Из уравнения (6) можно вывести следующие особенности функции Паттерсона:

- a. $P(\mathbf{u}) = P(-\mathbf{u})$, т.е. карта Паттерсона всегда центросимметрична.
- b. Максимумы $P(\mathbf{u})$ соответствуют межатомным векторам.
- c. Каждый пик имеет интенсивность, пропорциональную произведению атомных номеров двух атомов, связанных межатомным вектором \mathbf{u} . Более пары атомов могут иметь один и тот же межатомный вектор \mathbf{u} и потому могут вносить вклад в один и тот же пик Паттерсона.
- d. Элементарная ячейка, содержащая N атомов, генерирует N^2 межатомных векторов $\mathbf{r}_i - \mathbf{r}_j$. N из них будет локализовано в начальной точке, тогда как оставшиеся $N(N-1)$ распределены по ячейке. Это означает, что пики Паттерсона обычно перекрываются (также из-за того факта, что пики Паттерсона шире, нежели пики электронной плотности).
- e. Карты Паттерсона, вычисленные из порошковых данных, менее информативны, нежели полученные из монокристалльных данных в связи с неизбежными ошибками в дифракционных модулях, полученных в процедуре полного разложения дифракционной картины.

Вышеприведенные особенности могут усложнить получение положений атомов из анализа карты Паттерсона. Наиболее характерный подход включает использование так называемых *сечений Харкера*: они содержат межатомные векторы между атомом и его симметричными эквивалентами. Для примера, рассмотрим пространственную группу $P2_1$, с эквивалентными позициями:

$$(x, y, z), (-x, y + 0.5, -z).$$

Вектор Харкера $(2x, 0.5, 2z)$ лежит в сечении Харкера $(u, 0.5, w)$ и может быть использован для определения положения атома; действительно: $x=u/2$, y неопределен (имеет начало в $P2_1$, свободно изменяется вдоль оси z), $z=w/2$.

Использование сечений Харкера становится проще, когда в ячейке присутствуют несколько тяжелых атомов: их максимумы Харкера могут быть распознаны и использованы для локализации тяжелых атомов. Если у них достаточно большой атомный номер, они могут быть использованы как хорошая начальная модель, к которой может быть применен так называемый циклический метод уточнения с построением Фурье синтеза (*Method of Fourier Recycling*), для получения положений легких атомов и последующего восстановления всей структуры.

Для упрощения анализа карты Паттерсона могут быть полезны следующие приемы. Пики можно усилить, используя коэффициент $|E_h|^2$ вместо $|F_h|^2$; ко всему прочему, принимая во внимание, что процедура усиления может вызвать существенный эффект отсечения членов ряда Фурье, что вызывает заметные возмущения на карте Паттерсона, для ослабления которых можно использовать смешанные коэффициенты $(|E_h|/|F_h|)$ или $((|E_h|^3|F_h|)^{1/2})$ [45–47]. Также очень

удобно удалять пик в начале карты с карты Паттерсона: это достигается использованием в расчете коэффициентов:

$$|F'_h|^2 = |F_h|^2 - \sum_{j=1}^N f_j^2,$$

либо, равным образом, коэффициентов $(|E_h|^2 - 1)$.

Стратегия для определения кристаллической структуры методами синтеза Паттерсона с использованием больших известных молекулярных фрагментов была описана в [48]. Еще позднее, синтез Паттерсона была использован для более точного вычисления дифракционных модулей, нежели полные процедуры разложения дифракционного профиля [49–51]. Идея лежит в соотношении $(|F(\mathbf{r}^*)|^2 = T[P(\mathbf{u})])$: синтез Паттерсона первоначально обостряется положительностью карты и затем подвергается Фурье трансформации.

В качестве примера можно привести автоматическую процедуру, реализованную в программе EXPO [52], которая может быть описана последовательностью шагов:

1. Из порошковой дифракционной картины посредством метода Ле Бейла [33] извлекаются интегральные интенсивности $(|F|^2)$.
2. Вычисляется карта Паттерсона $P(\mathbf{u})$, которая преобразовывается в новую карту $P'(\mathbf{u})$. Используется критерий усечения. Для всех точек карты, имеющих плотность меньше определенного порога, устанавливается нулевое значение, для дальнейшего рассмотрения лишь основных пиков.
3. Инверсия $P'(\mathbf{u})$ дает новые дифракционные интенсивности, которые используются со значениями по умолчанию, заданными программой EXPO для получения нового набора $|F|^2$.
4. Процедура возвращается к шагу 2 и повторяется несколько циклов.

1.5.3. Прямые методы

Прямые методы называются *ab initio* подходы к решению кристаллических структур, способные рассчитывать фазы напрямую из значений структурного фактора. Начало этим методам было положено в 1948 г [53]. С приходом современных компьютеров и изощренных математических подходов они оказались способны на практике решать фазовую задачу для небольших молекул. Сегодня при помощи прямых методов решаются протеиновые структуры, содержащие вплоть до 2000 неводородных атомов в асимметричной ячейке, при условии доступности монокристалльных дифракционных данных с разрешением более чем 1.2 \AA .

Применение прямых методов к порошковым данным требует предварительной процедуры полного разложения дифракционной картины: предполагается, что отдельные дифракционные интенсивности доступны для каждого рефлекса в измеренном интервале углов 2θ . Вследствие

перекрывания пиков вычисление дифракционных модулей будет сопровождаться неизбежными ошибками: это ослабляет эффективность прямых методов (неправильные модули приводят к неправильным фазам), и по сей день делает определение кристаллической структуры из порошковых данных сложной задачей.

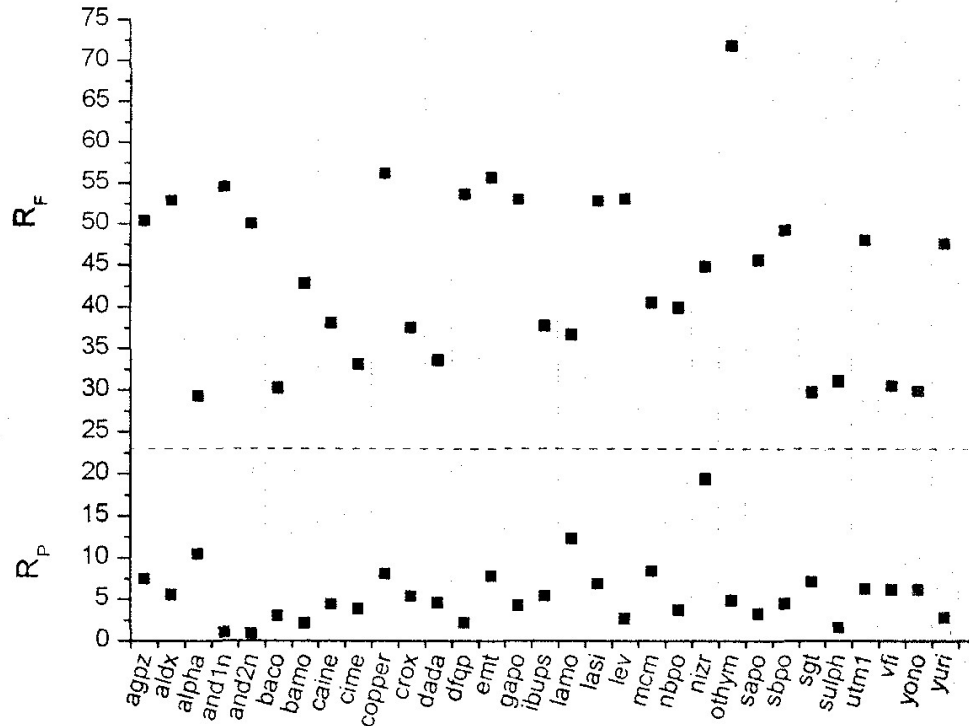


Рисунок 4 – R_F (верхняя часть) и R_P (нижняя), полученные программой EXPO с использованием функции профиля типа Пирсон VII, для набора тестовых структур. R_P – разница (в %) между двумя профилями: наблюдаемым и вычисленным при помощи выбранной профильной функции. R_F – разница (в %) между модулями, полученными методикой декомпозиции и модулями структурных факторов, вычисленными из опубликованных структурных моделей [54].

Недавно была исследована роль методик полного разложения дифракционной картины [54]: были применены два различных подхода извлечения, алгоритмы Паули [55] и Ле Бейла [56], использованы различные функции профиля дифракционного пика. Результаты можно свести к следующим:

- Систематические попытки улучшить подгонку профиля не обязательно приводили к большей точности экстрагированных интенсивностей. На рисунке 4 даны разность профиля R_P и ошибка вычисленного $|F|$ модуля (R_F) для нескольких тестовых структур: строгой корреляции для двух чисел не наблюдается.
- Выбор подходящей функции формы пика может улучшить подгонку профиля, но не обязательно эффективность разложения.

Высокие значения R_F являются, главным образом, следствием ограниченной информации, содержащейся в порошковой картине. Путь для преодоления этого ограничения, предложенный

Альтомаре и др. [56, 57], базируется на методе Монте-Карло: вычисляются несколько наборов экстрагированных интенсивностей, называемых «попытками разложения», которые подставляются один за другим для расчета прямыми методами. Каждая «попытка разложения» получается разложением общей интенсивности одного или более набора перекрывающихся рефлексов с тем условием, что интенсивность каждого набора постоянна.

В своей базовой форме прямые методы используют два типа предварительной информации: положительность карты электронной плотности (это условие можно опустить, например, в случае дифракции нейтронов, и атомарность (электроны не рассредоточены в элементарной ячейке, а сконцентрированы вокруг ядер). Эта информация, очевидно тривиальная, крайне полезна для достижения успеха на всех этапах процедуры вычисления современными прямыми методами: (1) масштабирование наблюдаемых интенсивностей и нормирование структурных факторов; (2) вычисление структурных инвариантов; (3) применению «формулы тангенса»; (4) дополнение и уточнение кристаллической структуры.

1.6. Методы определения кристаллических структур в прямом пространстве

Как было отмечено в предыдущей главе, решение кристаллической структуры из порошковых дифракционных данных традиционно достигалось первоначальным извлечением интенсивностей индивидуальных рефлексов из экспериментального профиля, и дальнейшим применением метода Паттерсона или прямых методов. Эффективность последних определяется сложностью извлечения однозначных значений индивидуальных интенсивностей из порошковой дифрактограммы. В качестве альтернативы в последнее время начали интенсивно развиваться т.н. *методы поиска структуры в прямом пространстве*, которые не требуют разложения дифракционной картины и напрямую используют экспериментальный дифракционный профиль. Некоторые данные об эффективности новых методов приведены в разделе 1.3.

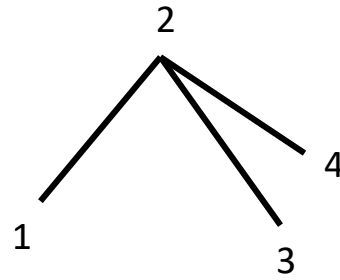
Методы в прямом пространстве используют *алгоритмы глобальной оптимизации* для ориентирования и локализации структурной модели. Эти алгоритмы способны находить глобальный минимум с любой случайной стартовой точки, и, таким образом, в состоянии избегать локальных минимумов (которые соответствуют некорректно определенным структурам). Каждой сгенерированной структурной модели ставится в соответствие значение *целевой функции (cost function, CF)*, которая является мерой сходства экспериментального и вычисленного порошковых дифракционных профилей. Методы глобальной оптимизации применяются для нахождения такой тестовой структуры, которой соответствует наименьшее значение CF. Это эквивалентно поиску глобального минимума гиперповерхности, определенной CF, по набору переменных, порождающих кристаллическую структуру. Поиск по сетке, метод Монте-Карло, метод имитации отжига и генетические алгоритмы являются наиболее распространенными для методами

поиска в прямом пространстве. Методы в прямом пространстве особенно выгодны в том случае, когда доступна значительная информация о молекулярной геометрии и/или кристаллохимии. Подобная информация вносит зависимость от предварительных химических ограничений: если она неверна, будут получаться только ошибочные решения, но в случае ее достоверности, генерируются только химически допустимые структурные модели. В особенности пригодны для решения методами в прямом пространстве (и сложны для прямых методов) молекулярные органические соединения: они обычно содержат структурные единицы с хорошо определенной геометрией и атомы с известной связностью.

Для построения химически правдоподобных структурных моделей может быть использован ряд программ, таких как Cerius [58], Chem 3D ultra [59], Sybyl [60] и Cambridge Structural Database (CSD) [61]. Получившаяся в результате модель часто описывается в терминах внутренних координат, т.е. длин связей, углов связей и торсионных (двугранных) углов. Обычно единственными параметрами, определяемыми методами поиска, являются двугранные углы (внутренние степени свободы) и параметры, определяющие положение и ориентацию одного или более молекулярных фрагментов (внешние степени свободы). Среди неорганических структур методами поиска в прямом пространстве могут хорошо определяться протяженные каркасные структуры, такие как цеолиты, содержащие четко определённые составные блоки. В этом случае химическая информация может быть вшита в составные части, например, многогранники атомов с известной геометрией. Тогда переменными будут только положения и ориентация рассеивающих блоков (внешние степени свободы) [62].

Использование заранее известной химической информации позволяет значительно уменьшить количество степеней свободы, необходимых для описания структуры, как для органических, так и для неорганических соединений, по сравнению с описанием через атомные координаты. В целом, можно добиться уменьшения числа степеней свободы примерно в два раза для неорганических структур и, по меньшей мере, в три раза для органических [62]. Более того, использование внутренних координат значительно уменьшает количество генерируемых пробных структур, так как исследуются только химически правдоподобные. Наиболее широко используемым для работы с пластичными моделями подходом, является представление их внутренних координат в т.н. Z -матрице (рисунок 5), где позиция каждого атома определена через длину связи, угол связи и двугранный угол по отношению к трем предшествующим атомам в матрице. В первом столбце матрицы на рисунке 5а присвоен номер всем атомам, изображенным на рисунке 5b, второй столбец содержит длины связей, соединяющих атом слева с атомом справа (третий столбец), четвертая колонка содержит угол, образованный атомами в столбцах 1, 3, 5 в одной и той же строке, а шестой столбец содержит двугранные углы, образованные атомами в столбцах 1, 3, 5, 7 в одной и той же строке.

1						
2	d_{12}	1				
3	d_{23}	2	θ_{123}	1		
4	d_{24}	2	θ_{124}	1	τ_{1234}	3



(a)

(b)

Рисунок 5 – (a) Описание Z -матрицей тетраэдра, показанного на (b). d_{ij} , θ_{ij} , τ_{ijkl} обозначают, соответственно, длину связи между атомами i и j , угол между атомами i , j и k и двугранный угол, образованный атомами i , j , k и l

В процессе вычисления методами в прямом пространстве часто требуются кристаллографические координаты для каждой модели. Алгебра и методики для преобразования внутренних координат в атомные координаты в декартовой системе и, наконец, кристаллографические координаты, описаны в [63].

1.6.1. Методы поиска по сетке

Наиболее прямолинейным подходом для решения кристаллической структуры методами в прямом пространстве является поиск перебором – задание сетки над интересующим пространством параметров и систематическое исследование каждой узловой точки. Значительное преимущество метода – в случае использования достаточно мелкой сетки он гарантирует нахождение глобального минимума. Кроме того, метод прост и легко реализуем в вычислительном коде. Однако он может быть применим только к относительно малому пространству параметров – когда число степеней свободы, необходимое для описания модели, очень ограничено, так как число точек сетки для исследования подвергается комбинаторному взрыву. При этом интервалы изменения позиционных и ориентационных параметров могут быть значительно сужены с учетом молекулярной и кристаллографической симметрии [64]. Кроме того, пространство переменных вращения (rotational space) может исследоваться эффективнее после переопределения связанных с ним параметров: число точек выборки может быть уменьшено на коэффициент $2/\pi \approx 0.64$, с равными объемом пространства для каждой точки, если обыкновенные эйлеровы углы θ , φ , ψ заменить квазиортогональными эйлеровыми углами [65]:

$$\theta_+ = \theta + \psi, \quad \theta_- = \theta - \psi.$$

Главным образом, процедуры сеточного поиска применяются к жестким структурным фрагментам [66–68]. В тех случаях, когда для конформационной пластичности необходимо описание внутренних степеней свободы, число точек сетки для исследования подвергается

комбинаторному взрыву, и поиск полным перебором становится невозможным. Некоторые авторы избегают данной трудности при помощи двухэтапной процедуры: вначале жесткая модель перемещается по шестимерной сетке, описанной внешними степенями свободы, а затем уточняется в положении, соответствующем минимуму оценочной функции, допуская изменения внутренней структуры в соответствии с ограничениями, наложенными на связи [69]. В целом, поиск по сети вычислительно труднообрабатываем для задач даже относительно низкой сложности и не сопоставим по эффективности с другими, ниже описываемыми методами поиска. Тем не менее, при ряде уместных модификаций, этот подход может оказаться полезным [2, с. 118].

1.6.2. Методы Монте-Карло

Для приблизительной локализации глобального минимума (т.е. получения решения кристаллической структуры хорошего качества) за разумное время, сеточные методы поиска приходится заменять вероятностными, основывающимися на случайной выборке пространства параметров. Подобный метод, называемый Монте-Карло (*МК*), широко применяется в различных сферах научной деятельности для имитации поведения сложных систем. Его применение к определению кристаллической структуры из порошковых дифракционных данных развивалось многими авторами, основные применяемые стратегии очерчены ниже.

Методы *МК* требуют построения марковских цепей структур, т.е. последовательности структур, каждая из которых зависит только от предыдущей. Каждая структура последовательности, называемая *конфигурацией*, определяется набором внешних и внутренних степеней свободы, необходимых для фиксирования положения, ориентации и внутримолекулярной геометрии предварительно заданной структурной модели. Данная конфигурация выводится из предыдущей путем случайными изменениями ее степеней свободы, и две последовательных конфигурации представляют шаг *МК* (*MC move*). На рисунке 6 в виде эскиза представлен единичный шаг работы метода *МК* для структурной модели, содержащей единственную внутреннюю степень свободы (неизвестный двугранный угол между двумя кольцами). Последовательность *МК* можно резюмировать следующим образом:

- а. Пробная конфигурация генерируется из данной старой конфигурации небольшими случайными смещениями набора $\{p_i\} = \{x, y, z, \theta, \varphi, \psi, \tau_1, \tau_2, \dots, \tau_n\}$, где первая тройка параметров являет собой положение центра масс структурной модели в ортонормированной относительной системе координат, вторая тройка параметров (обычно совпадающая с эйлеровыми углами) фиксирует ориентацию модели, а последние n параметров являются переменными двугранными углами. Новое значение p_i для каждого i -го параметра вычисляется из старого (назовем его p_i^{old}) в соответствии с выражением:

$$p_i = p_i^{\text{old}} + r_i s_i \Delta p, \quad (7)$$

где r_i есть случайный бит, принимающий значение $+1$ либо -1 , задающий направление смещения, s_i есть случайное число, постоянно берущееся из интервала $(0, 1)$, а $\Delta\rho$ – предварительно указанная величина максимального шага для i -го параметра.

- b. Для пробной структуры вычисляется порошковая дифракционная картина, которая сравнивается с экспериментальным профилем. Наиболее часто используемым фактором соответствия является профильный взвешенный индекс R_{wp} :

$$R_{wp} = \sqrt{\frac{\sum_t w_t (y_t^o - Ky_t^c)^2}{\sum_t w_t (y_t^o)^2}} \quad (8)$$

где y_t^o и y_t^c есть наблюдаемые (за вычетом фона) и вычисленные значения, соответствующие t -ой точке порошкового дифракционного профиля, K есть масштабный коэффициент, а w_t – весовой фактор, обычно берущийся пуассоновским ($w_t = 1/y_t^o$). R_{wp} используется как CF для оценки пригодности пробной структуры.

- c. Критерий принятия определяет, должна структура оказаться принятой или отвергнутой. В большинстве случаев он базируется на алгоритме выборки по значимости, разработанном Метрополисом [70]: тестовая конфигурация принимается, если $CF < CF_{old}$, в противном случае она принимается с вероятностью:

$$\exp[-(CF < CF_{old})/T], \quad (9)$$

где T есть подходящий масштабный фактор. Это завершается взятием случайного числа r между 0 и 1 и принятием пробной конфигурации, если $\exp[-(CF < CF_{old})/T] > r$.

- d. Если пробная конфигурация принята, она рассматривается как новая конфигурация в цепи Маркова и становится стартовой точкой для последующей генерации. Если пробная конфигурация отвергается, предпринимается новая попытка с предыдущей конфигурации до тех пор, пока критерий принятия не будет выполнен.

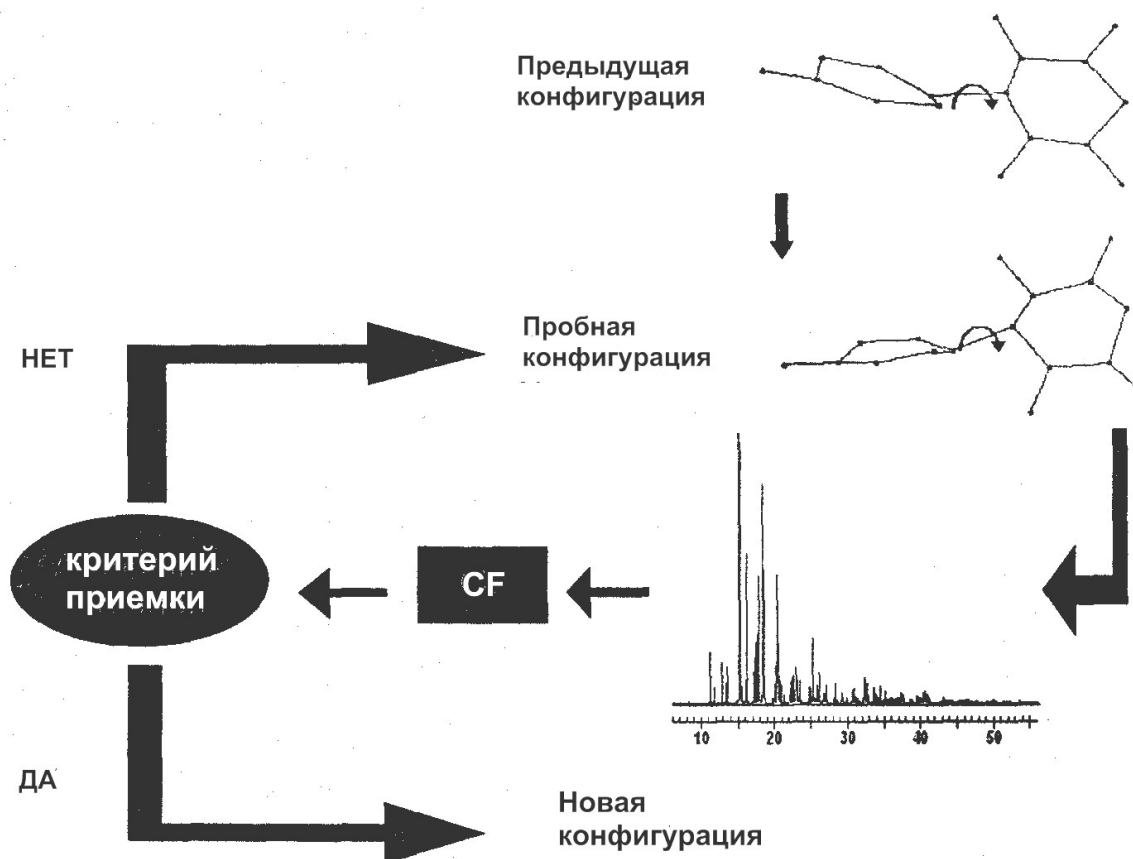


Рисунок 6 – Краткое изложение процедур, вовлеченных в одиночный шаг МК для структурной модели, содержащей единственный неизвестный двугранный угол между двумя кольцами

Последовательность шагов МК, которые могут быть представлены как случайное блуждание в пространстве параметров, повторяются до тех пор, пока не будет получена достаточно длинная цепочка структур. Алгоритмом выборки по значимости процесс направляется к глобальному минимуму гиперповерхности CF , который должен быть достигнут исследованием достаточно широкой области пространства параметров. Фактически, через уравнение (17) система избегает захвата локальным минимумом, так как оно позволяет, с некоторой вероятностью, принимать конфигурации с большей CF , и, следовательно, преодолевать барьеры на гиперповерхности CF . Скорость, с которой это происходит, может регулироваться температурным параметром T , который может изменяться в течение процесса минимизации согласно процедуре имитации отжига, описанной ниже.

Среди различных процедур МК, разработанных для решения фазовой задачи из порошковых данных, сошлемся на первопроходческие работы Соловьева и Кирика [71], Ньюмена, Харриса и др. [72], статью Андреева и др. [63], где были четко описаны математические процедуры для генерирования пластичных молекулярных структур. И на вклад Тримэйн и др. [73], разработавших компьютерную программу OSTOPUS [74], где вводится локальная минимизация

каждого шага *МК*, увеличивая эффективность алгоритма глобальной оптимизации. Последние две работы являются примерами того, как можно преодолеть один из главнейших изъянов методов в прямом пространстве: их зависимость от изначально полагаемой структурной модели. В статье Андреева и соавторов неизвестная часть структурной геометрии параметризуется и активно используется в течение глобальной оптимизации, но возникает эффект расширения пространства параметров поискового алгоритма. В статье Тримэйн и соавторов она фиксируется в алгоритме глобального поиска, уменьшая пространство параметров, но ослабляется в течение локальной минимизации. Затем вычисляется вероятность принятия, уравнение (9), используя CF , относящуюся к локально минимизированной конфигурации: если пробная структура принимается, генерируется новая, исходя из структуры со стандартной (а не смягченной) геометрией, во избежание существенных отклонений от стандартной молекулярной геометрии структурного фрагмента в ходе вычислений методом *МК*.

Методы *МК* могут быть полезны для вычисления потенциальной энергии структурной модели [75-83]. Они избегают либо значительно снижают возможность генерации нереалистичных структур, характеризуемых излишне короткими межмолекулярными контактами или невероятными длинами и углами связей. Основной целью введения в рассмотрение потенциальной энергии является не столько хорошее разграничение между различными правдоподобными структурами, сколько избегание неправдоподобных.

1.6.3. Методы имитации отжига

Фундаментальное отличие между стандартными методами Монте-Карло и методами имитации отжига заключается в способе, которым масштабный фактор T уравнения (9) используется для контроля алгоритма выборки. В первом случае, T фиксирован либо изменяется эмпирически, тогда как в последнем T медленно понижается в процедуре, согласно режиму отжига. В основе стратегии лежит аналогия между степенями свободы структурной модели, изменяющимися в процессе минимизации, и частицами в воображаемой физической системе. Последуем аргументации Киркпатрика и др. [84], которые впервые предложили методику имитации отжига. Целевая функция представляет собой энергию системы, соответственно глобальный минимум соотносится с ее основным состоянием. Вероятностный фактор критерия Метрополиса сходен с фактором Больцмана $\exp(-E/k_B T)$ для заселенности энергетического уровня E , где k_B – постоянная Больцмана, а T – температура. В процессе эволюции физической системы основное состояние может быть достигнуто медленным понижением ее температуры, через состояния термодинамического равновесия. Соответственно, корректная кристаллическая структура может быть получена при условии непрерывного уменьшения параметра T уравнения (9), избегая захвата системы в локальные минимумы, соответствующие ошибочным решениям структуры.

Оптимальная программа отжига, предложенная Киркпатриком [85], схема которой представлена на рисунке 7, может быть выведена путем рассмотрения свойств физических систем, составленных из большого числа взаимодействующих атомов. Чтобы придать текучесть низкоэнергетичному состоянию, система вначале должна быть нагрета до полного расплавления, затем необходимо медленно понижать температуру, особенно аккуратно вблизи точки затвердевания и, наконец, более резко охладить, дабы зафиксировать атомы. Той же последовательности можно придерживаться и в алгоритме глобальной оптимизации, рассматривая в качестве ключевых параметров температуру T и долю принимаемых решений (коэффициент принятия, acceptance ratio). Система начинает с данного значения T , которое возрастает до тех пор, пока коэффициент принятия не становится достаточно высоким (это гарантирует, что система расплавилась, т.е. степени свободы могут свободно изменяться в пространстве параметров); затем T уменьшается, сохраняя одно и то же значение, пока выполняется фиксированное количество шагов. Это наиболее долгий этап, в течение которого система должна достигнуть глобального минимума, избегая локальных минимумов. В заключение, положение глобального минимума уточняется быстрым понижением T . Итерация завершается, когда коэффициент принятия достигает некоторой выбранной малой величины, показывая отсутствие дальнейшего улучшения. В [86] показано, что сходимость к глобальному минимуму может быть достигнута логарифмической программой отжига, но нехватка времени для вычислений может подтолкнуть к выбору более быстрых программ отжига, успех которых уже не гарантирован.

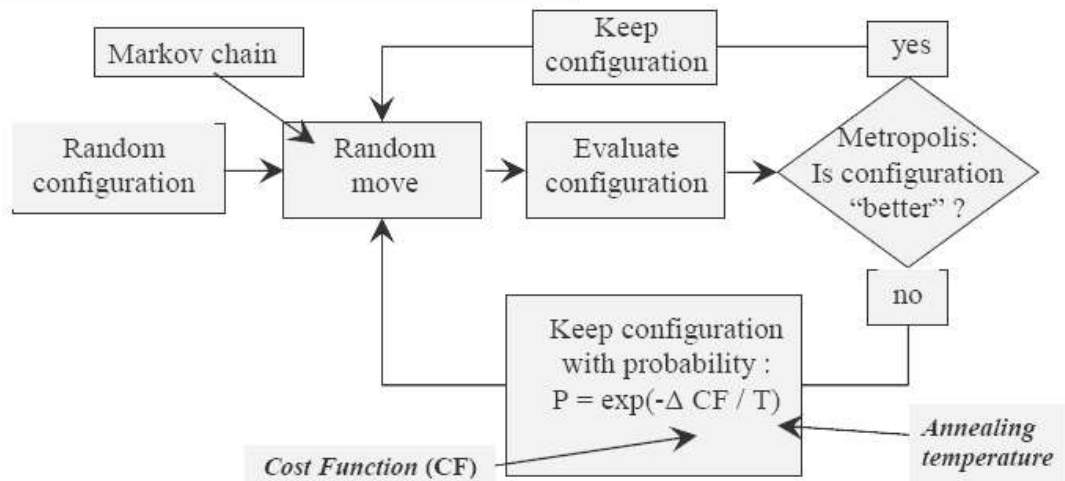


Рисунок 7 – Блок-схема алгоритма моделируемого отжига согласно реализации в программе FOX [63]. Из текущей точки совершается случайный шаг, вероятность принятия которого за следующую стартовую точку зависит от целевой функции (критерий Метрополиса). Управляющим параметром является «температура отжига» T

Было разработано несколько различных вариантов базового алгоритма имитации отжига, которые различаются выбором CF , построением программы отжига, либо процедурой генерирования пробных конфигураций. Например, Андреев и др. [87] уменьшают T с заданной скоро-

стью и, для каждой T , производят несколько перемещений; Дэвид и др. [88] представили следующие важные нововведения, включенные в компьютерную программу DASH [38]:

- а. CF основывается на сравнении между вычисленными модулями структурных факторов и интегральными интенсивностями, извлеченными согласно методике Паули по формуле:

$$CF = \sum_h \sum_k [(I_h - c|F_h|^2)(V^{-1})_{hk}(I_k - c|F_k|^2)] \quad (10)$$

где I_h и I_k – интегральные интенсивности, присвоенные, соответственно, рефлексам h и k уточнением методом Паули, V_{hk} – ковариационная матрица, полученная методом Паули, c – масштабный коэффициент, а $|F_h|$ и $|F_k|$ – модули структурного фактора, вычисленные из пробной структуры. Вычисление такой CF значительно быстрее, нежели профильного R -фактора, что приводит к быстрой идентификации корректной структуры. Для случаев, когда оптимизированная модель не дает полного описания асимметричной ячейки, CF выводится использованием методики максимального правдоподобия, приводящему к увеличению доли успеха алгоритма имитирования отжига [89].

- б. Изменение температуры связано с флуктуациями оценочной функции таким образом, что программа охлаждения замедляется, когда флуктуации CF велики. Это позволяет более подробное исследование пространства параметров, когда алгоритмом выбираются как хорошие, так и плохие решения: это эквивалентно более медленному охлаждению физической системы с большой теплоемкостью.
- с. Генерирование параметра происходит согласно экспоненциальному вероятностному распределению, в противоположность равномерному генерированию, подразумеваемому в уравнении (7). Сдвиг Δp параметра p_i извлекается в соответствии с вероятностным распределением:

$$P(\Delta p) = \frac{\Delta p}{\Delta p_0} \exp\left(-\frac{\Delta p}{\Delta p_0}\right)$$

где Δp_0 есть текущее характеристическое расстояние для параметра p_i . Это позволяет осуществлять более эффективную выборку прилегающего пространства параметров, пока существует конечный шанс на удачную выборку вблизи текущей точки [90].

- д. Для вращения модели используется кватернионное представление [91]. Три из четырех кватернионов фиксируют ориентацию осей вращения, оставшийся определяет вращение вокруг осей. Преимущество заключается в однозначной единообразной

выборке ориентаций, избегая сингулярностей, вносимых рассмотрением в виде эйлеровых углов [90].

Алгоритм моделирования отжига, реализованный в программе *PowderSolve* [92] включает похожие методики, с дополнением в виде зависимости ширины шага для изменений параметра [$\Delta\rho$ в уравнении (7)] от коэффициента принятия и флуктуации CF . Также введено локальное гашение, т.е. локальная оптимизация методом Ритвельда производится всякий раз, когда во время вычислений достигается перспективное решение. Это позволяет избежать скатывания к слишком низким температурам отжига, повышая эффективность метода.

Программа ENDEAVOUR [81] может искать кристаллические структуры ионных и интерметаллических соединений без использования начальной структурной модели, генерируя случайное распределение атомов в пределах элементарной ячейки. На каждом шаге MK произвольным перемещением 95% атомов (случайно выбранных) и парной перестановкой оставшихся 5% из предыдущей генерируется новая случайная атомная конфигурация. Также осуществляется локальная оптимизация, количество циклов которой зависит от числа атомов в элементарной ячейке. Данная зависимость представлена также в программе отжига, так что на общую процедуру сильно влияет содержимое элементарной ячейки, задаваемое пользователем. В этой программе часть *предварительно известной* структурной информации, обычно используемой в методах в прямом пространстве, добавляется в энергетическую часть оценочной функции, которая для металлических соединений в общем случае зависит от типа исследуемого соединения. Подобный подход был продолжен Коэлью [79] и реализован в программе *TOPAS* [34], где несколько энергетических членов, представляющих межатомное взаимодействие, суммируются для определения CF и, в дополнение, модулей случайных атомных перемещений, относящихся к ожидаемым средним значениям длины связей рассматриваемого соединения.

Программа ESPOIR [93] использует оригинальную CF , которая сочетает быстроту, свойственную обработке извлеченных интенсивностей (в отличие от расчетов профиля) с необходимостью учитывать перекрывание рефлексов. Интенсивности, извлеченные из наблюдаемых профилей, используются для воссоздания псевдо-картины дифракции, где фон, поляризация Лоренца, сложные формы профиля и кратность отражений не учитываются. Простой гауссовый пик с короткими хвостами, выбранный не более, чем по пяти точкам выше полной ширины на половине максимума, используется для грубого имитирования перекрывания на экспериментальном профиле. CF выражает соответствие между наблюдаемой и вычисленной псевдо-картинами.

Одной из наиболее популярных и широко используемых программ структурного анализа является свободно распространяемая программа FOX [62]. В ней для процесса определения структуры применяется усовершенствованная версия алгоритма имитирования отжига, называ-

емая *параллельной закалкой* (*parallel tempering*). Так как, в случае слишком быстрого уменьшения температуры, одиночная цепь конфигураций может быть поймана локальным минимумом, выполняется небольшое количество параллельных оптимизаций, каждая при отличной от других температуре. Между параллельными оптимизационными процессами введен обмен решениями, с использованием аналогичного с генерацией одиночной цепи критерия принятия. Данный подход особенно полезен для структурных моделей, заданных большим числом степеней свободы (обычно, более десяти). Метод адаптирован для GRID-вычислений.

С помощью FOX было решено большое количество новых структур значительной сложности. Примером может являться соединение $Al_2(H_3CPO_3)_3$, заданное фрагментами с общим числом степеней свободы 24 (рисунок 8); решение было найдено за 8 мин., потребовался перебор 500.000 вариантов. Кроме того, из 12 решений задач турнира *SDPD round robin-3*, 5 было получено именно этой программой.

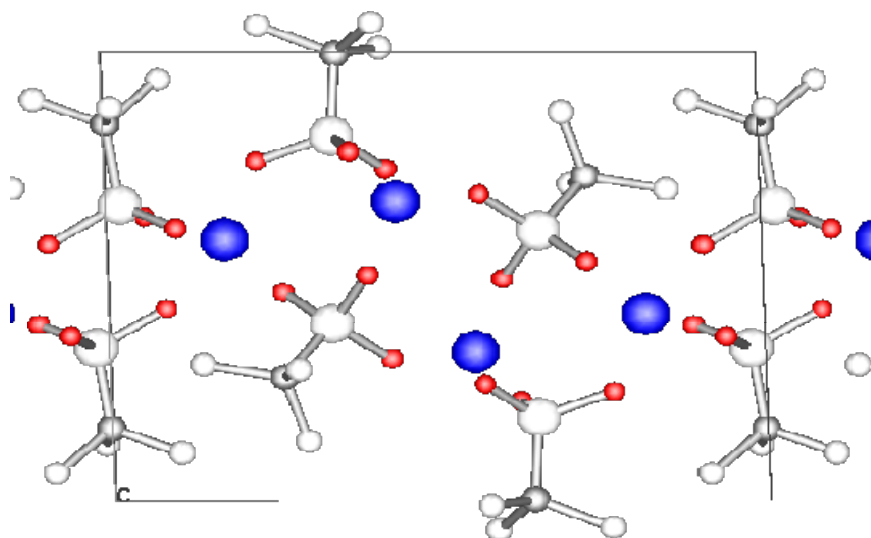


Рисунок 8 – Пространственная модель структуры $Al_2(H_3CPO_3)_3$, решенной с помощью FOX. В процессе поиска использовались отдельные структурные фрагменты:
3 H_3C-PO_3 и 2 Al (синим – Al, красным – O, белым – P, серым – C, белым (малые) – H)

1.6.4. Методы эволюционных и генетических алгоритмов

Генетический алгоритм [29] является альтернативным методом глобальной оптимизации по отношению к методам Монте-Карло и имитации отжига. Он основывается на принципах теории эволюции Дарвина и позволяет популяциям, состоящим из многих индивидов, эволюционировать по специальным правилам отбора к состоянию наилучшей *приспособленности* (*fitness*). Для соответствия нашей кристаллографической задачи терминологии биологической эволюции можно ввести соответствия, представленные в таблице 2.

Таблица 2 – Терминология эволюционных алгоритмов

Ген	Степень свободы (один из варьируемых параметров модели)
Хромосома	Последовательность степеней свободы (параметров модели), определяющая положение, ориентацию и внутреннее устройство структурной модели
Функция приспособленности (cost function - CF)	Критерий пригодности, сопоставляемый каждому индивиду
Эволюция	Поиск индивида с наилучшей приспособленностью (поиск наилучшей структуры), который является результатом последовательности эволюционных шагов от начальной популяции.

Операциями, включенными в единичный шаг эволюции являются:

- a. *Естественный отбор*. Он отбирает индивидов, предназначенных для выживания или подлежащих скрещиванию. Он подразумевает собой вычисление приспособленности для всех членов данной популяции и процедуру отбора, в некотором роде аналогичную критерию принятия метода имитации отжига, способную отмечать членов с наибольшей вероятностью выживания (таким образом, включая в себя некоторый элемент случайности).
- b. *Скрещивание*. Оно порождает новых потомков смешиванием генетической информации двух выбранных родителей.
- c. *Мутация*. Она генерирует новых индивидов (*мутантов*) произвольным изменением части генетической информации индивидов, случайно выбранных из популяции.

Все эти три процесса играют фундаментальную роль в методе генетического алгоритма: скрещивание важно для сохранения и улучшения выбранных черт в популяции, мутация важна для генетического разнообразия и предотвращает стагнацию популяции, естественный отбор ведет эволюцию к наиболее приспособленным популяциям, увеличивая продолжительность жизни индивидов, содержащих наиболее пригодные группы генов (*шаблоны*). Как общая стратегия, число членов одного поколения в будущем остается неизменным, таким образом, ключевым параметром является размер популяции: в слишком маленьких популяциях начинают доминировать отдельные индивиды, приводя к стагнации и скудной выборке пространства параметров, тогда как слишком большие могут замедлить сходимость к оптимальному решению.

Важной особенностью методов генетического алгоритма является свойственный им параллелизм, дающий возможность в одно и то же время обрабатывать различных членов популяции, таким образом, одновременно исследуя различные регионы пространства параметров.

Это делает генетический алгоритм, подобно вышеупомянутому методу параллельной закалки, особенно выигрышным для решения задач с многими степенями свободы.

Методику генетического алгоритма для решения структур из порошковых дифракционных данных независимо ввели две исследовательские группы: Кариуки и др. [94] и Шенкланд и др. [95]. Обе они использовали следующую стратегию:

- Начальная популяция произвольно выбранных индивидуумов составляется из хромосом следующего вида: $\{x, y, z | \theta, \varphi, \psi | \tau_1, \tau_2, \dots, \tau_n\}$, содержащих схемы трансляции, вращения и внутреннего строения.
- Процедура скрещивания выполняется в виде одноточечного перехода (single-point crossover), т.е. хромосомы двух родителей разрезаются в одной позиции и отрезанные элементы переставляются.
- Хромосомы подвергаются мутации присвоением нового произвольного значения одному или более случайно выбранных генов.

В деталях два подхода различаются многими другими аспектами реализации, наиболее значимым из которых является определение функции приспособленности: Шенкланд с коллегами используют функцию, основанную на интегральных интенсивностях, аналогично программе DASH [см. уравнение (21)], тогда как Кариуки с коллегами используют профильный фактор соответствия. В дополнение, они динамически масштабируют эту оценочную функцию преобразованием:

$$\rho = \frac{CF - CF_{\min}}{CF_{\max} - CF_{\min}}$$

где CF_{\min} и CF_{\max} есть минимальное и максимальное значения в пределах данной популяции, соответственно. ρ – аргумент \tanh или экспоненциальная функция, которая максимальна при $\rho = 0$ и минимальна при $\rho = 1$ [30]. в более поздних реализациях алгоритма [96] была введена концепция эволюции Ламарка, которая предполагает, что индивидуум может повышать свою приспособленность в течение жизни. Это достигается локальной минимизацией фактора соответствия каждой новой структуры, сгенерированной в процессе вычисления до применения операций естественного отбора, скрещивания и мутации. Так же как и для методов Монте-Карло, введение локальной минимизации приводит к существенному увеличению скорости сходимости алгоритма.

Новый эволюционный алгоритм, называемый дифференциальной эволюцией (ДЭ) [97] был успешно применен для решения кристаллических структур из порошковых данных [98]. ДЭ – более простой и детерминированный метод по отношению к генетическому алгоритму, и ба-

зируется на создании «детей» от единственного «родителя». В частности, каждый член популяции создает потомка, обладающего хромосомой:

$$parent + K(member_1 - parent) + F(member_2 - member_3), \quad (11)$$

где $parent$ есть хромосома его родителя, $member_i$, $i = 1 \dots 3$ соответствует хромосомам других произвольно выбранных членов популяции, а параметры K и F настраиваются в соответствии с уровнем рекомбинации и мутации, представленном в алгоритме. Операции, показанные в уравнении (11) применяются к индивидуальным генам хромосом, учитывают границы, соотношенные с каждым из них: если вновь сгенерированное значение какого-либо гена выходит за пределы этих границ, оно сбрасывается на усредненное значение между «родительским» и превышенным граничным. Новая популяция формируется детерминистским путем сравнения нового созданного потомка с его родителем. Алгоритм ДЭ показал себя более устойчивым, нежели генетический алгоритм, ценою более медленной сходимости.

Далее кратко рассмотрим ряд недавних разработок и характерные примеры определения с их помощью органических и неорганических кристаллических структур. Сразу отметим, что на текущий момент доступных реализаций немного, и большая часть разработок не является распространяемым программным продуктом.

Одной из первых реализаций ГА в структурном анализе стала программа *EAGER* разработки группы К.Д.М. Харриса [30], активно развиваемая и поныне. Однако, эта разработка не является распространяемым продуктом, и кроме того, нацелена на решение органических структур (в частности, фрагментов белков). Авторами было предложено использовать гибридный ГА с эволюцией по Ламарку (см. рисунок 9). Алгоритм использует промежуточную популяцию, объединяющую всех родителей и потомков (образованных скрещиванием родителей), из которой с помощью пропорциональной селекции определяется часть следующей популяции. Оставшаяся часть формируется путем мутации случайно выбранных индивидов промежуточной популяции. При этом все потомки и мутанты подвергаются локальному поиску по процедуре Пауэлла.

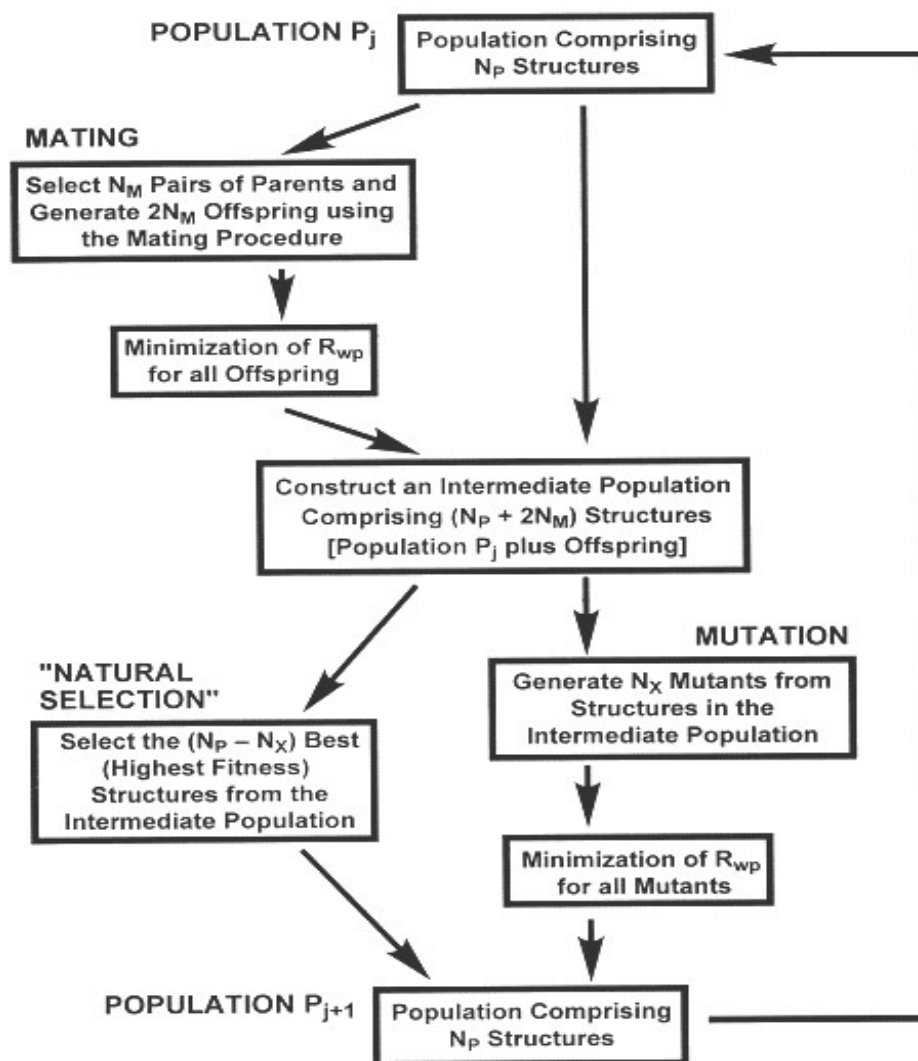


Рисунок 9 – Блок-схема гибридного генетического алгоритма

Примером работы алгоритма может служить задача определения структуры α -фазы L-глутаминовой кислоты (L-glutamic acid) $\text{HO}_2\text{C}(\text{CH}_2)_2\text{CH}(\text{NH}_2)\text{CO}_2\text{H}$ с 10 степенями свободы (см. рисунок 10). Популяция ГА содержала 100 индивидов; решение было найдено уже на 8 поколении, что оказалось существенно быстрее предыдущих реализаций, не содержащих процедуры локального поиска [96].

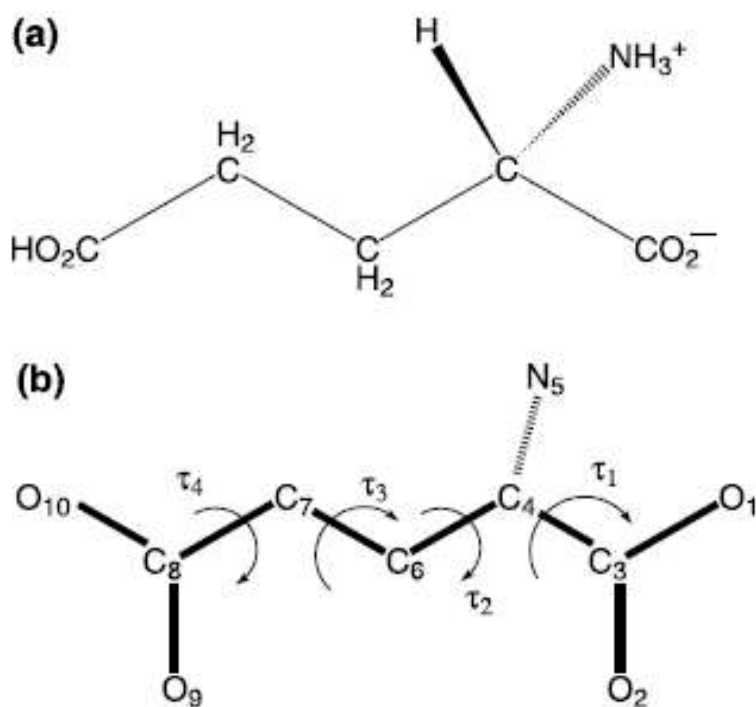


Рисунок 10 – (а) Молекулярная структура α -фазы L-глутаминовой кислоты (L-glutamic acid) $\text{HO}_2\text{C}(\text{CH}_2)_2\text{CH}(\text{NH}_2)\text{CO}_2\text{H}$;

(б) Соответствующий структурный фрагмент для оптимизации по ГА с 10 степенями свободы (указаны варьируемые торсионные углы)

В 2007 году была разработана свободно распространяемая программа структурного анализа *GEST* [99], также использующая ГА. В ней применена реализация алгоритма GENOCOP4.0 [100] (рисунок 11), в котором в качестве генов используется вещественное представление параметров. Операторы локального поиска не применяются. Целевая функция основана на брэгговском R-факторе и использует интегральные интенсивности для упрощения расчета. С помощью программы была определена структура нового полимера $\text{PbC}_7\text{H}_3\text{NO}_4 \cdot \text{H}_2\text{O}$ с 14 степенями свободы при размере популяции 500 и числе поколений 1000.

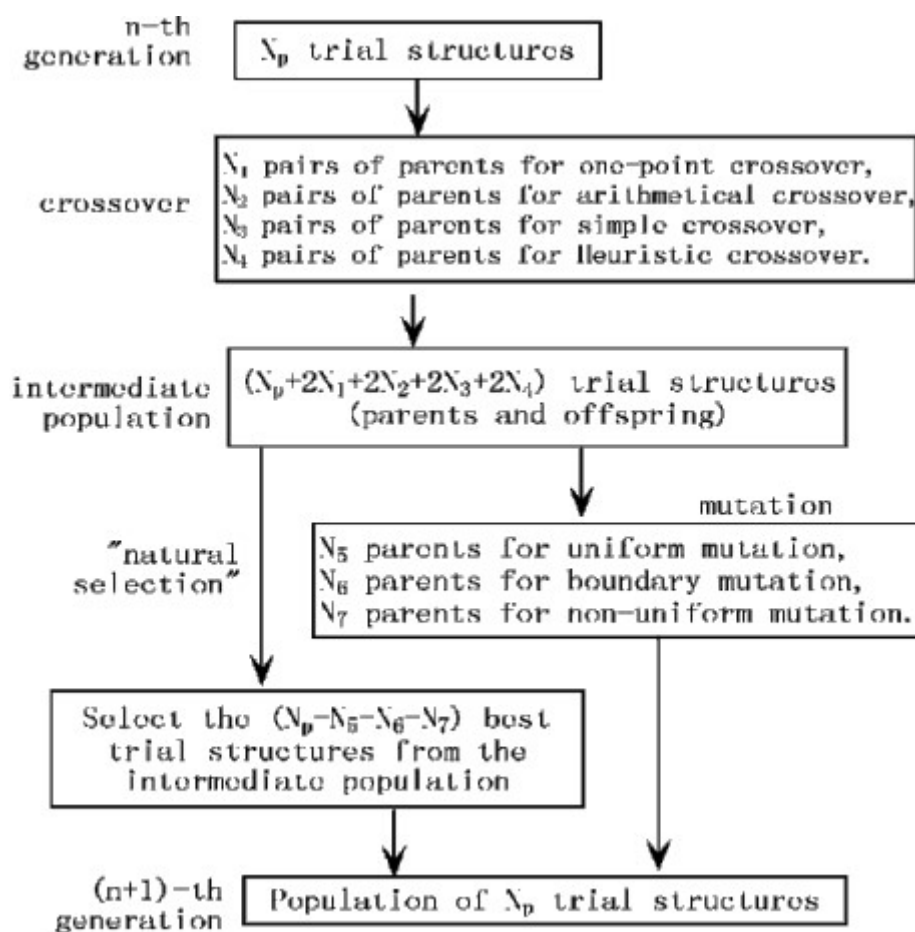


Рисунок 11 – Блок-схема генетического алгоритма из GEST. Для воспроизводства популяции применяются 4 типа операторов кроссовера и 3 – мутации; используется стратегия элитизма

Свободно распространяемая программа MAUD [101] также позволяет проводить полный процесс решения структуры «с нуля» с использованием стандартного ГА либо моделируемого отжига. Однако эта программа ориентирована на изучение текстуры порошкового образца и количественный фазовый анализ, поэтому методам определения структуры, их эффективности и применению к решению новых структур, разработчиком уделялось второстепенное внимание.

1.6.5. Гибридные подходы

Недавние разработки методов глобальной оптимизации привели к гибридным алгоритмам, которые способны совмещать наилучшие особенности двух различных методик. В качестве примера рассмотрим алгоритм, выведенный из методов Монте-Карло и молекулярной динамики (МД) [102]. Ориентация и положение модели молекулы здесь описывается как задача о дрейфе гипотетической частицы по шкале потенциальной энергии, заданной в пространстве параметров. Координаты частицы соответствуют внешним и внутренним степеням свободы структурной модели, а компоненты ее импульса представляют смещение соответствующих степеней свободы. Молекулярная динамика управляет движением частицы в пространстве пара-

метров посредством уравнений Гамильтона, где начальные компоненты импульса произвольно извлекаются из распределения Гаусса, а потенциальная энергия соответствует CF , приведенной в уравнении (10). После заданного количества шагов молекулярной динамики (MD steps), частица описывает траекторию в пространстве параметров, но величина ее полной энергии (кинетической + потенциальной) стремится отклониться от начальной вследствие систематических погрешностей, вносимых конечным размером шага симуляции методом молекулярной динамики. Поэтому выполняются шаги MK , с тем, чтобы принудительно сохранить полную энергию частицы, а траектория испытывается при помощи критерия принятия Метрополиса, основывающимся на сравнении ее начальной (E_{old}) и конечной (E) полной энергии. Траектория принимается, если $E < E_{old}$, в противном случае она принимается с вероятностью:

$$\exp[-(E - E_{old})/T],$$

по аналогии с уравнением (9). Если траектория принимается, новая начинается с ее конечной точки, в противном случае она стартует с начальной точки отвергнутой траектории; в обоих случаях новые компоненты импульса произвольным образом извлекаются из гауссова распределения. Гибридный Монте-Карло останавливается либо когда достигается заранее установленная нижняя граница потенциальной энергии, либо когда превышает максимальное число шагов метода MD + метода MK . В случае одиночного теста структуры данный подход повышает долю успешных попыток на 35% по отношению к алгоритму имитации отжига, реализованному в программе DASH.

Другой гибридный подход включает использование структурной огибающей. Она генерируется с использованием нескольких сильных рефлексов с низким разрешением, которые меньше подвержены перекрыванию и чьи фазы требуют извлечения, и производит разбиение элементарной ячейки на области с высокой и низкой электронной плотностью. Это приводит к сужению пространства параметров, используемого в методах глобальной оптимизации. Процедура сеточного поиска, ограниченная точками сетки, попадающими в область асимметричной ячейки, заданной структурной огибающей, преуспела в поиске решения цеолитовых [103] и органических [104] структур. В более поздних версиях процедуры, информация об огибающей была внедрена в рамки алгоритма имитации отжига, в особенности определения ее CF : $CF = wR_{wp} + (1 - w)P$, где w – весовой коэффициент, который может устанавливаться в зависимости от надежности сгенерированной огибающей, а P есть функция штрафа, оценивающая соответствие модели в текущей конфигурации огибающей [105].

Альтомаре и др. попытались соединить прямые методы с методом имитации отжига [106]: процедура реализована в программном пакете EXPO [41]. Это объясняется необходимостью использования предварительной информации о молекулярной геометрии для случаев, в которых классические прямые методы *ab initio* подхода к порошковым данным не способны добить-

ся полностью интерпретируемой карты электронной плотности. В качестве первого шага была разработана процедура дополнения неорганических кристаллических структур через метод Монте-Карло. Преимущество заключается в предварительном знании координации многогранников вокруг тяжелых атомов, локализованных при помощи прямых методов и, для каждого многогранника, среднего расстояния катион-анион [106]. Ожидаемые типы координации многогранника около каждого катиона в асимметричной ячейке определяются пользователем путем директив, вместе с ожидаемыми средними значениями длин связей и параметрами отклонений. Затем выполняются следующие шаги:

- Параметры положения тяжелых атомов, полученные прямыми методами, подвергаются автоматическому уточнению методом Ритвельда для улучшения их точности.
- Расстояния между тяжелыми атомами анализируются для установления (или подтверждения) катионной связности (тетраэдральной или октаэдральной). Находятся разумные позиции для анионов, проверяется их симметрия положения, генерируются их симметричные эквиваленты и контролируется полнота многогранника.
- Процесс циклизуется: процедура останавливается, когда было определено месторасположение всех анионов. Для каждой разумной модели вычисляется разностный профиль для соотнесения с наилучшими факторами соответствия для ритвельдовского уточнения.

Затем процедура была обобщена на случаи, когда корректно локализованы не все тяжелые атомы, присутствующие в элементарной ячейке, давая возможность локализации недостающих катионов и окружающих анионов, когда катионы координированы тетраэдрально либо октаэдрально [107].

Дополнение органической структуры является трудной задачей, так как карты электронной плотности, получающиеся при применении прямых методов, обычно содержат неполные и запутанные молекулярные фрагменты. Танахаси и др. [108] дополняют их комбинированием прямых методов и методов Монте-Карло, заимствованным из процедуры, описанной Харрисом и др. [72]. Позиции недостающих атомов могут быть найдены варьированием их координат на каждом шаге *МК* (MC move).

Альтомаре и др. [109] комбинируют прямые методы с имитацией отжига. Пики электронной плотности, полученные прямыми методами, используются для уменьшения числа степеней свободы глобального оптимизационного алгоритма в соответствии с тремя протоколами:

- а. Если на карте электронной плотности присутствуют триплеты или квадруплеты, они по очереди ассоциируются со всеми (если это возможно) похожими мультиплетными в структурной модели. За каждым назначением следует последовательность имитации отжига с внешними степенями свободы, строго ограниченными их начальными зна-

чениями (положение и ориентация структурной модели практически неизменна в течение процесса имитации отжига).

- b. Пары связанных пиков по очереди ассоциируются с парами связанных атомов модели. Последовательность имитации отжига, следующая за каждым назначением, обладает пространством параметров, ограниченным требованиями, чтобы центры тяжести пар карты и модели совпадали, а модель вращалась предпочтительно вокруг оси пары.
- c. Каждый пик карты по очереди ассоциируется со всеми атомами модели. В этом случае фиксируется только положение модели, тогда как ее ориентация и внутренняя структура должны быть найдены в процедуре имитации отжига.

В течение последовательностей имитации отжига, выбирается комбинированная *функция пригодности*, определяемая как произведение обычного профильного фактора соответствия (8) на коэффициент достоверности, который измеряет степень перекрытия между данной конфигурацией и пиками карты. После соотнесения всех мультиплетов модели с соответствующими мультиплетами карты по одному из трех протоколов, выбираются наилучшие полученные из различных последовательностей имитации отжига решения, подвергающиеся затем локальной минимизации. Ограничения, введенные для описания ассоциирования, ослабляются с целью уточнения местоположения оцениваемой модели. Можно отметить, что благодаря такому подходу, к корректному решению можно прийти из различных последовательностей имитации отжига, что делает поиск глобального минимума значительно более надежным.

1.6.6. Перспективы развития

Последнее десятилетие характеризуется интенсивной разработкой новых алгоритмов, которые упрощают решение кристаллических структур из порошковых данных. Теперь возможно определение кристаллических структур, состоящих из 20–40 атомов в асимметричной элементарной ячейке. Успехи прямых методов сопровождаются значительным развитием методов в прямом пространстве: их интеграция может явиться одним из самых многообещающих механизмов ближайшего будущего.

Прямые методы особенно подходят для случаев, когда молекулярная геометрия неизвестна. Их эффективность зависит от количества атомов в асимметричной элементарной ячейке, качества дифракционной картины и разрешения данных. Прямые методы извлекут пользу из любых усовершенствований (детерминистических или стохастических) методик разложения полного профиля.

Методы в прямом пространстве наиболее конкурентоспособны для органических соединений, которые более устойчивы к традиционным методикам и структурные модели которых

легко можно угадать. В настоящее время сложность кристаллических структур, решаемых методами в прямом пространстве, главным образом ограничивается числом степеней свободы, с которым могут справиться алгоритмы глобальной оптимизации за приемлемое время. В перспективе, улучшение, как алгоритмов поиска, так и вычислительных мощностей может преодолеть данное ограничение. Главными подводными камнями при использовании методов в прямом пространстве являются: (а) затратность по времени; (б) зависимость от наличия достоверной предварительной структурной информации. Частично некорректные модели могут подорвать успех процедуры вне зависимости от затраченного на нее времени; (с) чувствительность к точности параметризации профиля пика через функции его формы и ширины [110].

В соответствии с рассмотренными выше представлениями, наилучшие перспективы, в целом, имеют гибридные подходы, включающие прямые методы и методы в прямом пространстве. Прямые методы сужают пространство поиска и допускают оптимизацию более чем одной молекулы в асимметричной ячейке; методы в прямом пространстве упрощают решение структур, состоящих только из легких атомов, даже при использовании лабораторного дифрактометра.

1.7. Генетические алгоритмы как средство решения оптимизационных задач

Теория генетических алгоритмов является одним из основных инструментов реализации данного предложенного метода, поэтому рассмотрим здесь достаточно подробно ее современное состояние.

1.7.1. Основы теории генетических алгоритмов

Одним из самых популярных направлений многоагентных стохастических методов оптимизации являются эволюционные алгоритмы (ЭА). Идея использования эволюционных принципов для машинного обучения присутствовала еще в знаменитой работе Алана Тьюринга, посвященной проблеме создания мыслящих машин [111]. Это достаточно обширное множество разнообразных подходов в оптимизации [112]. Объединяет их общая идея – все они моделируют процессы, происходящие в эволюции: естественный отбор, воспроизводство, мутацию. ЭА базируются на коллективном обучающем процессе внутри популяции индивидуумов, каждый из которых представляет собой поисковую точку в пространстве допустимых решений данной задачи. Популяция случайно инициализируется, и затем охватывает лучшие регионы поискового пространства посредством случайных процессов селекции, мутации и рекомбинации. Окружающая среда представляет качественную информацию (степень пригодности) о поисковых точках (индивидуумах), а процесс селекции отбирает тех индивидуумов, у которых значение пригодности выше. Отобранные потомки являются, в свою очередь, родителями в следующем

поколении. Механизм рекомбинации перемешивает генетическую информацию родителей (тем самым рождается один или несколько потомков), и, наконец, механизм мутации способствует в некоторой степени обновлению генетической информации потомков [113]. Различные подходы эволюционных алгоритмов рассматривают эволюцию на разных уровнях – так генетический алгоритм [29] представляет решения в виде хромосом, состоящих из двоичных генов, алгоритм генетического программирования [114, 115] работает с решениями, представленными в виде деревьев, заимствуя свою терминологию из ботаники, эволюционные стратегии в качестве хромосомы используют строку вещественных чисел [116, 117].

Особое место среди эволюционных алгоритмов занимает генетический алгоритм. Основоположником его является Джон Холланд [118]. Термин "генетические алгоритмы" ввел в 1975 г. Д. Голдберг [29].

Генетические алгоритмы (ГА) применяются для решения таких задач, как [119]:

- Искусственная жизнь,
- Автоматическое обучение,
- Извлечение данных (знаний),
- Инженерное проектирование,
- Планирование и управление,
- Моделирование, идентификация,
- Численная и комбинаторная оптимизация.

В классическом генетическом алгоритме каждое решение кодируется в виде бинарной строки – такое представление называется хромосомой, а само решение – индивидом. Множество решений называется популяцией, итерации алгоритма – поколениями, выбор лучшего решения – селекцией, поисковые операторы – скрещиванием и мутацией. Таким образом, эволюционные алгоритмы заимствуют из биологии не только основные идеи, но и терминологию.

В общем виде работу генетического алгоритма можно представить следующим образом [29]:

1. Инициализировать случайным образом популяцию решений.
2. С помощью оператора селекции выбрать часть популяции (родителей) для порождения потомков.
3. Применить оператор скрещивания и получить потомков.
4. Новые решения (потомки) подвергаются мутации.
5. Формируется новая популяция: выбрать решения из родителей и потомков.
6. Повторять 2–5 пока не выполнится условие остановки.

На шаге инициализации задаются параметры алгоритма: длина хромосомы, размер популяции и др., а также типы и вероятности применения основных генетических операторов (скрещивания, мутации и селекции). Если априорные сведения о пространстве поиска отсутствуют, начальная популяция генерируется случайным образом.

Подобно тому, как природный хромосомный материал представляет собой линейную последовательность различных комбинаций четырех нуклеотидов, решения в генетическом алгоритме представляются в виде хромосом (генотипов). Генотип – строка конечной длины, состоящая из генов, представленных символами некоторого алфавита.

В генетическом алгоритме существует строгое различие между фенотипом (решением, выраженным в терминах поставленной задачи) и генотипом (хромосомой, представлением решения). Генетический алгоритм работает с генотипом, фенотип служит для определения пригодности индивида (оценки качества решения поставленной задачи), поэтому для работы алгоритма необходимо определить некоторую функцию кодирования ($e: D \rightarrow S$, где D – пространство поиска, S – пространство представлений решений) и функцию декодирования ($e^{-1}: S \rightarrow D$). Таким образом, на самом деле генетические алгоритмы решают не задачу $f(d) \rightarrow \underset{d \in D}{opt}$, где $f: D \rightarrow R^1$, а задачу $\mu(s) \rightarrow \underset{s \in S}{opt}$, где $\mu: S \rightarrow R^1$ и $s = e(x)$, $\mu(s) = f(e^{-1}(s)) = f(x)$.

На практике наибольшее распространение получили алгоритмы с бинарным представлением решений. Формально они решают задачу псевдобулевой оптимизации, т.е.

$$\chi(X) \rightarrow \underset{X \in B_{2^n}}{opt}, \text{ где } \chi: B_{2^n} \rightarrow R^1.$$

К такой задаче сводятся практически любые задачи с дискретными переменными (возможно выраженных в разных шкалах), а также задачи с непрерывными переменными (заменяя непрерывные переменные дискретными с заданной точностью). Наиболее часто используются стандартное бинарное кодирование и бинарные коды Грея, однако существуют и другие способы бинарного кодирования, эффективность которых еще только предстоит оценить [120].

Рассмотрим метод бинаризации для задачи безусловной параметрической оптимизации.

Метод бинаризации. В общем виде задачу безусловной оптимизации можно представить в виде (12), где переменные могут быть выражены в различных шкалах измерений.

$$\chi(X) \rightarrow \underset{X \in R \times F \times N \times P}{extr},$$

$$\chi: R \times F \times N \times P \rightarrow R^1, \quad (12)$$

где R – область изменения метрических переменных, F – область изменения порядковых переменных, N – область изменения переменных наименований, P – область изменения переменных перестановок.

Предложенный Л.А. Растригиным метод бинаризации позволяет сводить задачи типа (12) к задачам псевдобулевой оптимизации (13). Следует отметить, что к задаче (13) сводятся любые задачи дискретного программирования, а также непрерывного, путем дискретизации пространства поиска.

$$\chi(X) \rightarrow \underset{X \in B_{2^n}}{extr},$$

$$\chi: B_{2^n} \rightarrow R^1, B_{2^n} = (X: x_j \in B_2, j = \overline{1, n}), B_2 = \{0, 1\}. \quad (13)$$

Основная идея бинарного кодирования заключается в следующем. Булев вектор n компонент, может принимать 2^n - значений. Необходимо каким-либо образом установить соответствие между значениями переменных исходной задачи и значениями булевого вектора (задать алгоритм кодирования). В случае, если число значений переменных исходной задачи не равно 2^n , то либо переопределяют область изменения исходных переменных, либо одному значению исходной переменной ставят в соответствие несколько значений булевого вектора.

Перевод порядковых переменных в булевы.

Пусть дана порядковая переменная $f \in \{a_1, \dots, a_q\}$. Бинарным представлением переменной f будет вектор $X = (x_1, \dots, x_n)$, $x_i = 0 \wedge 1, i = \overline{1, n}$, где $n = \log_2 q$. Находим значение булевого

вектора, переведя его в целое число $k = \sum_{j=1}^n x_j \cdot 2^{j-1}, k \in [1; q]$. Тогда полагаем $f = a_k$.

Перевод метрических переменных в булевы.

Кодирование метрической шкалы начинается с перевода ее в порядковую. Для этого интервал $[a, b]$ разбивается на $h = \frac{(b-a)}{\varepsilon}$ равных частей (ε - точность решения задачи оптимизации). Теперь переменная r измеряется в порядковой шкале, $r \in \{y_1, \dots, y_{h+1}\}$, где $y_i = a + \varepsilon \cdot (i-1), i = \overline{1, h+1}$. Далее порядковая переменная преобразуется в булеву как показано выше.

Перевод переменных наименований и перестановок в булевы.

Эффективных алгоритмов перевода порядковых переменных и переменных наименований до сих пор не предложено. Непосредственный перевод их в булевы переменные дает результат аналогичный полному перебору.

Предложено по мере накопления информации о функционале в процессе оптимизации, определять отношения порядка на множестве перестановок и наименований. Далее работать так же, как и с порядковыми шкалами.

При кодировании вещественных параметров возникает проблема противоречия точность-размерность. Динамическое кодирование позволяет решить данную проблему следующим образом: в начале осуществляется грубый поиск с малой точностью (малой размерностью булевого вектора), затем поиск ограничивается некоторой перспективной областью (оператор увеличения), параметры перекодируются с более высокой точностью. Оператор увеличения может применяться многократно.

Код Грея это бинарный код целого числа, такой, что для смежных целых чисел расстояние между их кодами в метрике Хемминга равно 1.

Стандартное рефлексивное Грей-кодирование целого числа получается применением оператора исключающего ИЛИ к стандартному бинарному кодированию целого числа и к тому же бинарному коду, смещенному на одну позицию вправо, последний бит отсекается.

Кодирование и декодирование Грея можно выразить в виде операций с матрицами. Пусть x и y n -битовое бинарное и Грей-кодирование целого числа соответственно, и G – матрица трансформации, состоящая из 1 на главной диагонали и диагонали верхнего минора и 0 в других позициях. Грей-кодирование и декодирование тогда можно представить как $x^T G$ и $y^T G^{-1}$ соответственно. Можно показать, что любая перестановка столбца в матрице G приводит к другой матрице трансформации.

На практике можно использовать следующий алгоритм перевода бинарного кода в код Грея и наоборот: начать со старшего (правого) бита для получения Грей-кода и с левого для получения бинарного кода, последовательно заменять биты по правилу:

$$(\dots, 1, 1, \dots) \rightarrow (\dots, 1, 0, \dots)$$

$$(\dots, 1, 0, \dots) \rightarrow (\dots, 1, 1, \dots)$$

$$(\dots, 0, 1, \dots) \rightarrow (\dots, 0, 1, \dots)$$

$$(\dots, 0, 0, \dots) \rightarrow (\dots, 0, 0, \dots)$$

На каждом шаге заменяется два бита и производится смещение на один бит таким образом, что на последующем шаге заменяется один новый бит и один старый.

Пример:

$X = (10011)$ – бинарное представление целого числа, Y – его Грей-код.

а) Применение исключающего ИЛИ (XOR).

$$XOR = \begin{cases} 1, & \text{if } x_1 = 1, x_2 = 0 \\ 0, & \text{if } x_1 = 0, x_2 = 1 \\ 0, & \text{if } x_1 = 1, x_2 = 1 \\ 1, & \text{if } x_1 = 0, x_2 = 0 \end{cases}$$

$$\begin{array}{r} 10011 \\ xor \quad 01001 \\ \hline \end{array} \quad \begin{array}{r} 11010 \\ xor \quad 01001 \\ \hline \end{array}$$

$Y = (11010)$ и обратно $x = (10011)$.

b) Матричное кодирование.

$$G = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$Y = X^T \cdot G = (10011) \cdot \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} = (11010)^T$$

$$X = Y^T \cdot G^{-1} = (11010) \cdot \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} = (10011)^T$$

c) Алгоритм последовательной замены битов.

$$\begin{array}{r} X = (100\underline{11}) \quad Y = (\underline{11}010) \\ \quad \underline{100}10 \quad \quad \underline{100}10 \\ \quad \underline{100}10 \quad \quad \underline{100}10 \\ \quad \underline{100}10 \quad \quad \underline{100}10 \\ \hline \underline{Y} = (\underline{11}010) \quad \text{и} \quad \underline{X} = (100\underline{11}) \end{array}$$

Для многих практических задач код Грея оказывается более эффективным, чем бинарное кодирование. Код Грея частично сохраняет структуру окрестностей пространства поиска в бинаризованном пространстве. В результате, Грей-код не может образовать оптимумов больше, чем у исходной целевой функции. Более того, т.к. у текущей точки в коде Грея соседних точек больше, чем в дискретной целевой функции, то в пространстве поиска Грей-кода число оптимумов обычно меньше. В противоположность, бинарный код часто создает новые оптимумы,

которых у исходной целевой функции не было. Пример различных систем окрестностей показан на рисунке 12.

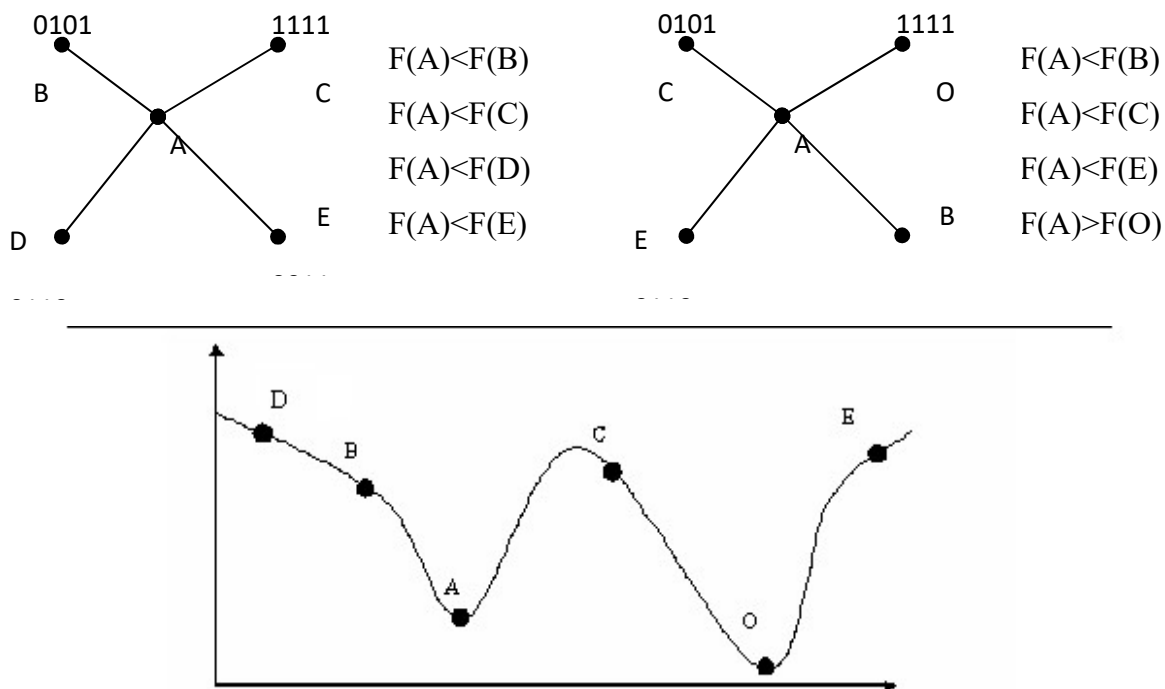


Рисунок 12 – Различные системы окрестностей при бинарном кодировании

1.7.2 Получение фенотипа из генотипа

Иногда получение фенотипа из генотипа – это простой и очевидный процесс (например, вычисление значения функции декодирования). В то же время генотип может быть рядом параметров для алгоритма, который, обрабатывая исходные данные, производит фенотип (рисунок 13).

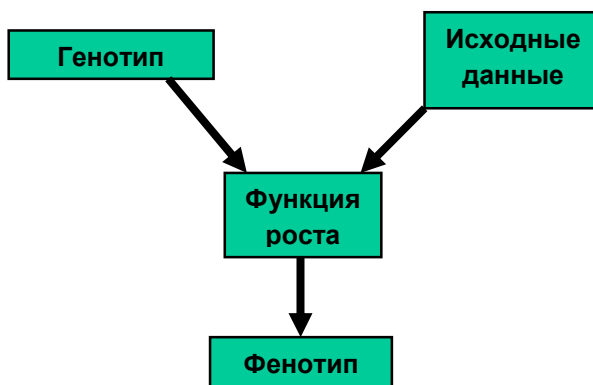


Рисунок 13 – Получение фенотипа из генотипа

Инициализация

Если априорная информация о пространстве поиска отсутствует, начальная популяция должна быть инициализирована равномерно в пространстве поиска (если это возможно). Причем, индивиды могут быть распределены равномерно как в пространстве генотипов, так и в пространстве фенотипов. Примеры:

- Двоичное представление: генерировать 0 или 1 с вероятностью $\frac{1}{2}$;
- Вещественное представление: генерировать решение равномерно на заданном интервале.

Если существуют некоторые предположения о пространстве поиска или известны некоторые варианты решения задачи, можно распределить популяцию в соответствии с имеющимися эвристиками или результатами предыдущих опытов. Однако, возможны следующие негативные последствия:

- Потеря генетического разнообразия;
- Непоправимое смещение хода эволюции.

Оператор селекции – оператор, посредством которого индивиды выбираются для порождения потомков или для перехода в следующее поколение. Наиболее приспособленные особи должны выбираться с большей вероятностью для сохранения своих генов в следующем поколении. Т.о., оператор селекции позволяет сконцентрировать поиск на наиболее многообещающих регионах пространства.

Предложено множество различных схем отбора в генетическом алгоритме. Конкретный тип оператора селекции (а также и др. генетических операторов) проектируется исходя из решаемой задачи (например, когда требуется контроль допустимости решений, учет ограничений, многокритериальности и т.д.), однако наиболее распространены следующие базовые типы селекции.

В *пропорциональной селекции* вероятность индивида быть отобранным пропорциональна его пригодности. Вероятность вычисляется следующим образом (для задачи минимизации):

$$P(X^i) = \frac{-fitness(X^i) + C}{r * C - \sum_{j=1}^r fitness(X^j)},$$

где r – размер популяции, $C : P(X^i) \geq 0, \forall i, \sum_{j=1}^r P(X^j) = 1$.

Пропорциональная селекция обладает следующими недостатками: преждевременная сходимость и стагнация.

Стагнация возникает, когда на определенном этапе поиска все индивиды получают относительно высокую и примерно равную пригодность, что приводит к очень низкому селективному давлению (наилучшее решение лишь немного предпочитается худшему).

Преждевременная сходимость (проблема супериндивида) возникает, когда на ранних этапах появляется индивид с пригодностью намного большей, чем у других индивидов в популяции, но очень плохой с точки зрения решаемой задачи. Вероятность супериндивида быть отобранным стремится к единице, в то время как вероятности других членов популяции – к нулю. В итоге он копирует себя в следующее поколение и вскоре «широкий» поиск прекращается.

При применении *ранговой селекции* индивиды популяции ранжируются в соответствии с их пригодностью: $R_i < R_j$ если $f(X^i) \leq f(X^j)$. Тогда

$$P(X^i) = \frac{R_i}{\sum_{k=1}^r R_k} = \frac{i}{\sum_{k=1}^r i} = \frac{2i}{r(r+1)}, \text{ где } \sum_{j=1}^r P(X^j) = 1.$$

Ранговая селекция устраняет недостатки пропорциональной: нет стагнации, т.к. даже к концу работы алгоритма $P(X^1) \neq P(X^2) \neq \dots$, нет преждевременной сходимости, т.к. нет индивидов с вероятностью отбора близкой к единице.

В *турнирной селекции* для отбора индивида создается группа из t ($t \geq 2$) индивидов, выбранных случайным образом. Индивид с наибольшей пригодностью в группе отбирается, остальные – отбрасываются. Параметр t называется размером турнира. Наиболее популярным является бинарный турнир. Этот тип селекции не требует сортировки популяции и вычисления пригодности для всех индивидов. Недостатки: худший индивид никогда не выбирается.

Селекция с усечением. В процессе селекции с усечением порогом τ , только доля τ из всех лучших индивидов может быть отобрана, причем в этой доле каждый имеет одинаковую вероятность отбора.

Элитарная селекция. Как минимум одна копия лучшего индивида всегда переходит в следующее поколение. Преимущества: гарантия сходимости. Недостатки: большой риск захвата локальным оптимумом.

Оператор скрещивания (рекомбинации) – генетический оператор поиска. При скрещивании отобранные индивиды (родители), обычно двое, по заданному правилу передают потомку части своих хромосом. Потомок может унаследовать только те гены, которые есть у его родителей. Существуют различные схемы скрещивания: одно- и двухточечное, равномерное, скрещивание более чем двух родителей [121, 122], скрещивание всех индивидов родительской популяции и другие. При точечном скрещивании для выбранных родителей выбираются точки разрыва хромосомы, родители определенным образом обмениваются соответствующими участ-

ками хромосом. При равномерном – соответствующий ген потомка может быть унаследован от любого родителя с равной вероятностью. Скрещивание осуществляется с вероятностью p_{xover} , иначе с вероятностью $(1 - p_{xover})$ родители клонируются в следующее поколение [123, 124].

Наиболее популярным типом скрещивания является *одноточечное скрещивание* – случайно выбирается точка разрыва, родительские хромосомы разрываются в этой точке и обмениваются правыми частями (рисунок 14).

При двухточечном скрещивании хромосома как бы замыкается в кольцо, выбираются 2 точки разрыва, родители обмениваются частями (рисунок 15).

При равномерном скрещивании потомок может унаследовать с равной вероятностью гены любого из родителей (рисунок 16).

Равномерное скрещивание по всей популяции (uniform gene pool recombination) получается применением равномерного скрещивание ко всем членам популяции, т.е. потомок может унаследовать любой ген, имеющийся в популяции в заданной позиции хромосомы [125].

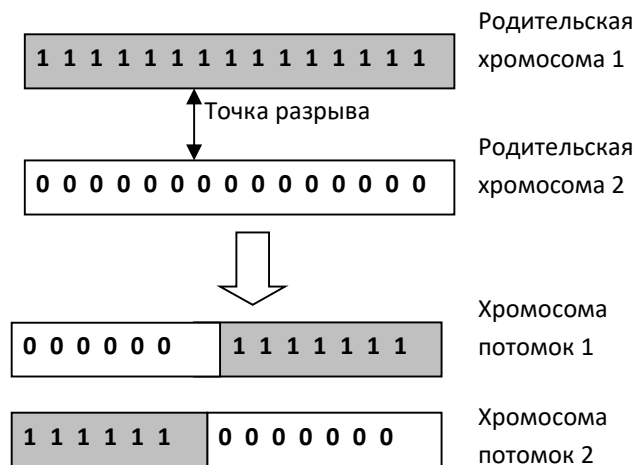


Рисунок 14 - Одноточечное скрещивание

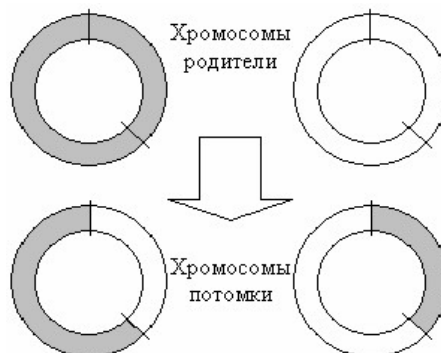


Рисунок 15 - Двухточечное скрещивание

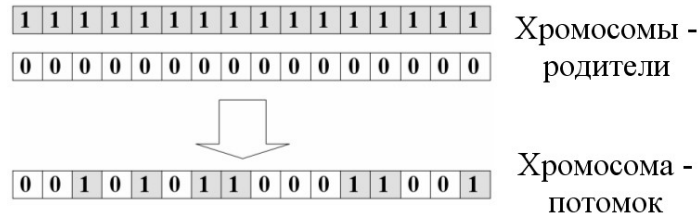


Рисунок 16 - Равномерное скрещивание

Оператор мутации – одноместный оператор поиска, случайное изменение в одном или нескольких генах индивида. В генетическом алгоритме мутация рассматривается как метод восстановления потерянного генетического материала, а не как поиск лучшего решения. Мутация применяется к генам с очень низкой вероятностью $p_m \in [0.001; 0.01]$. Хорошим эмпирическим правилом считается выбор вероятности мутации равным $p_m = \frac{1}{n}$, где n - число генов в хромосоме (в среднем хотя бы один ген будет подвержен мутации). В случае бинарного алфавита мутация состоит в инвертировании случайно выбранных битов (рисунок 17).

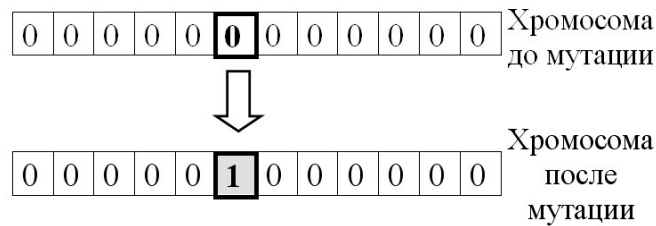


Рисунок 17 - Пример мутации в ГА.

Пригодность индивида – это некоторая количественная оценка качества решения поставленной задачи. Функция пригодности может быть спроектирована с учетом особенностей решаемой задачи (что делает генетический алгоритм довольно гибким, универсальным). Обычно функция пригодности принимает положительные значения, и значение пригодности максимизируют.

В общем случае при решении задач безусловной оптимизации в качестве значения функции пригодности может выступать значение оптимизируемого функционала ($fitness(X^i) = \chi(X^i)$).

Одним из главных недостатков генетических алгоритмов, на которые часто указывают их критики, является большое количество настраиваемых параметров. Неспециалисту в информационных технологиях не просто разобраться, какие значения параметров генетического алгоритма больше подходят для конкретной задачи. Эта проблема успешно решается применением коэволюционного генетического алгоритма.

1.8. Полнопрофильный анализ методом Ритвельда

Метод Ритвельда был разработан для уточнения кристаллической структуры по профилю дифрактограммы еще в 1969 году [126]. Широкое применение метода Ритвельда отмечено выше. Метод сводится к минимизации квадрата разности между экспериментальным и рассчитанным профилем с оптимизацией набора профильных и структурных параметров фаз-компонентов анализируемого материала по нелинейному методу наименьших квадратов (МНК). Функционал МНК имеет вид:

$$\Phi(\mathbf{P}) = \sum_j w_j (Y^{mod}(\bar{\mathbf{P}}, \theta_j) - Y^{эксн.}(\theta_j))^2 = \sum_j w_j (Y(\bar{\mathbf{P}}, \theta_j) - Y_j)^2 \rightarrow \min_{\mathbf{P}} \quad (14)$$

где $Y^{эксн.}(\Theta_j)$ и $Y^{mod}(\bar{\mathbf{P}}, \Theta_j)$ - интенсивности экспериментальной и вычисленной дифрактограмм в точках $j = 1, 2, \dots, N$, образующих профиль рентгенограммы; w_j - вес измерения.

Интенсивность Y^{mod} вычисляется по формуле:

$$Y^{mod}(\bar{\mathbf{P}}, \Theta_j) = K \sum_j S_i \sum_i I_{i,hkl}(\bar{P}_{i,str}) \cdot \Omega_i(\bar{P}_{i,prof}, \Theta_{i,hkl}, \Theta_j) \cdot A_i \cdot t_{i,hkl} + y^\phi(\Theta_j), \quad (15)$$

где K - коэффициент приведения к единой шкале;

Y_j^ϕ - модельная функция фона рентгенограммы;

S_i - масштабный коэффициент для расчетной рентгенограммы фазы i ;

$\bar{\mathbf{P}} = (\bar{P}_{prof}, \bar{P}_{str})$ - вектор параметров, включающий профильные и структурные параметры всех фаз образца,

Ω_i - функция профиля i -й фазы для рефлексов hkl_i (псевдо-Войгта, Пирсона и т.д.);

t_{ih} - набор мультипликативных поправок на интенсивность рефлекса $(hkl)_i$ фазы i (текстура и т.п.);

A_i - набор мультипликативных поправок на профиль рефлексов (hkl) фазы i (асимметрия, и т.п.);

I_{ih} - интегральная интенсивность рефлекса (hkl) фазы i , являющаяся функцией структурных параметров \bar{P}_{str_i} фазы i ; для монохроматического излучения от плоского образца в дифрактометре с фокусировкой по Брэггу-Бретано она равна [134]:

$$I_{hkl,i} = \left[\frac{I_0 \lambda^3}{32 \pi r} \frac{e^4}{m_e^2 c^4} \right] \cdot \left[\frac{M_{hkl}}{2V_i^2} |F_{(hkl)_i}|^2 \left(\frac{1 + \cos^2 2\theta \cos^2 2\theta_m}{\sin^2 \theta \cos \theta} \right) \right] \cdot \left[\frac{W_i}{\rho_i \mu_m^*} \right] \quad (16)$$

где W_i и ρ_i - массовая доля и плотность фазы i ;

μ_m^* - массовый коэффициент поглощения образца m ;

V_i - объем элементарной ячейки фазы i ;

$|F_{(hkl)_i}|^2$ - структурный фактор рефлекса hkl фазы i , определяемый из координат атомов:

$$|F_{hkl}|^2 = \left| \sum_j \eta_j \cdot f_j \sum_m e^{-2\pi i(hx_{jm} + ky_{jm} + lz_{jm})} \cdot t_{jm} \right|^2 \quad (17)$$

где j, m - суммирование по атомам, соответственно, независимой и зависимой части ячейки; η_j - коэффициент заполнения атомом j кристаллографической позиции; f_j - атомный фактор рассеяния; x, y, z - координаты атомов в долях ячейки; t_{jm} - анизотропный тепловой фактор атома j позиции m (или изотропный t_j).

Для работы метода Ритвельда нужно задать исходные приближения P_0 уточняемых параметров P . Минимум функционала МНК достигается при некотором $P^* = P_0 + \Delta P$. Найдем это P^* .

Необходимым условием минимума функционала МНК является:

$$\frac{\partial \Phi(\bar{P}^*)}{\partial P_k} = \sum_j^m (Y(\bar{P}^*, \theta_j) - Y_j) \frac{\partial Y(\bar{P}^*, \theta_j)}{\partial P_k} = 0, \quad k = 1, \dots, N.$$

Решить эту систему уравнений относительно P^* не представляется возможным ввиду ее существенной нелинейности, но ее можно решать приближенно, линеаризовав относительно ΔP . Для этого разложим $Y(P_0 + \Delta P, \theta_j)$ по формуле Тейлора:

$$\sum_j^m \left(Y(\bar{P}_0, \theta_j) + \sum_i^N \frac{\partial Y(\bar{P}_0, \theta_j)}{\partial P_i} \Delta P_i - Y_j \right) \frac{\partial Y(\bar{P}_0, \theta_j)}{\partial P_k} \approx 0, \quad k = 1, \dots, N,$$

откуда

$$\sum_i^n \sum_j^m \frac{\partial Y(\bar{P}_0, \theta_j)}{\partial P_i} \frac{\partial Y(\bar{P}_0, \theta_j)}{\partial P_k} \Delta P_i \approx \sum_j^m (Y_j - Y(\bar{P}_0, \theta_j)) \frac{\partial Y(\bar{P}_0, \theta_j)}{\partial P_k},$$

$$k = 1, \dots, N.$$

Получаем систему линейных уравнений от n неизвестных, которую можно решить, например, методом Гаусса. Последовательно находя $\Delta P_1, \Delta P_2, \dots$ (т.е. уточняя значения P_0), будем приближаться к искомому P^* .

В качестве характеристики «качества» решения применяются один или несколько R -факторов.

1. Профильный (R_p):

$$R_p = \frac{\sum_j |Y_{Oj} - Y_{Cj}|}{\sum_j Y_{Oj}}, \quad (19)$$

где Y_{Oj} – измеренные в точке θ_j интенсивности, поправленные на фон; Y_{Cj} – расчетные значения.

2. Взвешенный профильный (R_{wp}):

$$R_{wp} = \sqrt{\frac{\sum_j w_j (Y_{O_j} - Y_{C_j})^2}{\sum_j w_j (Y_{O_j})^2}}, \quad (20)$$

где w_j – весовой коэффициент, отражающий надежность проведенного измерения, обычно равный $1/Y_{O_j}$.

3. R -фактор по интегральным интенсивностям (R_I – брэгговский) – то же, что и R_p но с использованием интегральных интенсивностей.
4. Ожидаемый R -фактор (R_{exp}):

$$R_{exp} = \sqrt{\frac{(N-P)}{\sum_i^N w_i Y_{O_i}^2}}, \quad (21)$$

где N – число измерений, P – число параметров.

R_{exp} – статистически ожидаемый R -фактор, к которому должен приблизиться взвешенный профильный фактор.

Итерационный метод Ритвельда имеет три существенные проблемы.

1). При недостаточно точных начальных приближениях P_0 и большом числе параметров n система уравнений расходится, либо устойчиво попадает в локальный минимум. Чем больше n , тем точнее требуются P_0 .

2). Сходимость и точность метода существенно зависит от качества моделирования формы линий и фоновой составляющей. Хорошими аппроксимациями формы линий признаны асимметричные варианты распределений Пирсона VII и псевдо-Войгта, а фона – полиномами. Тем не менее, анализ материалов с заметным количеством аморфной фазы, дефектными или структурно разупорядоченными фазами и сложной микроструктурой представляет существенные трудности.

3) Метод недостаточно автоматизирован. На практике его использование трудоемко и требует длительного подбора «хороших» начальных приближений для части основных параметров и постепенного вовлечения в уточнение всех оставшихся.

1.9 Сравнительная оценка вариантов возможных решений и выбор направления развития генетического алгоритма эволюционной оптимизации

На основании обзора, приведенного в первой главе диссертации, можно сделать вывод, что методы структурного решения из порошковых данных интенсивно развиваются и в данное время позволяют стабильно получать результаты с количеством степеней свободы структурных параметров порядка 30-40, а в некоторых случаях – до ~ 100 . Неуклонный рост числа опублико-

ванных результатов, полученных методами глобальной оптимизации относительно числа результатов от прямых и др. методов, показывает, что именно методы глобальной оптимизации в прямом пространстве (Монте-Карло, имитации отжига, генетические алгоритмы) становятся наиболее эффективными при получении структур из порошковых данных. Эти методы комбинируют поиск позиций независимых атомов и целых молекул, представляющих априори известные структурные фрагменты. Их позиции задаются 3-мя координатами и торсионными углами, так, что общее число независимых и фиксированных или уточняемых атомов в структуре может достигать до сотни. Во многих случаях применение методов глобальной оптимизации комбинируется с прямыми или Паттерсоновскими методами. Уточнение, как правило, выполняется по методу Ритвельда. В целом, основным направлением развития методов решения структуры из данных порошковой дифракции должно быть увеличение количества степеней свободы определяемых параметров. Для этого требуется, во-первых, развитие методов в направлении повышения их устойчивости с ростом размерности задачи, во-вторых, вовлечение в решение доступной дополнительной информации, в 1-ю очередь, кристаллохимической.

В результате обзора современного состояния методов порошковой дифракции определен оптимальный вариант направления исследований, заключающийся в развитии и гибридизации генетического алгоритма с другими подходами. Гибридный генетический алгоритм должен быть комбинирован с эффективными математическими методами оптимизации и подходящими дифракционными, кристаллохимическими и физическими критериями оценки испытываемых структур. Это должно обеспечить эволюцию тестовых структур с устойчивой сходимостью в глобальный минимум, отвечающий истинной структуре, при существенном увеличении размерности структурных задач.

Разрабатываемый подход отличается от других методов глобальной оптимизации в прямом пространстве следующим. Во-первых, в нем применен гибридный генетический алгоритм, интегрированный с методом полнопрофильного анализа. При этом в качестве инструмента полнопрофильного анализа наряду с классическим методом Ритвельда использован новый метод «минимизации производной разности» (МПР или DDM) расчетной и экспериментальной дифрактограмм, не требующий моделирования фона дифрактограммы и улучшающий экстракцию очень слабых дифракционных рефлексов, что уменьшает размерность задачи ГА и повышает его сходимость. В качестве одного из вариантов в нем применен 2-х уровневый ГА. На 1-м уровне используется классический ГА, работающий с двоичным кодированием округленных значений параметров, инкапсулируемых в хромосомы, и предназначенный для генерации испытываемых структур и их Ламарковской эволюции по дифракционному критерию, основанному на динамике снижения R-фактора. Лучшие в смысле дифракционного критерия испытываемые структуры передаются на 2-й уровень ГА, где подвергаются эволюционной локальной оптими-

зации. Процесс оптимизации основан на испытании и эволюции вариантов уточнения различных групп профильных и структурных параметров по методу МПР. Критерием эволюции служит сходимость метода МПР. В процессе эволюции 2-го уровня, по сути, происходит поиск пути спуска на гиперповерхности R-фактора из точек, заданных испытываемыми структурами, полученными с 1-го уровня ГА. Лучшие из уточненных таким способом структур включаются в генерацию новой популяции испытываемых структур на 1-м уровне ГА и процесс эволюции продолжается. Преимуществами применения такой эволюционной оптимизации являются увеличение областей локальной сходимости и повышение вероятности глобальной сходимости одной из испытываемых структур к истинным значениям структурных параметров.

Однако, в качестве основного в работе предложен и экспериментально исследован новый вариант мультипопуляционного ГА (МПГА) для решения кристаллических структур. В нем индивидуальные генетические алгоритмы выполняются на разных узлах вычислительного кластера или многоядерного ПК. Лучшие структурные модели из всех узлов подвергаются локальной оптимизации по методу полнопрофильного структурного анализа и накапливаются на управляющем узле. Он управляет их выборочной миграцией обратно в популяции на вычислительных узлах. Стохастические процессы ГА на разных ядрах попадают в разные локальные минимумы на гиперповерхности R-фактора. Периодическое «разбавление» плохо сходящихся популяций лучшими структурными моделями из других популяций приводит к «конкуренции локальных минимумов». Это обеспечивает выходы из них и улучшает общую сходимость эволюционного процесса в глобальный минимум. Таким образом, этот подход имеет существенный потенциал развития за счет оптимизации параллельной организации эволюции при работе МПГА на суперкомпьютерных кластерах. На данный момент специфика библиотеки, реализующей параллельные вычисления, не позволила использовать в нем алгоритм DDM для локального поиска. Было принято решение о построении МПГА с включением кода метода Ритвельда из открытой библиотеки ObjCryst++, которая является основой программы FOX. Соответственно, в МПГА второй уровень ГА был заменён на периодический запуск локального поиска по методу Ритвельда по всем уточняемым координатам сразу.

Глава 2. Мультипопуляционный генетический алгоритм моделирования кристаллической структуры из рентгенодифракционных данных

Разработан мультипопуляционный параллельный генетический алгоритм (МППГА) для моделирования кристаллической структуры химических соединений из данных рентгеновской порошковой дифракции, предназначенный для работы на многоядерных ПК и многопроцессорных вычислительных кластерах. В нём на разных узлах вычислительного кластера индивидуальные генетические алгоритмы выполняют эволюцию различных популяций структурных моделей. При этом лучшие структурные модели периодически подвергаются локальной оптимизации по интегрированному в МППГА методу полнопрофильного структурного анализа и накапливаются на управляющем узле (УЯ). УЯ управляет их выборочной миграцией обратно в популяции на вычислительных узлах. Повышение эффективности поиска структур происходит за счет полнопрофильной оптимизации и миграции, обеспечивающей выход из локальных минимумов путем вовлечения в индивидуальные эволюционные процессы структурных моделей из других популяций.

2.1 Схема работы индивидуальных процессов ГА в составе МППГА

Структурные модели в индивидуальных ГА представлены в виде строк (индивидов ГА), составленных из параметров, используемых в программах полнопрофильного анализа. Это позволяет интегрировать ГА с полнопрофильным анализом по методам Ритвельда и МПР. Для межпопуляционной синхронизации смены поколений эволюции в МППГА индивидуальные процессы ГА должны быть высокопроизводительными.

2.1.1 Предварительная установка профильных параметров

В качестве исходных данных алгоритму МППГА требуются:

- параметры элементарной ячейки $\{a, b, c, \alpha, \beta, \gamma\}$ и пространственная группа симметрии всех фаз образца;
- профильные параметры всех фаз образца;
- данные об атомном составе для искомой фазы, и точное положение атомов для остальных фаз.

Параметры элементарной ячейки и пространственная группа искомой фазы определяется на этапе индицирования, выполняемом предварительно. Определение профильных параметров происходит в МППГА автоматически. При первом запуске по встроенному в МППГА методу Ле

Бейла [44, с. 134-165] выполняется декомпозиция дифрактограммы с определением профильных параметров, которые затем фиксируются.

Алгоритм определения профильных параметров такой:

1. Загружается дифрактограмма из текстового файла, указывается тип излучения и длина волны.
2. Задаётся линия фона из 20 точек, затем по МНК уточняется их положение для данной дифрактограммы (Bayesian Optimization).
3. По методу Ле Бейла последовательно уточняются параметры:
 - а) сдвиг нуля дифрактограммы, коэффициент полуширины W;
 - б) коэффициенты полуширины U и W, коэффициент гауссина Eta0;
 - в) коэффициент гауссиана Eta1;
 - г) коэффициенты асимметрии, displacement, transparency;
 - д) доуточнение коэффициентов полинома линии фона;
 - е) доуточнение параметров ячейки (если требуется).

Поправки на смещение образца описываются следующими формулами:

$$\textit{Sample displacement correction: } 2\Theta_{\text{obs}} = 2\Theta_{\text{real}} + \text{Displacement} / \cos(\Theta)$$

$$\textit{Sample transparency correction: } 2\Theta_{\text{obs}} = 2\Theta_{\text{real}} + \text{Transparency} * \sin(\Theta)$$

Полуширина пиков описывается по закону Каглиоти:

$$\text{FWHM} = [W + V * \tan(\Theta) + U * \tan^2(\Theta)]^{1/2}$$

В качестве функции формы пика используется функция псевдо-Войгта, которая является суммой вкладов лоренсовой и гауссовой функций:

$$PV(x) = \text{eta} * L(x) + (1 - \text{eta}) * G(x)$$

$$\text{eta} = \text{Eta0} + \text{Eta1} * 2\Theta$$

Коэффициент асимметричности пиков вычисляется по формуле:

$$A = A0 + A1 / \sin(2\Theta) + A2 / \sin^2(2\Theta)$$

2.1.2 Описание генетических операторов

Каждая из генерируемых тестовых структурных моделей в МПГА кодируется хромосомой в виде строки из вещественных чисел, представляющих собой параметры искомой фазы. Это могут быть координаты атомов, факторы тепловых колебаний, заполняемость атомной позиции и т.п. Каждая из величин варьируется в своём диапазоне, границы варьирования задаются для каждого параметра автоматически. Например, для атомных координат границами варьирования будут величины сторон независимой части ячейки. Вещественное представление параметров является основным в МПГА, но предусмотрена также и возможность стандартного для ГА бинарного кодирования параметров (обычно, по 5-ти битной вычислительной сетке).

В начале работы МПГА выполняет случайную генерацию n различных популяций структурных моделей вещества (индивидов ГА) и на каждом из n вычислительных ядер ПК или кластера запускает индивидуальный однопопуляционный ГА. Схема работы индивидуальных процессов ГА в МПГА показана на рисунке 18.

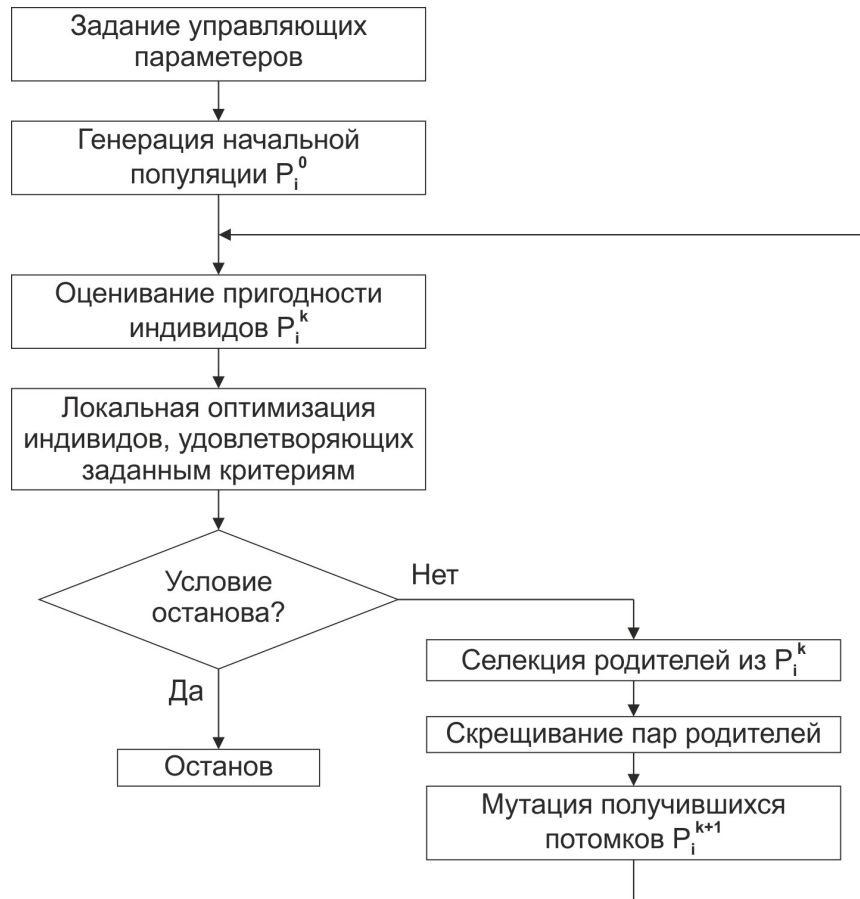


Рисунок 18 – Схема работы индивидуального процесса ГА

При **мутации** каждый из индивидов подвергается следующему алгоритму:

Для каждого параметра V_i выполнить :

$$\begin{cases}
 \text{если } (rand() * L / k_{mut} \leq 1) \text{ то} \\
 \quad V_i := RandNorm(V_i, B_i / 4) \\
 \text{пока } (V_i < B_i^{low}) \text{ то } V_i := V_i + B_i \\
 \text{пока } (V_i \geq B_i^{up}) \text{ то } V_i := V_i - B_i \\
 \}
 \end{cases}$$

Здесь V_i – это каждый из параметров в строке-хромосоме; L – общая длина строки (количество искомых параметров); k_{mut} – коэффициент мутации (задаётся в настройках ГА, обычно имеет значение от 0,5 – низкая мутация, до 1,5 – высокая мутация); $rand()$ – функция, возвращающая вещественное случайное число в диапазоне от 0 до 1; $RandNorm()$ – функция, возвращающая

ющая случайную величину по закону нормального распределения с заданными мат. ожиданием (первый параметр) и стандартным отклонением (второй параметр); V_i – диапазон варьирования параметра V_i ; V_i^{low} – нижняя граница диапазона варьирования параметра V_i ; V_i^{up} – верхняя граница диапазона варьирования параметра V_i .

Селекция (отбор) индивидов для скрещивания осуществляется по следующему алгоритму:

- 1) случайным образом выбираются `tourn_size` индивидов из популяции;
- 2) из отобранных индивидов выбирается лучший.

Здесь `tourn_size` – величина турнира, которая задаётся в настройках ГА. Размер турнира обычно берётся равным 3, 5 или 7.

Скрещивание индивидов осуществляется по алгоритму:

- 1) оператором селекции из популяции выбираются два индивида – «родителей»;
- 2) если (`rand() > kcross`), то в качестве потомка копируется один из родителей;
- 3) иначе, случайно выбирается точка разрыва на границе параметров строки-хромосомы;
- 4) разрываются оба индивида в этой точке;
- 5) полученные «хвосты» индивидов переставляются местами;
- 6) случайно выбирается один из двух полученных таким образом индивидов.

Здесь k_{cross} – коэффициент скрещивания, который задаётся в настройках ГА. Обычно он устанавливается равным 0,75.

2.1.3 Алгоритмы штрафования генерируемых структурных моделей

2.1.3.1 Функция штрафования за нарушение минимальных межатомных расстояний

В МПГА дополнительным критерием пригодности сгенерированной структуры является оценка того, насколько она удовлетворяет критериям адекватных длин межатомных связей. А именно, для каждой пары атомов задаются минимальные межатомные расстояния, которые исследователь считает допустимыми.

Для каждой из сгенерированных структур вычисляется сумма относительных нарушений минимальных межатомных расстояний по формуле:

$$AB = \sum_{p=1}^n \frac{(D_{p_min} - D_{p_calc})}{D_{p_min}}, \text{ при } D_{p_calc} < D_{p_min},$$

где AB – величина относительного нарушения межатомных расстояний для данной структуры (сокращение от термина *anti-bump*); n – количество имеющихся пар атомов; D_{p_min} – минимальное значение для каждой пары атомов; D_{p_calc} – вычисленное значение межатомного расстояния для данной пары атомов в данной структуре.

Затем значение функции пригодности корректируется в соответствии с этим значением по формуле:

$$Fitness = R_{wp} + AB \cdot coef \cdot R_{wp},$$

где R_{wp} – значение взвешенного профильного R-фактора; AB – значение суммы из предыдущей формулы; $coef$ – «коэффициент штрафа при сближении» из таблицы 4 (раздел 2.3.2 Задание параметров процесса МПГА).

2.1.3.2 Функция штрафования на количество межатомных связей

Пользователь может задать требуемое количество межатомных связей, которые должны иметь атомы данного типа с атомами других типов. Эта информация задается в виде таблицы, пример которой приведён в таблице 3.

Таблица 3 – Пример задания для штрафования на количество межатомных связей

Атом А	Атом Б	Кол-во связей	Критерий
N	Pd	1	1
N	C	1	1
C	Pd	0	1
Br	Pd	1	1
Pd	X	4	1

В первой строке записывается количество последующих правил штрафования. Далее описываются эти правила. Каждое правило (каждая строка) состоит из 4-х элементов:

- 1) Атом А – для какого типа атомов действует данное правило.
- 2) Атом Б – с атомами какого типа будет задано ограничение.
- 3) Количество связей, которые каждый атома типа А должен иметь с атомами типа Б.
- 4) Критерий штрафования. Если стоит 1 – то количество связей должно быть равно третьей позиции, если стоит 0 – то количество связей должно быть не меньше, чем указано в третьей позиции.

Например, правило «N - Pd - 1 - 1» означает, что каждый атом типа N должен иметь ровно 1 связь с атомами типа Pd. Правило «Pd - X - 4 - 1» означает, что каждый атом палладия должен иметь ровно 4 связи с другими атомами, независимо от их типа. А правило «B - O - 3 - 0» означает, что каждый атом типа B должен иметь не менее трёх связей с атомами типа O.

Вычисление того, связаны атомы или нет, происходит на основе определения перекрытия ковалентных радиусов атомов: если расстояние между атомами больше суммы их ковалентных радиусов, то программа считает, что между ними нет химической связи. Значения ковалентных

радиусов взяты из [127]. Каждое нарушение количества межатомных связей, заданных таким образом, увеличивает значение штрафа на $coef_2$ ($coef_2 \sim 0,1$ или 0). Соответственно, с учётом вклада этого вида штрафа, формула вычисления пригодности структурной модели приобретает вид:

$$Fitness = R_{wp} + (AB + coef_2 \cdot BC) \cdot coef \cdot R_{wp},$$

где BC – значение суммы нарушений заданного количества межатомных связей (сокращение от английского «bonds count»).

В текущей реализации – это грубый экспериментальный критерий, требующий введения дополнительных условий на типы возможных химических связей для данной пары атомов. Однако и в таком исполнении он позволяет сужать пространство поиска для некоторых структур, что положительно сказывается на эффективности работы МПГА.

2.1.4 Локальная оптимизация индивидов в популяциях МПГА

При изучении механизма сходимости ГА в [128] показано, что уточнение структурных моделей статистически обеспечивает установку отдельных атомов в истинные позиции в структуре (с понижением R-фактора) и является важным фактором ускорения сходимости. Поэтому лучшие, в смысле R-фактора, структурные модели с каждого эволюционного цикла ГА автоматически уточняются по атомным координатам с помощью встроенного метода полнопрофильного анализа и добавляются в популяцию моделей следующего поколения в случае понижения R-фактора. Следует отметить, что для ускорения и синхронизации межпопуляционной эволюции здесь исключен использовавшийся ранее и хорошо зарекомендовавший себя, но вычислительноемкий 2-й уровень ГА [3], оптимизирующий выбор уточняемых структурных параметров для полнопрофильного уточнения. Вместо него включен менее эффективный, но в десятки раз более быстрый алгоритм уточнения по МНК сразу всех параметров (указанных в задании на поиск, см. рис. 20). Уточнение может выполняться либо с помощью метода Ритвельда (из объектной библиотеки оптимизированных программ ObjCryst++ [62]), либо метода МПР (реже, т.к. работает почти на порядок дольше из-за вычисления в МПР производных). При этом, к структурным параметрам могут быть добавлены профильные для их совместного уточнения.

Локальная оптимизация может установить атомы разного сорта в позиции друг друга, если порядковые номера этих атомов в периодической таблице близки (рядом). Для компенсации этого эффекта в МПГА имеется механизм случайной замены атомов. После расчёта R-факторов вновь сгенерированной популяции и после выборочной локальной оптимизации части индивидов, может запускаться цикл, который для части индивидов (\sim для 20% популяции - для экономии времени) меняет два разнородных атома местами и проверяет, не уменьшился ли R-

фактор. Если R-фактор стал меньше, то данные об этой структурной модели перезаписываются. Аналогичный алгоритм перестановки атомов используется в программе FOX [62]. Однако там перестановка выполняется не для 20%, а для 2% популяции, и вместо мутации.

2.2 Схема взаимодействия индивидуальных ГА в МПГА

Схема МПГА изображена на рисунке 19. Предварительно, по встроенному в МПГА методу Ле Бэйла выполняется декомпозиция дифрактограммы с определением профильных параметров, таких как форма и ширина дифракционных линий и т.п. (эта процедура описана в гл. 2.1.1), которые затем фиксируются.

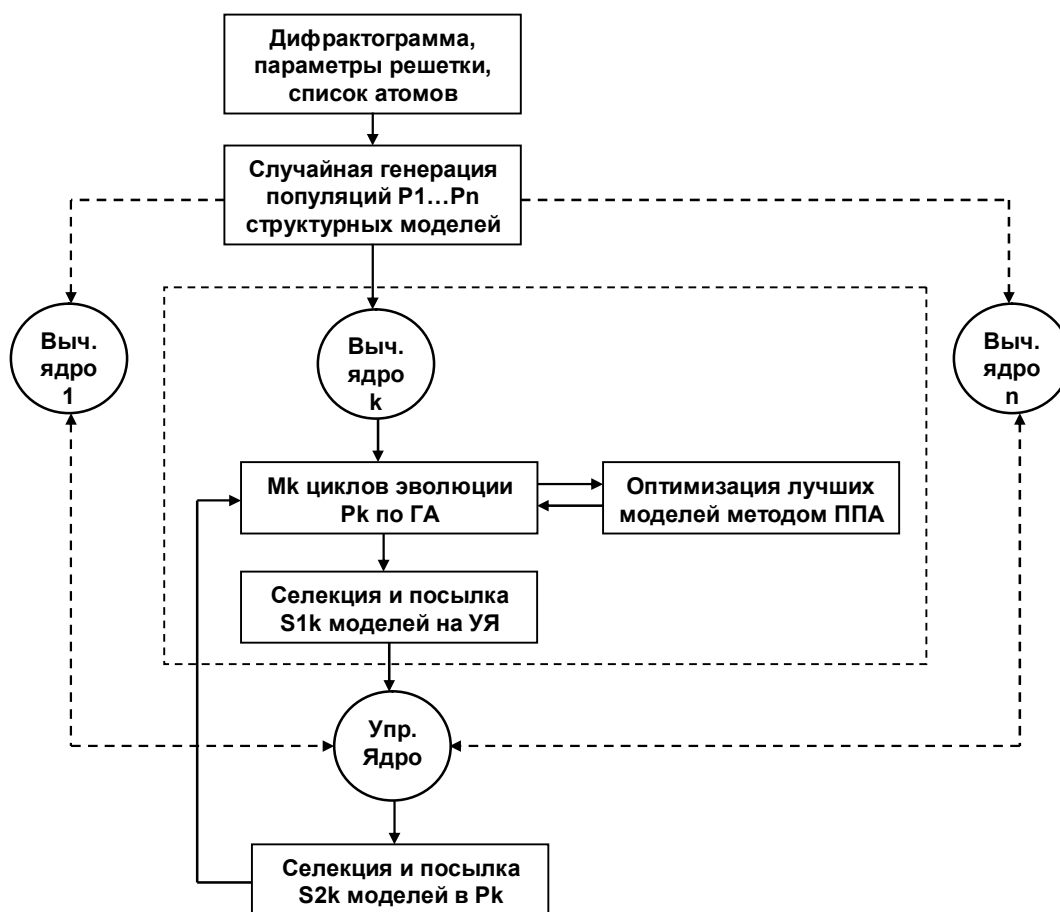


Рисунок 19 – Схема мультипопуляционного параллельного генетического алгоритма

Далее МПГА выполняет случайную генерацию n различных популяций структурных моделей вещества, и на каждом из n вычислительных ядер ПК или кластера запускает индивидуальный однопопуляционный ГА. Лучшие, в смысле R-фактора, структурные модели с каждого эволюционного цикла ГА уточняются с помощью метода полнопрофильного анализа и добавляются в популяцию моделей следующего поколения.

Основной операцией МПГА является случайная миграция структурных моделей между популяциями, эволюционирующими на разных вычислительных ядрах. Для этой цели на управляющее ядро (УЯ) с вычислительных ядер передаются по одной структурной модели с самым низким R-фактором (и, возможно, еще несколько случайно выбранных с R-фактором ниже глобального среднего, всего S1 штук). Все эти модели запоминаются и накапливаются на УЯ. На каждом поколении на УЯ производится оценка глобального среднего значения R-фактора накопленных структурных моделей. Периодически, через заданное число поколений (обычно, через 3), УЯ рассылает на вычислительные ядра S2 моделей (обычно от 1 до 3 штук), случайно отобранных из накопленных, у которых R-фактор ниже глобального среднего значения. Период рассылки и количество отправляемых с УЯ структурных моделей задаётся пользователем в настройках алгоритма.

На процесс миграции также оказывает влияние реализованный в МПГА временный элитизм. Его суть в том, чтобы сохранять на рабочих ядрах хорошие решения от разрушения операторами скрещивания и мутации в ходе нескольких поколений эволюции (количество поколений элитизма обычно задаётся равным 3 или 5). Так, если на очередном поколении рабочее ядро ГА находит решение с более низким R-фактором, чем на предыдущем поколении, то это решение запоминается в отдельной ячейке памяти и заданное количество раз передаётся в популяцию следующего поколения без изменений. Если на следующем поколении найдена структура с ещё более низким R-фактором, то тогда она становится «элитной» и передаётся в следующие поколения. Механизм элитизма также способствует тому, чтобы на управляющем ядре накапливались структуры с более низким R-фактором, чем без использования элитизма.

При достижении одной из структурных моделей на УЯ целевого значения R-фактора (определенного в результате декомпозиция дифрактограммы) МПГА завершает работу, а модель может подвергаться заключительному полнопрофильному уточнению по всем профильным и структурным параметрам.

После завершения вычислений МПГА сохраняет данные о найденной структурной модели в файл. Затем исследователь может провести кристаллохимическую верификацию структуры: оценку корректности межатомных расстояний, валентных углов и т.п.

2.3 Программное обеспечение, реализующее алгоритм МПГА

Описанный выше алгоритм МПГА реализован в двух компьютерных программах. Первая – реализует непосредственно сам алгоритм МПГА. Программа написана на языке C++ и имеет кроссплатформенную архитектуру, что позволяет запускать её на операционных системах Microsoft Windows и Unix. Вторая программа – MPGА_Control – реализует интерфейс управления МПГА. Она написана на языке программирования Delphi и предназначена для работы в ОС Microsoft Windows. MPGА_Control позволяет в удобной форме задавать параметры работы МПГА,

и просматривать результаты работы МПГА, запоминаемые на УЯ и индивидуальных ядрах. Это также обеспечивает задание с ПК параметров работы на МПГА кластере и визуально-графическую оценку результатов этой работы.

2.3.1 Задание профильных и структурных параметров

На рисунке 20 представлен пример интерфейса МПГА. В его верхней части представлена форма МПГА для установки профильных параметров и параметров решетки, внизу – форма для установки структурных параметров: координат атомов (X, Y, Z), их изотропных или анизотропных тепловых факторов (V_{iso} или V_{aniso}) и коэффициентов заполнения атомами кристаллографических позиций (N). Визуализирован список всех параметров, которые могут использоваться в методе Ритвельда, справа от каждого параметра расположена управляющая кнопка. Ее нажатие приводит к включению данного параметра в индивиды ГА, и кнопка при этом становится зеленой (повторное нажатие исключает параметр, и кнопка становится серой). Возможно, также устанавливать группы функционально связанных параметров – кнопки коричневого цвета с номером группы (например, сверху на рисунке 20 для задания равенства осей ячейки a и b). Это может использоваться и для задания жестко связанных атомных фрагментов.

Для выбора режима поиска и его запуска используется подменю «ГА» в строке меню (сверху). Существует возможность интерактивного полнопрофильного доуточнения полученной по МПГА структурной модели (кнопка с голубым треугольником ниже строки меню), при этом для выбора его режимов и контроля R-фактора визуализируется отдельная форма (внизу справа). Найденные при полнопрофильном уточнении в процессе МПГА (или после доуточнения) значения параметров структурных моделей динамически устанавливаются в соответствующих им окнах интерфейса. При наведении курсора на окна и кнопки визуализируются подсказки с их описанием и назначением.

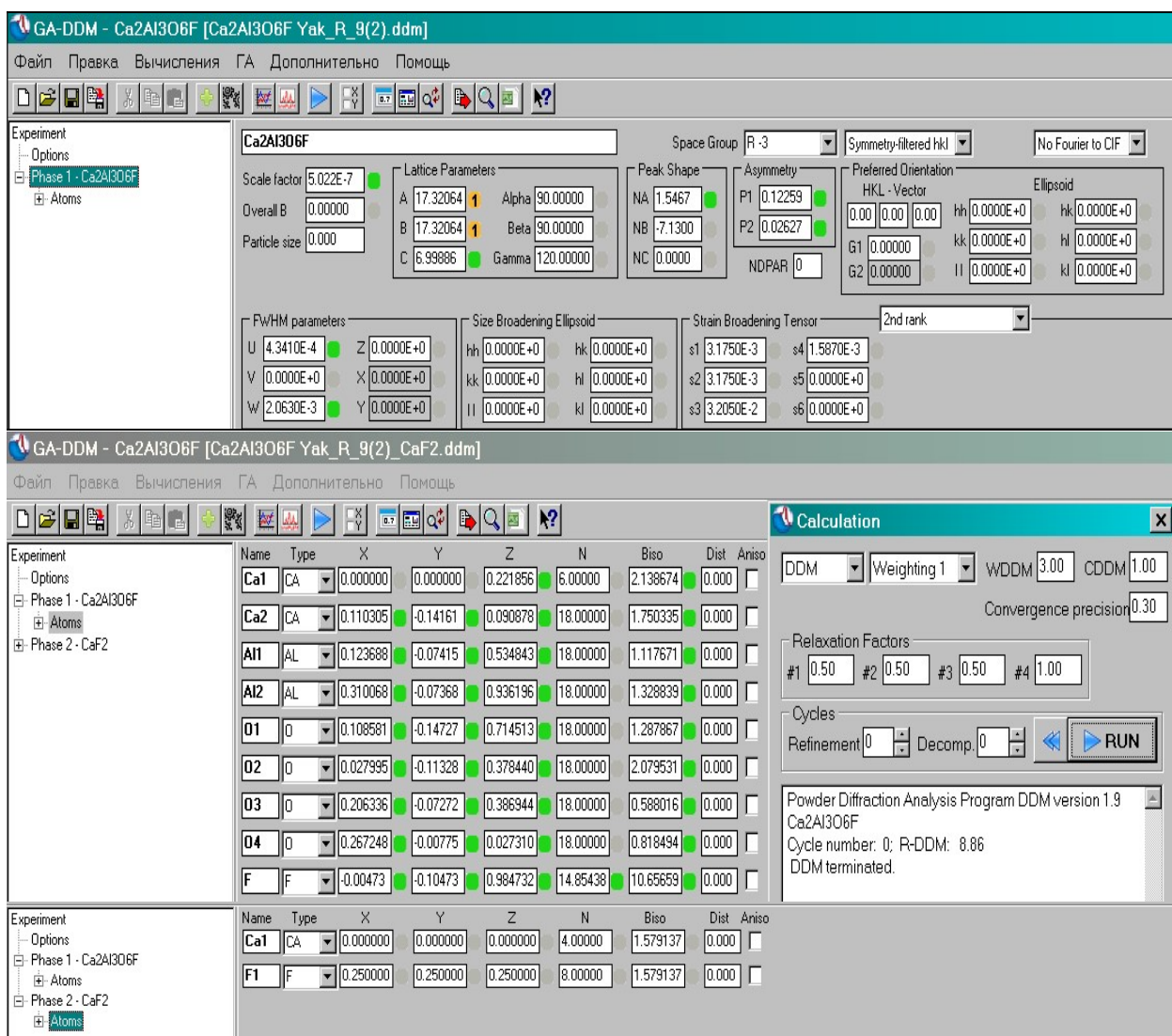


Рисунок 20 – Пример интерфейса МПГА с заданием на поиск структуры $\text{Ca}_2\text{Al}_3\text{O}_6\text{F}$; сверху – форма для профильных, внизу – для структурных параметров (с учетом примесной фазы флюорита); зелеными кнопками отмечены искомые параметры; рядом в окнах найденные по МПГА значения

2.3.2 Задание параметров процесса МПГА

После выбора в подменю «ГА» режима «МПГА» визуализируется форма управления МПГА, приведенная на рисунке 21. Пояснения к параметрам МПГА приведены в таблице 4. На форме задаются общие параметры генетических операций (слева сверху), индивидуальные параметры генетических операций на конкретных ядрах (справа сверху) и ограничения на параметры структурных моделей (слева внизу).

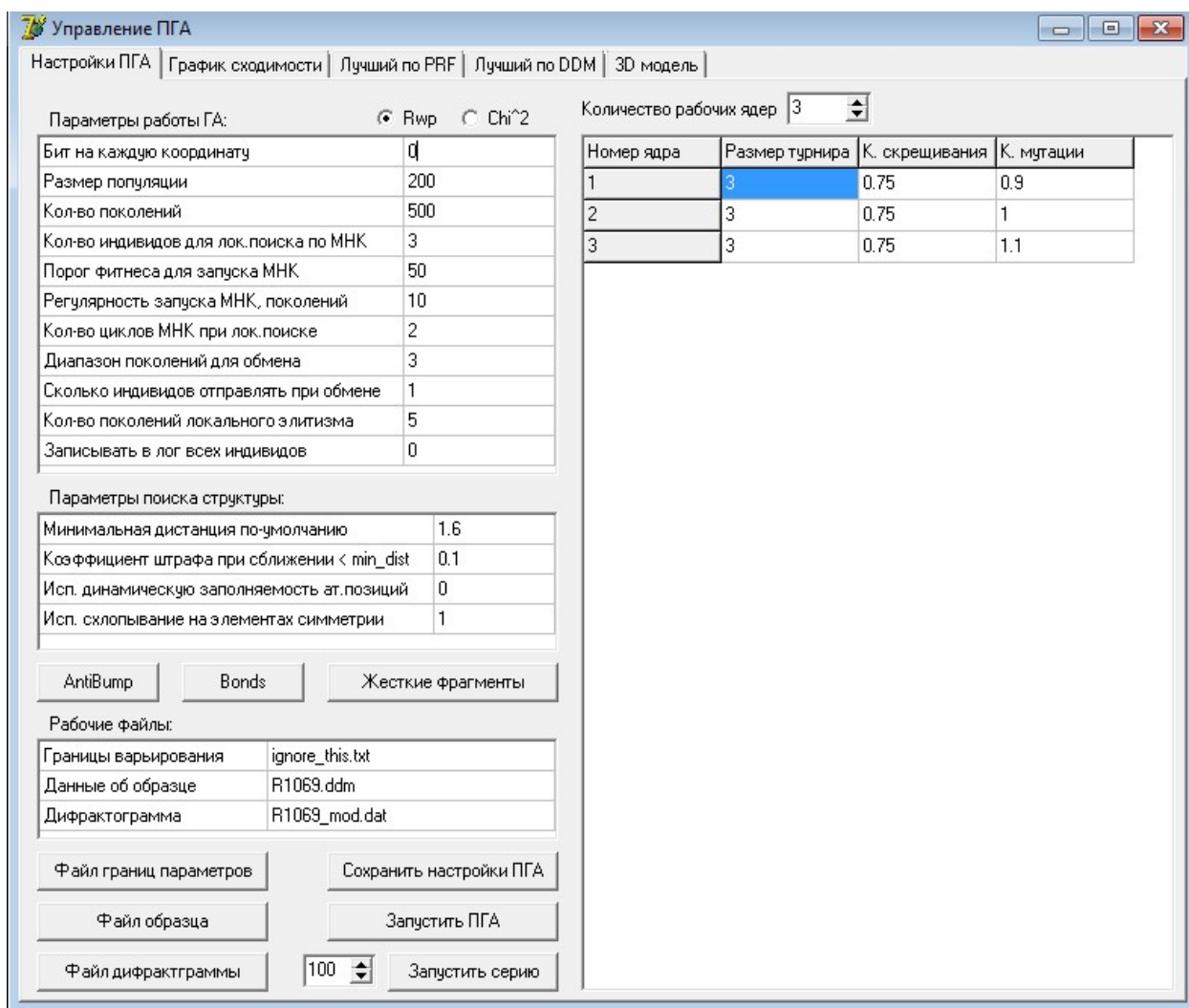


Рисунок 21 – Окно задания параметров МПГА

Таблица 4 – Описание параметров поиска кристаллической структуры в МПГА

Обозначение параметра	Типичное значение	Описание параметра
Бит на каждую координату	5 или 0 (0 - для запуска вещественного ПГА)	Кол-во бит для кодирования каждого параметра – для варианта бинарного ГА. В варианте МПГА с вещественным кодированием этот параметр не используется.
Размеры популяций индивидуальных ГА	200	Должен превышать количество степеней свободы структурных параметров в несколько раз.
Количество поколений	500	Чем больше, тем лучше; ограничено только общим (разумным) временем вычислений.
Количество индивидов для локального поиска по МНК	3	Лок. поиску будет подвергаться такое количество индивидов из каждой популяции, для которых значение функции пригодности ниже порогового (см. след. параметр). Если это число больше 1, то лок. поиску будет подвергнут 1 лучший индивид, и остальные – случайные, в пределах порогового значения фитнеса.
Порог фитнеса для запуска МНК	80	Значение функции пригодности (fitness – приспособленность), ниже которого индивид имеет шанс подвергнуться локальному поиску
Регулярность запуска МНК	1 - 5	Раз в сколько поколений будет запускаться МНК

Количество циклов МНК при локальном поиске	2	В МПГА встроен алгоритм усиления влияния МНК при снижении значения функции пригодности. Так, при достижении фитнеса в $\frac{3}{4}$ от порогового, количество циклов МНК увеличивается вдвое, при $\frac{1}{2}$ от порогового значения – втрое, при $\frac{1}{4}$ от порогового значения – вчетверо.
Диапазон поколений обмена (частота рассылки с УЯ на ядра)	3	С какой регулярностью управляющее ядро будет рассылать лучшие индивиды на вычислительные ядра
Сколько индивидов отправлять при обмене (с УЯ на вычислительные ядра)	1	Управляющее ядро накапливает лучшие решения со всех ядер за все поколения. При рассылке индивидов с УЯ на вычислительные ядра выбирается заданное количество случайных индивидов из накопленных – от самого лучшего, до глобального среднего по фитнесу из лучших накопленных за все предыдущие поколения.
Количество поколений локального элитизма	3 - 5	В течение заданного количества поколений рабочие ядра гарантированно не будут терять лучшее из найденных ими решений.
Записывать в лог всех индивидов	0	Это служебная функция, позволяющая сохранять в текстовые файлы данные обо всех сгенерированных в процессе работы ГА индивидов, при необходимости.
Минимальная дистанция по-умолчанию	1,0	Минимальная дистанция между атомами, при нарушении которой на найденную структуру будет накладываться пропорциональный штраф. Это значение имеет влияние, когда не задано минимальное меж-атомное расстояние для атомов отдельных типов (подробнее смотри в разделе 2.2.1 Функция штрафования за сближения атомов).
Коэффициент штрафа при сближении	0,2 - 0,6	Чем больше это значение, тем сильнее на ГА будет влиять размер штрафа межатомных расстояний и количества связей. Чем оно меньше, тем больше будет влиять профильный R-фактор.
Использовать динамическую заполняемость атомных позиций	1	Если два однотипных атома будут сближаться друг с другом, и если эта функция включена, то заполняемость атомной позиции каждого из них будет постепенно снижаться, и при полном сближении будет равна 0,5 у каждого. Это нужно в случае, если мы не знаем точного количества атомов в ячейке, и указываем атомов больше, чем в истинном решении.
Использовать схлопывание на элементах симметрии	1	Если атом сближается сам с собой возле какого-то элемента симметрии, то, если эта функция включена, программа попытается принудительно посадить его на этот элемент симметрии.
Файл с границами варьирования	Sample_bounds.txt	Имя файла с границами варьирования уточняемых параметров. Если уточняются лишь атомные координаты и заполняемость атомной позиции, то этот файл не нужен, границы этих параметров выбираются программой автоматически.
Файл с данными об образце	Sample.ddm	
Файл с дифрактограммой	Sample.dat	

Для МПГА задаётся количество рабочих ядер (оно должно быть на 1 меньше, чем количество используемых вычислительных ядер процессора, так как одно из них будет управляющим).

И устанавливаются коэффициенты скрещивания, мутации и размер турнира для каждого из рабочих ядер.

Кроме того, имеется возможность запустить серию последовательно выполняемых независимых процессов МПГА (т.к. из-за стохастичности ГА не каждый запуск МПГА может сойтись к истинной структуре). Для этого служит кнопка «Запустить серию», рядом с которой можно указать количество запусков в серии.

2.3.3 Работа с молекулами, заданными в виде жестких фрагментов

В МПГА имеется возможность задавать молекулу в виде жесткого фрагмента, который можно перемещать внутри ячейки и вращать в процессе поиска. Ориентация фрагмента задается при помощи кватерниона из 4-х параметров. На рисунке 22 приведено окно задания фрагментов. Координаты атомов в жестком фрагменте задаются в ангстремах относительно геометрического центра фрагмента. Вместе со списком атомов в этом же окне рисуется и графическое представление редактируемого фрагмента (для наглядности его можно вращать при помощи курсора).

The screenshot shows the 'FragmentsForm' window. At the top, there is a dropdown menu for 'Выбор фрагмента:' set to 'SnO4' and a 'Добавить фрагмент' button. Below are input fields for X, Y, Z coordinates and Q0, Q1, Q2, Q3 quaternions, each with a 'R' checkbox. There is also an 'Occupancy' field with a 'R' checkbox and a 'Добавить атом' button. At the bottom, there are buttons for 'Сместить в геом.центр', 'Сместить центр к выделенному атому', and 'Удалить атом'. A table lists the atoms with their titles, types, and coordinates. To the right is a 3D visualization of the SnO4 tetrahedron with a central Sn atom and four O atoms.

Title	Type	X, A	Y, A	Z, A	B_iso
Sn1	Sn	0	0	0	1.5
O1	O	1.60590505599	0.99117970466	-0.5152928829	1.5
O2	O	0.57928740978	-1.5977206230	0.97140359878	1.5
O3	O	-1.2167828083	-0.3844114542	-1.4628195762	1.5
O4	O	-0.9919385910	1.1867649741	1.2170010805	1.5

Рисунок 22 – Окно работы с параметрами жесткого фрагмента на примере тетраэдра SnO₄

2.3.4 Контроль за процессом работы программы МПГА

Для запуска процесса поиска параметров по МПГА в ОС Microsoft Windows нужно нажать на кнопку «Запуск ПГА». Чтобы запустить МПГА в ОС Unix, необходимо нажать на кнопку «Сохранить настройки ПГА», и затем скопировать рабочую папку программы на компьютер с ОС Unix и запустить там исполняемый файл процесса МПГА. При первом запуске МПГА с данным образцом будет автоматически проведён поиск профильных параметров при помощи

процедуры декомпозиции Ле Бейла (пример приведён на рисунке 23). Найденные профильные параметры будут сохранены в специальный файл, и при последующих запусках программа будет считывать их из этого файла.

```

C:\Program Files (x86)\MPICH2\bin\mpiexec.exe
>> File StartStructure.xml not found
>> File StartStructure.xml not found
>> File StartStructure.xml not found
>> File StartStructure.xml not found
Imported powder pattern: 3385 points, 2theta= 3.013 -> 90.997, step= 0.026
Imported powder pattern: 3385 points, 2theta= 3.013 -> 90.997, step= 0.026
Imported powder pattern: 3385 points, 2theta= 3.013 -> 90.997, step= 0.026
Imported powder pattern: 3385 points, 2theta= 3.013 -> 90.997, step= 0.026
> Wave length = 1.54188
> Wave length = 1.54188
> Wave length = 1.54188
> Wave length = 1.54188
> Le Bail Extraction - Step 1
> Le Bail Extraction - Step 1
> Le Bail Extraction - Step 1
> Le Bail Extraction - Step 1
> Target Rw = 8.79623
> Le Bail Extraction - Step 2
> Target Rw = 8.79623
> Le Bail Extraction - Step 2
> Target Rw = 8.79623
> Le Bail Extraction - Step 2
> Target Rw = 8.79623
> Le Bail Extraction - Step 2
> Target Rw = 8.79623
> Le Bail Extraction - Step 2
> Target Rw = 8.08303
  
```

Рисунок 23 – Отображение процесса профильной декомпозиции в окне МПГА при работе в 4-процессном режиме

В процессе работы МПГА можно наблюдать за ходом расчётов при помощи встроенных инструментов: график сходимости по значению пригодности индивидов (рисунок 24), график разности между лучшей найденной структурой и экспериментом (рисунок 25), просмотр 3D модели лучшей найденной структуры (рисунок 26).

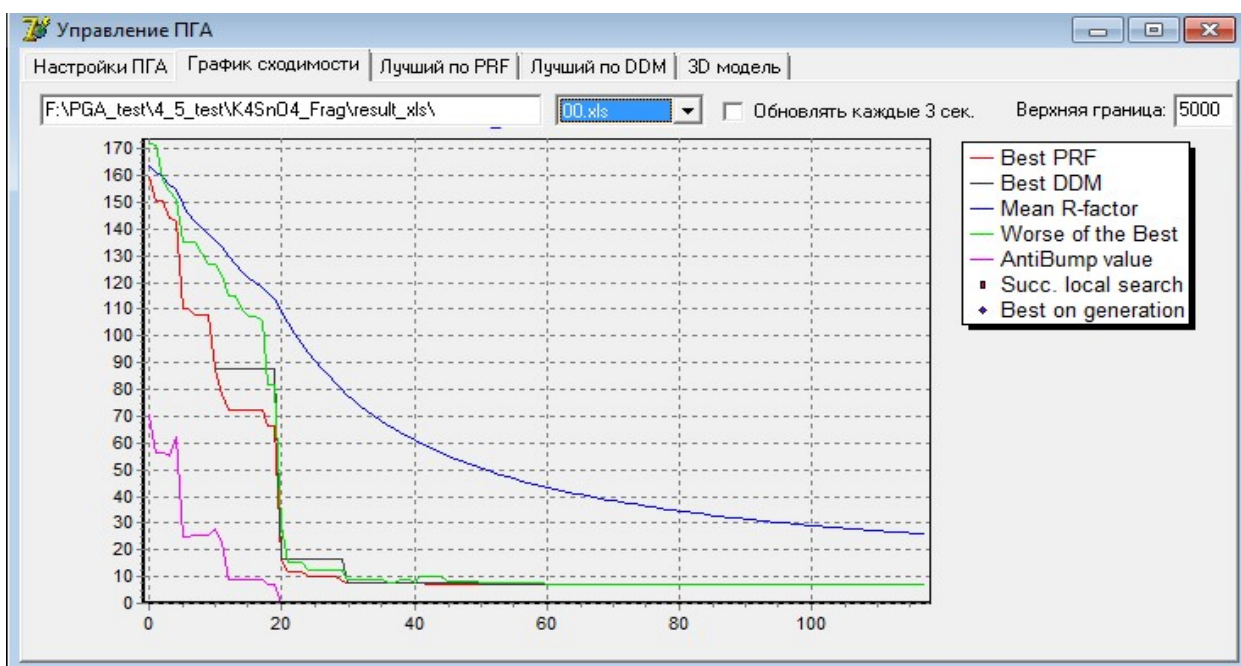


Рисунок 24 – Пример графика сходимости процесса МПГА при поиске структуры K_4SnO_4

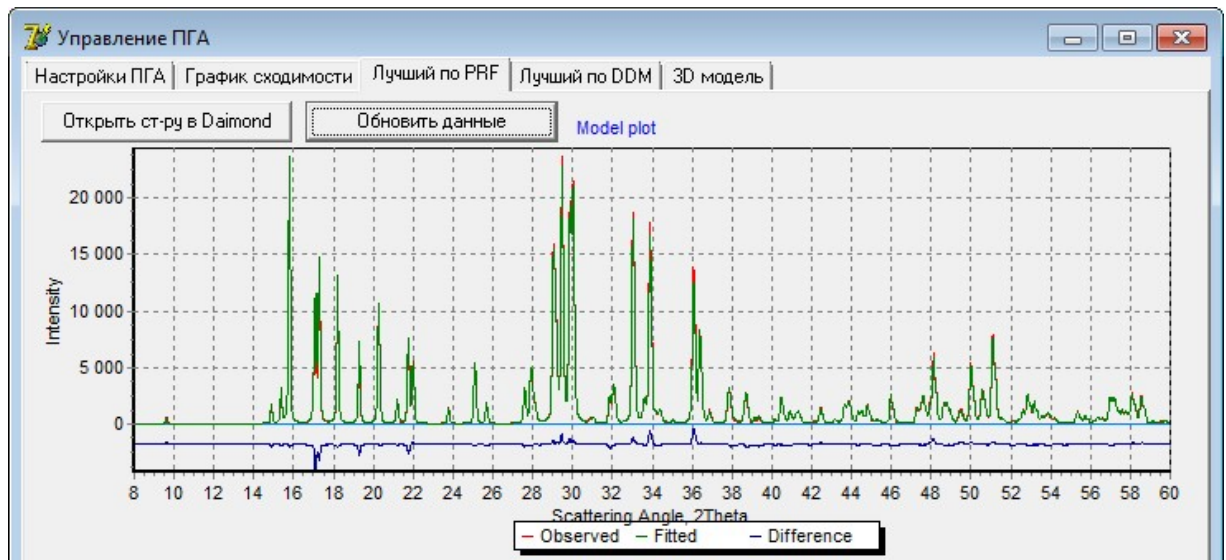


Рисунок 25 – График разности между лучшей найденной структурой и экспериментом.

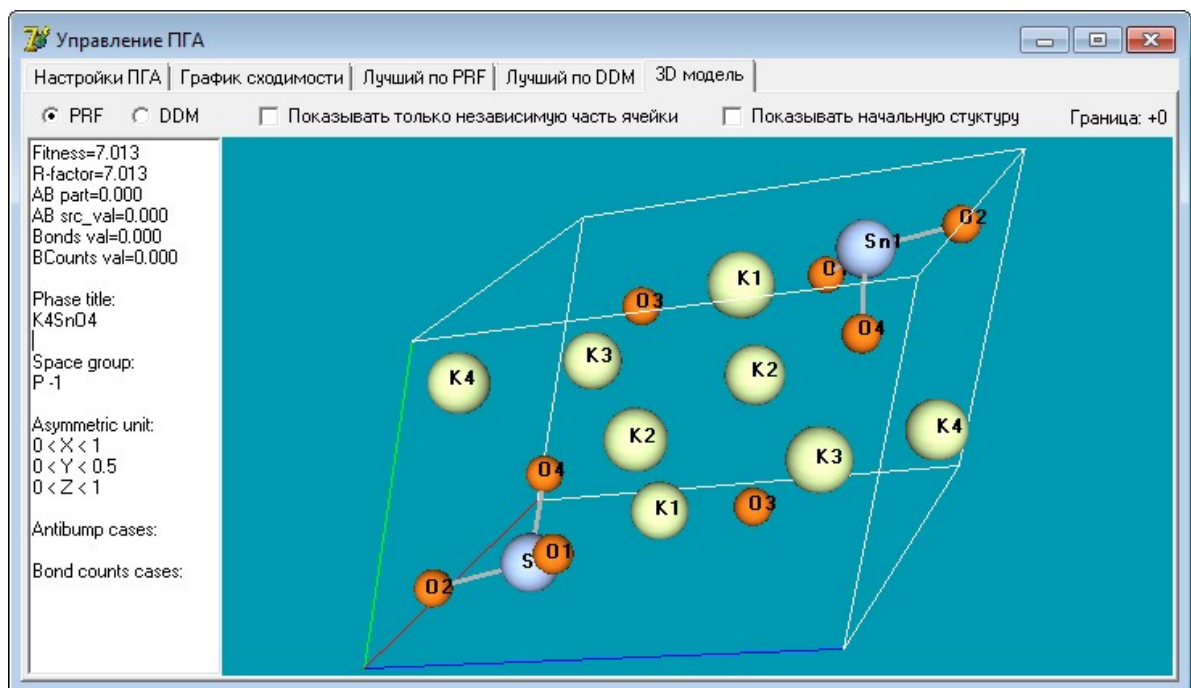


Рисунок 26 – Окно просмотра 3D модели лучшей найденной структуры

2.3.5 Сохранение результатов работы МПГА

Все данные о ходе вычисления автоматически сохраняются в папке, из которой был произведён запуск. Результаты работы МПГА сохраняются в трёх форматах:

- BestSolve.ddm – для верификации и доуточнения в программе DDM.
- BestSolve.xml – для верификации и доуточнения в программе FOX.
- BestSolve.cif – для просмотра кристаллической структуры в программе Diamond.

2.4 Результаты апробации МПГА на тестовых структурах

2.4.1 Описание процесса тестирования МПГА

Важными характеристиками любого стохастического метода структурного анализа являются скорость и надежность сходимости, и сложность определяемых структур. Надежность выражает долю успешных пусков алгоритма, а сложность – количество независимо определяемых структурных параметров (степеней свободы структуры). Для статистической оценки этих характеристик был выполнен многократный поиск по МПГА хорошо известных структур химических соединений различной степени сложности, указанных в таблице 5. Выполнялся поиск координат атомов в общих позициях, температурные факторы были заданы априори, т.к. на практике они могут быть первоначально взяты из известных, близких по составу и кристаллическому строению структур и уточнены на заключительных этапах поиска.

Таблица 5 – Список тестовых кристаллических структур для МПГА

Химическая формула соединения	Параметры решетки; простр. группа симметрии; число формульных единиц Z в ячейке	Число атомов в независимой части ячейки	Число степеней свободы атомных координат	R _{wp} -фактор структуры, найденной по МПГА и уточненной по МПР	Целевой R _{wp} -фактор из процедуры Ле Бэйла
Ba ₂ CrO ₄	a= 7.671Å, b= 10.391Å, c= 5.891Å; P n a 21; Z=4;	7	21	0.79	0.52
K ₄ SnO ₄	a=6,48Å, b=6,51Å, c=9,70Å, α=71,82°, β=99,89°, γ=113,13°; P -1; Z=2;	9	27	0,78	0.50
K ₂ PbO ₂	a=10,9Å, b=7,6Å, c=7,32Å, α=119,3°, β=88,4°, γ=117,7°; P -1; Z=2;	10	30	3.2	2.23
Ca ₂ Al ₃ O ₆ F *	a=7,3205Å, c=6,9988Å; R -3; Z=12;	9	25	8.93	8.51
[Pd(NH ₃) ₄](C ₂ O ₄)	a= 7.081Å, b= 7.081Å, c=3.801Å, α= 91.91°, β= 98.665°, γ= 97.283°; P -1; Z=2;	6	15	5.93	4.54
Pd(NH ₃) ₂ (NO ₂) ₂	a= 5.003Å, b= 5.42Å, c= 6.32Å, α= 91.384°, β= 111.81°, γ= 100.358°; P -1; Z=2;	8	14	5.42	4.16
Rh(NH ₃) ₃ (NO ₃) ₃	a= 7.464Å, c= 5.225Å; P 3; Z=3;	10	25	6.97	5.77
Er ₁₀ W ₂ O ₂₁	a= 15.82Å, b= 10.481Å, c= 10.541Å; P b c n; Z=4;	18	54	3.32	2.9
EuNdCuS ₃	a= 11.044Å, b= 4.066Å, c= 11.415Å; P n m a; Z=4;	6	12	15.9	13.21

CaC ₄ H ₄ O ₆ ·4H ₂ O**	a= 6.249Å, b= 8.224Å, c= 10.438Å, α= 94.93°, β= 105.97°, γ= 107.52°; P-1; Z=2;	15	22	8.02	7.4
Bi ₃ FeO ₁₁ Sb ₂ ***	a = 9.51521 Å; Pn-3; Z=4	7	25	6.98	5.76

*) из работы [129] по дифрактограмме, представленной авторами;

**) из конкурса SDPDRR-3 [3], с заданием молекулы тартрата C₄H₄O₆ как жесткого фрагмента;

***) из работы [130] по дифрактограмме синхротронного излучения, представленной авторами.

Первые 3 структуры и структура Er₁₀W₂O₂₁ определялись по их расчетным и зашумленным дифрактограммам, имитирующим экспериментальные. При этом использовались профильные параметры, заданные при расчете. Целью исследования было изучение «чистой» эволюционной сходимости координат атомов в структурах различной степени сложности, без влияния на сходимость МПГА погрешностей, вносимых за счет определения профильных параметров по экспериментальным дифрактограммам. В результате были определены наиболее подходящие вышеуказанные установки генетических операторов. Остальные структуры определялись по экспериментальным дифрактограммам веществ, снятым с шагом 2 тета = 0,02 град. в монохроматизированном CuKα излучении. При определении последней структуры использовалась экспериментальная дифрактограмма из [3].

С целью получения наиболее объективных характеристик по надежности и скорости сходимости МПГА для поиска некоторых из структур было выполнено по 100 независимых пусков экспериментальной программы МПГА на 4-х ядерной ПК. Дополнительно для сравнительной оценки сходимости МПГА и однопопуляционного ГА в программе было выполнено по 20 пусков поиска каждой из структур с отключенным режимом пересылки структурных моделей с УЯ на вычислительные ядра (в этом случае на ядрах выполняются независимые однопопуляционные ГА). При этом в каждом пуске программы на 3-х вычислительных ядрах выполнялся независимый однопопуляционный ГА, так что в 20 пусках было получено по 60 независимых результатов.

Число структурных моделей, отбираемых для миграции, составляло 2% от размера популяций (пояснения по параметрам миграции S1 и S2 смотри в разделе 2.2 Схема взаимодействия индивидуальных ГА в МПГА). Параметры генетических операций: турнирного парного скрещивания, мутации атомных координат и селекции структурных моделей в очередные поколения эволюции по R-фактору, были идентичны. Кроме того, использовалась операция временного элитизма – сохранения в популяции лучшей структуры на протяжении 3-х поколений. Характеристики поиска структур сведены в таблице 6.

Таблица 6 – Надежность поиска структур по МПГА на 4-х ядерной ПК и по ГА

Тестовая структура	Размер популяции; кол. поколений эволюции	Параметры миграции		Процент сходимости поиска, %	Среднее время сходимости, мин.	Средний R_{wp} -фактор и $\sigma(R)$ при сходимости
		S1	S2			
МПГА						
K_4SnO_4	50; 200	1	1	76	1,72	5,53 (0,01)
K_2PbO_2	100; 200	1	2	64	2,77	7,76 (0,03)
$Ca_2Al_3O_6F$	100; 300	1	2	27	4,21	7,09 (0,01)
Однопопуляционный ГА						
K_4SnO_4	50; 200	1	0	50	3,08	5,53 (0,001)
K_2PbO_2	100; 200	1	0	27	5,20	7,75 (0,001)
$Ca_2Al_3O_6F$	100; 300	1	0	6,7	4,50	7,09 (0,006)

Экспериментальная оценка эффективности от использования параллельных вычислений выполнена в результате усреднения характеристик многократных пусков поиска различных тестовых структур по МПГА на 4-х ядерном ПК, суперкомпьютерном кластере и в режиме однопопуляционного ГА (при отключении режима миграции). Оцениваемыми характеристиками эффективности являлись надежность и скорость поиска, и сложность определяемых структур. Результаты сведены в таблице 7, из которой можно сделать следующие выводы. Надежность (процент сходимости) структурного анализа по МПГА на 4-х ядерном ПК в 2 – 3 раза выше, чем в однопопуляционном ГА, а на кластере с использованием 24 ядер в 2 – 4 раза выше, чем на 4-х ядерном ПК. Среднее время работы программы (на данном ПК) наоборот уменьшается, примерно, в полтора – два раза. Главным же выводом является то, что увеличение числа вычислительных ядер позволяет определять более сложные структуры.

Таблица 7 – Сравнение характеристик эффективности определения структур по МПГА на 4-х ядерном ПК, вычислительном кластере и в однопопуляционном ГА

Тестовая структура	Число степеней свободы атомных координат	Размер популяции	Число пусков программы	Процент сходимости, %	Среднее время сходимости, мин.	Средний R_{wp} -фактор МПГА и $\sigma(R_{wp})$ при сходимости
Однопопуляционный ГА на ПК						
Ba_2CrO_4	21	50	100	70	2,0	2,13
K_4SnO_4	27	100	100	33	3,08	5,53 (0,001)
K_2PbO_2	30	100	100	27	5,2	7,75 (0,001)
$Ca_2Al_3O_6F$	25	300	50	8	19,6	11,15 (0,006)
$CaC_4H_4O_6 \cdot 4H_2O$	22	300	20	0		-
$Er_{10}W_2O_{21}$	54	500	20	0		-

МПГА на 4-х ядерном ПК						
Ba ₂ CrO ₄	21	50	100	100	1,2	2,1
K ₄ SnO ₄	27	100	100	76	1,8	5,53 (0,01)
K ₂ PbO ₂	30	100	100	64	2,8	7,76 (0,03)
Ca ₂ Al ₃ O ₆ F	25	300	50	28	12,5	11,15 (0,01)
CaC ₄ H ₄ O ₆ ·4H ₂ O	22	300	20	30	4,5	9.87
Er ₁₀ W ₂ O ₂₁	54	500	20	10	55,8	4.52
МПГА на кластере (24 ядра)						
K ₄ SnO ₄	27	100	20	100		5,53 (0,01)
CaC ₄ H ₄ O ₆ ·4H ₂ O	22	300	20	60		9.87
Er ₁₀ W ₂ O ₂₁	54	700	20	40		4.52

2.4.2 Обсуждение результатов тестирования МПГА

Основным недостатком методов глобальной оптимизации, включая ГА, является частая преждевременная сходимость и стагнация в локальных минимумах гиперповерхности функции фитнеса. Перспективным направлением является использование многопопуляционных параллельных генетических алгоритмов, которые выполняют обмены индивидов между разными популяциями, чтобы предотвратить стагнацию. Этот подход используется для различных задач глобальной оптимизации, а его эффективность зависит от размерности задачи оптимизации, специфической формы гиперповерхности и конкретной конструкции МПГА [131].

Первое приложение подхода МПГА для структурных исследований было сделано Хариссом в [132]. Этот вариант МПГА был создан на базе успешно действующей программы EAGER однопопуляционного ГА [94] и дополнен средствами случайного обмена структурными моделями непосредственно между разными популяциями ГА, эволюционирующих на разных узлах вычислительного кластера. На примере многократного определения тестовой структуры органического соединения, включающего поиск 13-ти структурных координат, показано, что 4-х популяционный МПГА с размером популяций из 100 структурных моделей обеспечил сходимость 42% (в 10 случаях из 24 пусков), а однопопуляционный ГА – 18% (в 3 случаях из 17 пусков) при одинаковых вычислительных ресурсах (на 20-ти ядерном вычислительном кластере). Т.е. было получено ускорение сходимости в ~ 2,4 раза относительно однопопуляционного ГА, что показало эффективность межпопуляционной миграции.

Однако, данный подход не получил достаточного развития, возможно потому, что для структуры с 13-ю степенями свободы сходимость не достаточна и заметно уступает широко используемым методам имитации отжига, особенно, при их работе на суперкомпьютерном кластере [132]. В частности, не были достаточно изучены особенности сходимости МПГА, увеличение скорости структурного поиска и возможность определения более сложных структур. Тем не менее, во многих работах [4, 134-136] отмечается перспективность развития именно генетических алгоритмов для структурного анализа.

Главным отличием нашего подхода к МПГА является более эффективная организация миграции. В частности, постоянное накопление на управляющем ядре МПГА лучших, в смысле R-фактора, структурных моделей из всех популяций, и обратная миграция моделей, выбираемых для этого среди лучшей половины моделей УЯ. Такая обратная миграция обеспечивает постоянное присутствие во всех популяциях структурных моделей с R-фактором лучше среднего. Для предотвращения преждевременного разрушения лучших моделей, мигрировавших с УЯ или собственных, вследствие неудачной мутации или скрещивания, введена также операция временного элитизма. Кроме того, установлено, что для предотвращения чрезмерного распространения одной из них в популяциях и обеспечения разнообразия структурных моделей необходимо использовать небольшой коэффициент миграции с УЯ (1-2% от размера популяции). Также установлено, что размеры популяций зависят от сложности искомой структуры и должны превышать количество степеней свободы структурных параметров в несколько раз. Число поколений эволюции должно быть выше размера популяций. Из данных таблицы 6 видно, что вариант МПГА на 4-х ядерной ПК при поиске структур с 25 – 30 степенями свободы структурных переменных показывает среднюю сходимость 56% при повышении сходимости МПГА относительно однопопуляционного ГА в 1,5 - 4 раза (при одинаковых вычислительных ресурсах). Рассмотрим некоторые типичные примеры *ab initio* поиска по МПГА тестовых структур, указанных в таблице 5.

Определение по МПГА структуры $Ca_2Al_3O_6F$

Кристаллическая структура $Ca_2Al_3O_6F$ (пр. гр. $R-3$, $a = 17.3237$, $c = 7.00017$, $Z = 6$) была недавно определена по данным порошковой дифракции с дополнительным использованием Рамановской спектроскопии [129]. Для решения и уточнения структуры была использована программа TOPAS 4.2. Модель структуры была найдена в 2 этапа с помощью процедуры имитации отжига. На первом определялись координаты двух Са и ориентация двух жестко заданных фрагментов - тетраэдров AlO_4 (т.е. 18 степеней свободы), был достигнут $R_{wp}=16.3\%$. На 2-м этапе определялись координаты фтора, другие атомы фиксировались, достигнут $R_{wp}=13.8\%$. Суммарная химическая формула $Ca_2Al_3O_6F$ требует оккупации 2/3 из 18 позиций атомов фтора (т.е. он разупорядочен). После финального уточнения по методу Ритвельда был достигнут $R_{wp} = 7.38\%$. Значения атомных параметров из [129] представлены в таблице 8. Эта структура при исследовании была принята нами за истинную.

Таблица 8 – Параметры атомов кристаллической структуры $\text{Ca}_2\text{Al}_3\text{O}_6\text{F}$, приведённые в статье авторов.

	x	y	z	B_{iso}	Occ.
Ca1	0.22316(10)	0.41486(10)	0.0760(2)	2.24(5)	1.0
Ca2	0.0	0.0	0.2771(4)	2.63(8)	1.0
Al1	0.59279(16)	0.45660(17)	0.2984(3)	1.85(6)	1.0
Al2	0.64325(17)	0.05015(16)	0.1027(4)	1.82(6)	1.0
O1	0.5943(3)	0.5403(3)	0.4458(6)	1.57(11)	1.0
O2	0.1400(3)	0.1113(3)	0.1173(6)	2.59(13)	1.0
O3	0.5582(3)	0.0773(3)	0.1175(6)	2.16(11)	1.0
O4	0.6009(3)	-0.0578(3)	0.1941(6)	1.98(12)	1.0
F	0.3411(8)	0.5681(6)	0.1799(12)	7.6(3)	0.6667

Для поиска по МПГА была использована рентгенограмма (Rigaku Ultima IV, $\text{CuK}\alpha$), предоставленная для его апробации авторами статьи. Искали одновременно координаты всех атомов в независимой части ячейки (один из атомов Ca находится в частной позиции на плоскости XY), без задания жестких фрагментов (т.е. 25 степеней свободы). Была найдена структурная модель с R_{wp} -фактором 11,15 %. На рисунке 27 представлены установки поиска по МПГА, а на рисунке 28 – графики сходимости МПГА на УЯ при поиске структуры $\text{Ca}_2\text{Al}_3\text{O}_6\text{F}$. На оси абсцисс указан номер поколения эволюции (нулевой номер – начальная случайная генерация всех популяций), на оси ординат – соответствующие ему значения R-фактора.

Параметры работы ГА: R_{wp} χ^2

Бит на каждую координату	0
Размер популяции	300
Кол-во поколений	500
Кол-во индивидов для лок. поиска по МНК	3
Порог фитнеса для запуска МНК	120
Регулярность запуска МНК, поколений	10
Кол-во циклов МНК при лок. поиске	2
Диапазон поколений для обмена	3
Сколько индивидов отправлять при обмене	1
Кол-во поколений локального элитизма	3
Записывать в лог всех индивидов	0

Параметры поиска структуры:

Минимальная дистанция по умолчанию	1.6
Коэффициент штрафа при сближении < min_dist	0.8
Исп. динамическую заполняемость ат. позиций	0
Исп. схлопывание на элементах симметрии	1
Фактор сохранения начальных значений	0

Количество рабочих ядер

Номер ядра	Размер турнира	К. скрещивания	К. мутации
1	3	0,75	0,9
2	3	0,75	1
3	3	0,75	1,1

AntiBump

Bonds

Жесткие фрагменты

Рисунок 27 – Окно задания параметров МПГА при поиске структуры $\text{Ca}_2\text{Al}_3\text{O}_6\text{F}$

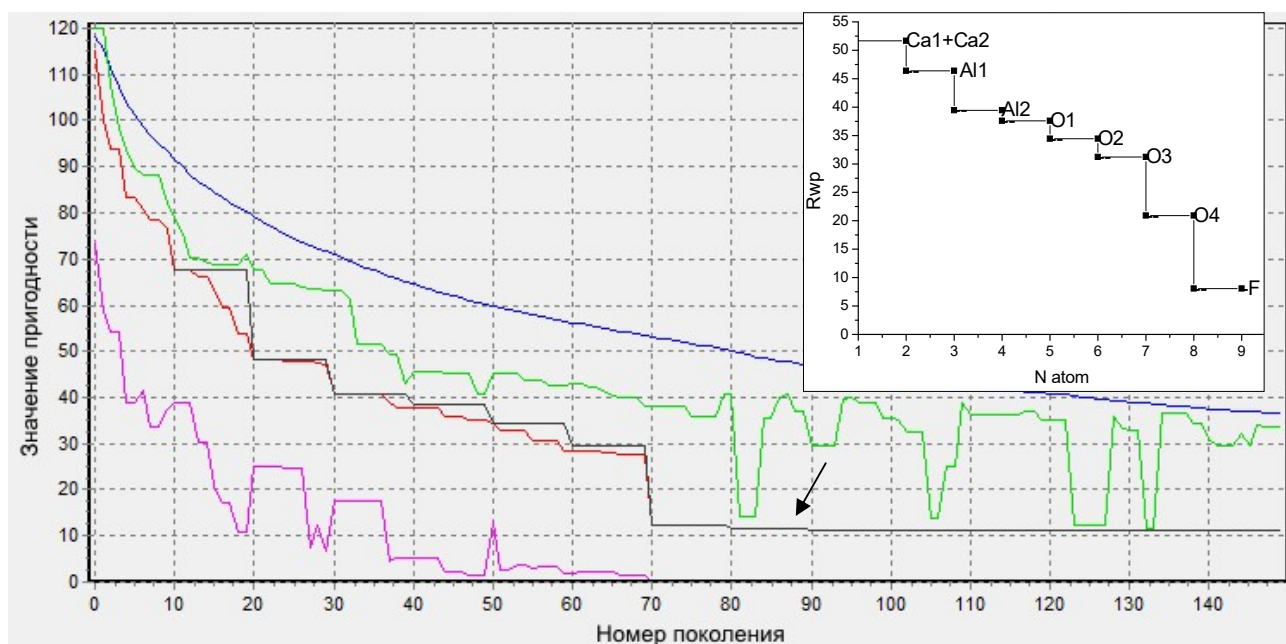


Рисунок 28 – Графики сходимости МПГА на УЯ при поиске структуры $\text{Ca}_2\text{Al}_3\text{O}_6\text{F}$.

Красная линия – значение пригодности для лучшего индивида на данном поколении; чёрная – лучшая пригодность, полученная при доуточнении части индивидов по МНК; фиолетовая – показывает вклад штрафа в значение лучшей пригодности; зелёная – показывает лучшую пригодность для наихудшего из рабочих ядер; синяя – средняя пригодность индивидов, накопленных на управляющем ядре.

Вверху справа приведён сравнительный график снижения R_{wp} при последовательной расстановке вручную атомов в истинные позиции в этой структуре.

На начальных этапах работы МПГА, когда генерируемые структуры ещё далеки от искомой, локальный поиск не используется, а вычисляются только значение пригодности (сумма профильного R-фактора и штрафов). На графике сходимости – это значение пригодности для лучшего найденного решения, поступившее к текущему поколению на УЯ со всех рабочих узлов, отображается красной линией. Аналогично, черная линия отражает значение лучшей пригодности локального поиска, получаемого в результате уточнения методом МПР в течение заданного числа циклов лучших индивидов на рабочих узлах с пригодностью ниже порогового значения. Очевидно, оба этих графика не могут повышаться. При поиске данной структуры пороговое значение было равно 70, а уточнение по МПР выполнялось раз в 10 поколений. Поэтому красная линия на большинстве поколений ниже черной, а ступенчатая форма последней отражает роль локального поиска для достижения результата. Фиолетовая линия показывает размер штрафа, наложенного на R-фактор за чрезмерные (нефизичные) сближения атомов в лучшей найденной модели. Синяя линия изображает среднее значение R-фактора для лучшей половины структурных моделей, накопленных на УЯ. В случае ее сближения с черной до достижения последней целевого значения R-фактора это послужило бы индикатором стагнации процесса эволюции. Зеленая линия показывает худшее значение R-фактора для всех лучших моде-

лей на всех вычислительных ядрах на данном поколении. Ее скачки вверх иллюстрируют разрушение лучшей модели в какой-либо из популяций на вычислительных ядрах, а скачки вниз – включение одной из лучших моделей во все популяции. Из-за случайного характера миграции сближение зеленой линии с черной почти всегда запаздывает, а расстояние между ними характеризует скорость распространения лучших структурных моделей с УЯ по популяциям (может регулироваться коэффициентами миграции $S1$ и $S2$).

Процесс сходимости к истинной структуре $\text{Ca}_2\text{Al}_3\text{O}_6\text{F}$ можно описать с помощью графиков рисунка 28 следующим образом. На рисунке вверху справа для сравнения приведен график снижения R_{wp} -фактора при последовательной ручной расстановке атомов в истинные позиции в структуре (заметим, что влияние изменения порядка расстановки легких атомов O и F на вид этого графика влияет несущественно). Этот график хорошо интерпретирует ступенчатый вид черного графика пригодности при сходимости в МПГА лучшей структурной модели на УЯ. Первая черная ступенька слева ($Y=68$) соответствует размещению атомов кальция вблизи их истинных позиций в структуре. Следующая ступенька ($Y=48$) – автоматическому доуточнению по МНК позиций кальция и нахождению 1-го атома алюминия. Следующие 5 ступенек – последовательному определению позиций 2-го алюминия и, далее, 4-х легких атомов. Скачки вверх зеленых графиков показывают периодическое разрушение мутациями уже найденных лучших моделей на некоторых ядрах (что не влияет на общую сходимость на УЯ, но обеспечивает разнообразие структурных моделей в МПГА). Наконец, последняя ступенька с резким уменьшением значения пригодности ($Y=12$) соответствует нахождению последнего легкого атома. С этого момента фиолетовый график штрафов равен нулю, что показывает отсутствие некорректных межатомных расстояний (т.е. здесь $Y=R_{\text{wp}}$). Две последние малые черные ступеньки (к $R_{\text{wp}}=11.15\%$) соответствуют автоматическому доуточнению по МНК позиций легких атомов. Момент сходимости МПГА к минимальному значению R_{wp} -фактора (11,15%), а лучшей структурной модели к истинной структуре (таблица 9), отмечен стрелкой (на 89-м поколении). Заметим, что здесь было задано всего по 2 цикла уточнения лучших структурных моделей по МПР. Поэтому, для повышения точности требовалось финальное ручное доуточнение найденной структуры по МПР в программе DDM.

Таким образом, динамика изменения этих графиков позволяет анализировать процесс сходимости МПГА. По завершению работы МПГА для найденной структуры строится график соответствия расчетной и экспериментальной дифрактограмм (рисунок 29).

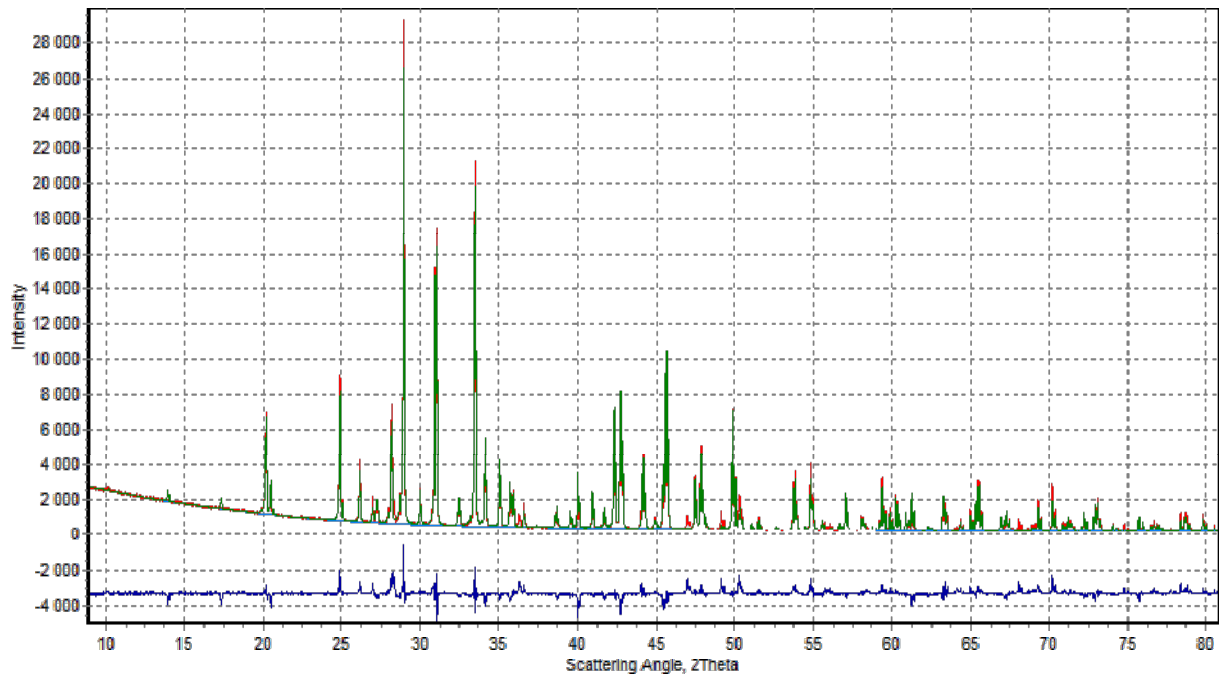


Рисунок 29 – Расчетная (зеленая) и экспериментальная (красная) дифрактограммы, и их разность (синяя) для найденной по МПГА структуры $\text{Ca}_2\text{Al}_3\text{O}_6\text{F}$; R_{wp} -фактор 11,15 %.

Решение этой структуры в МПГА было найдено в другой кристаллографической установке. Для сравнения полученного решения с истинным (таблица 8) оба этих набора атомных координат были нормализованы при помощи функции трансляции атомных позиций в независимую часть ячейки, которая входит в состав библиотеки ObjCryst++ [62]. В таблице 9 приведено сравнение нормализованных атомных координат полученного решения данной структуры по МПГА с нормализованными истинными значениями атомных координат.

Таблица 9 – Сравнение нормализованных атомных координат для истинной структуры (столбцы со *) $\text{Ca}_2\text{Al}_3\text{O}_6\text{F}$, и найденной при помощи МПГА ($R_{\text{wp}} = 11,15$)

Ат.	X*	X _{МПГА}	\Delta	Y*	Y _{МПГА}	\Delta	Z*	Z _{МПГА}	\Delta
Ca1	0.14163	0.14348	0.00185	0.88983	0.89163	0.00180	0.59070	0.59289	0.00219
Ca2	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.27710	0.27608	0.00102
Al1	0.07388	0.07479	0.00091	0.87673	0.87570	0.00103	0.03490	0.03627	0.00137
Al2	0.25977	0.26093	0.00116	0.97658	0.97806	0.00148	0.23066	0.23693	0.00627
O1	0.26096	0.26489	0.00393	0.87363	0.87479	0.00116	0.77911	0.77314	0.00597
O2	0.11132	0.11314	0.00182	0.97131	0.97155	0.00024	0.88275	0.87576	0.00699
O3	0.14758	0.14918	0.00160	0.89152	0.89193	0.00041	0.21582	0.21186	0.00396
O4	0.32532	0.32467	0.00065	0.93420	0.93937	0.00516	0.13927	0.14344	0.00416
F	0.00780	0.00585	0.00195	0.90146	0.90335	0.00189	0.51323	0.51760	0.00437

На рисунке 30 приведено сравнение расчётной и экспериментальной дифрактограмм после финального полнопрофильного уточнения по МПР найденной по МПГА структуры.

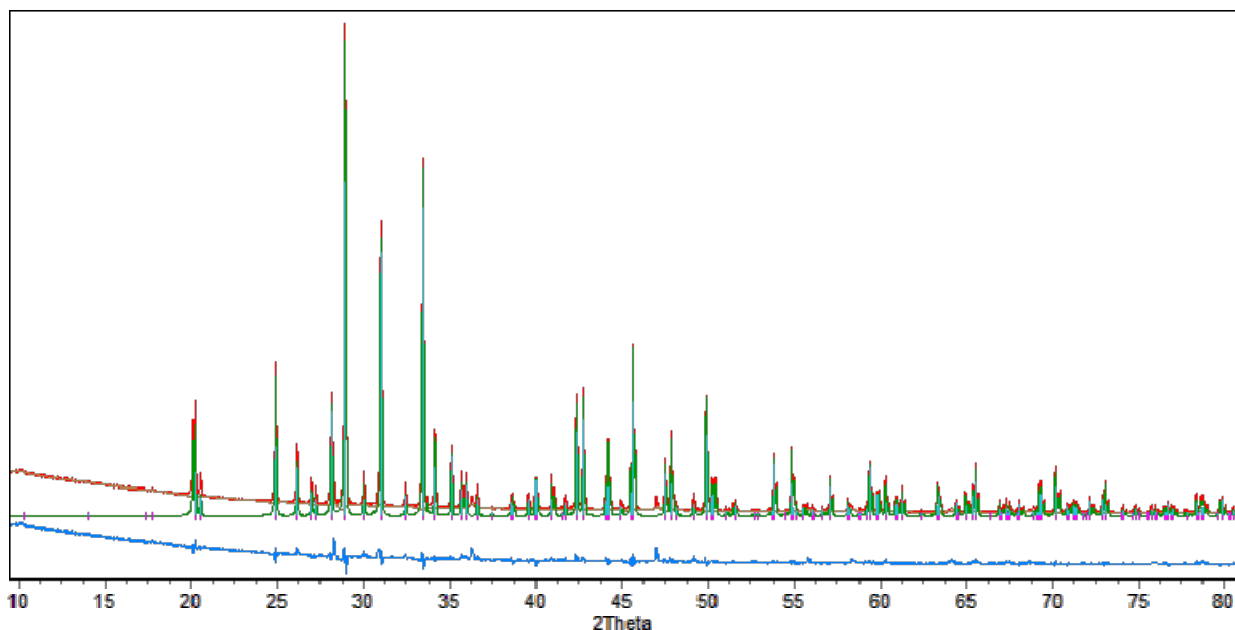


Рисунок 30 – Расчетная (зеленая) и экспериментальная (красная) дифрактограммы, и их разность (синяя) для найденной по МПГА и уточнённой по МПР структуры $\text{Ca}_2\text{Al}_3\text{O}_6\text{F}$; R_{wp} -фактор 8,03 %.

В таблице 10 приведено аналогичное сравнение для финального решения структуры, полученного после полнопрофильного уточнения по МПР.

Таблица 10 – Сравнение нормализованных атомных координат* для истинной структуры $\text{Ca}_2\text{Al}_3\text{O}_6\text{F}$, и найденной при помощи МПГА после ее уточнения по МПР ($R_{\text{wp}} = 8,03$)

Ат.	X^*	$X_{\text{МПГА}}$	$ \Delta $	Y^*	$Y_{\text{МПГА}}$	$ \Delta $	Z^*	$Z_{\text{МПГА}}$	$ \Delta $
Ca1	0.14163	0.14167	0.00003	0.88983	0.88998	0.00015	0.5907	0.59062	0.00008
Ca2	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.2771	0.27795	0.00085
Al1	0.07388	0.07377	0.0001	0.87673	0.87713	0.00039	0.0349	0.03551	0.00061
Al2	0.25977	0.26002	0.00025	0.97658	0.97712	0.00054	0.23066	0.2301	0.00056
O1	0.26096	0.26055	0.0004	0.87363	0.8729	0.00074	0.77911	0.78087	0.00176
O2	0.11132	0.11322	0.0019	0.97131	0.9717	0.00039	0.88275	0.88164	0.00111
O3	0.14758	0.14737	0.0002	0.89152	0.89194	0.00041	0.21582	0.21345	0.00237
O4	0.32532	0.32516	0.00016	0.9342	0.93429	0.00009	0.13927	0.1358	0.00347
F	0.13927	0.1358	0.00347	0.90146	0.9015	0.00003	0.51323	0.5119	0.00134

Определение по МПГА структуры тартрата кальция $\text{CaC}_4\text{H}_4\text{O}_6 \cdot 4\text{H}_2\text{O}$

Эта структура была представлена в качестве конкурсного задания в Round Robin SDPDRR-3 [3], а дифрактограмма соединения была скачана с сайта организаторов конкурса. На рисунке 31 представлены графики сходимости МПГА на УЯ при поиске структуры $\text{CaC}_4\text{H}_4\text{O}_6 \cdot 4\text{H}_2\text{O}$ (пр.гр. Р -1). При поиске структуры этого соединения также, как и участниками Round Robin, применявших методы имитации отжига для поиска структуры в прямом пространстве (ГА никто не применял), был использован жесткий фрагмент молекулы тартрата $\text{C}_4\text{H}_4\text{O}_6$. Для упрощения расчетов были исключены атомы Н и по МПГА определялись положение и ориентация фрагмента C_4O_6 , атома Са и 4-х атомов О молекул воды. Ориентация молекулы задаётся в виде кватерниона из 4 элементов, каждый из которых при поиске варьируется от 0 до 1, и чтобы найти положение и ориентацию молекулы в элементарной ячейке кристалла требуется определить значения 7 параметров. Окно с заданием в МПГА этого фрагмента приведено на рисунке 32. Таким образом, поисковая задача имела 22 степени свободы атомных координат.

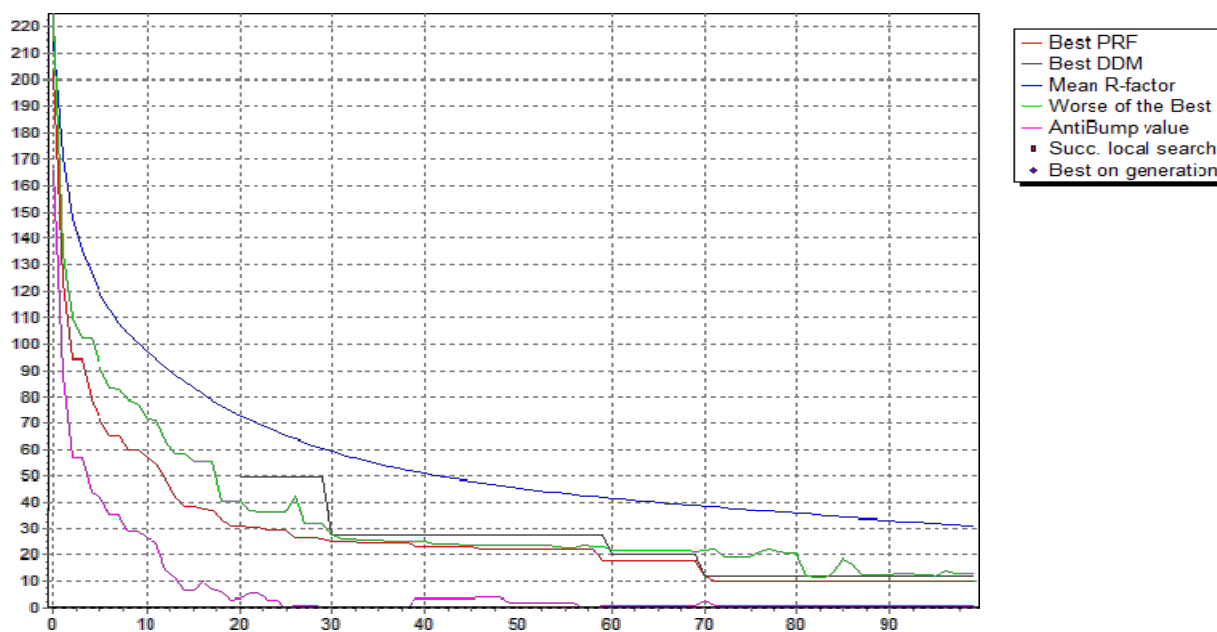


Рисунок 31 – Графики сходимости МПГА при поиске структуры $\text{CaC}_4\text{H}_4\text{O}_6 \cdot 4\text{H}_2\text{O}$; ось абсцисс – номер поколения эволюции, ось ординат – значения R-фактора

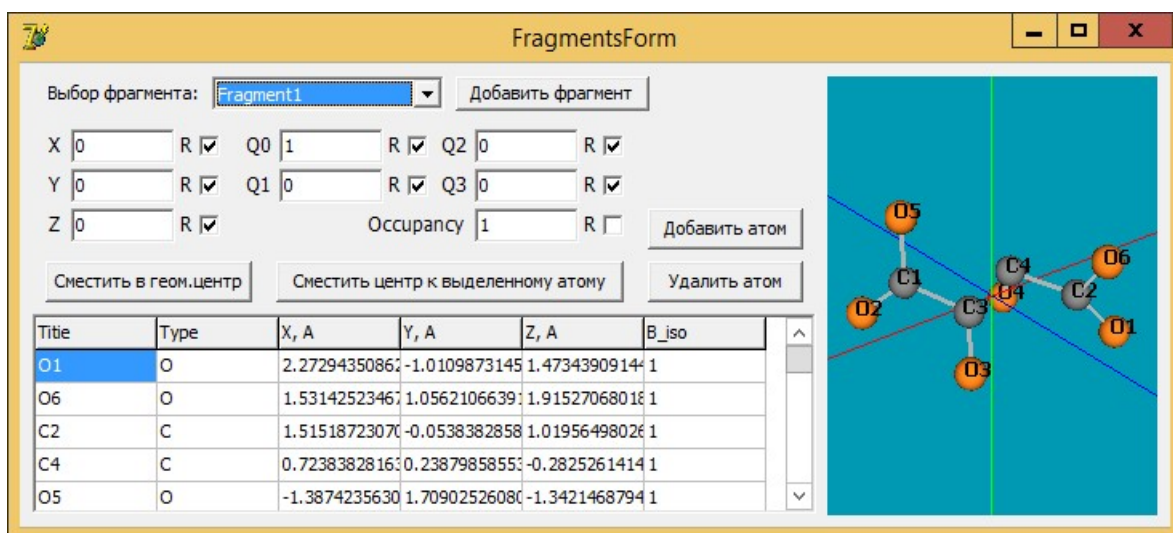


Рисунок 32 – Окно задания молекулы тартрата в виде жесткого фрагмента для поиска структуры $\text{CaC}_4\text{H}_4\text{O}_6 \cdot 4\text{H}_2\text{O}$.

На рисунке 33 показано соответствие расчетной и экспериментальной дифрактограмм для финального решения структуры тартрата кальция по МПГА. В таблице 11 приведено сравнение полученного решения данной структуры с истинными значениями атомных координат.

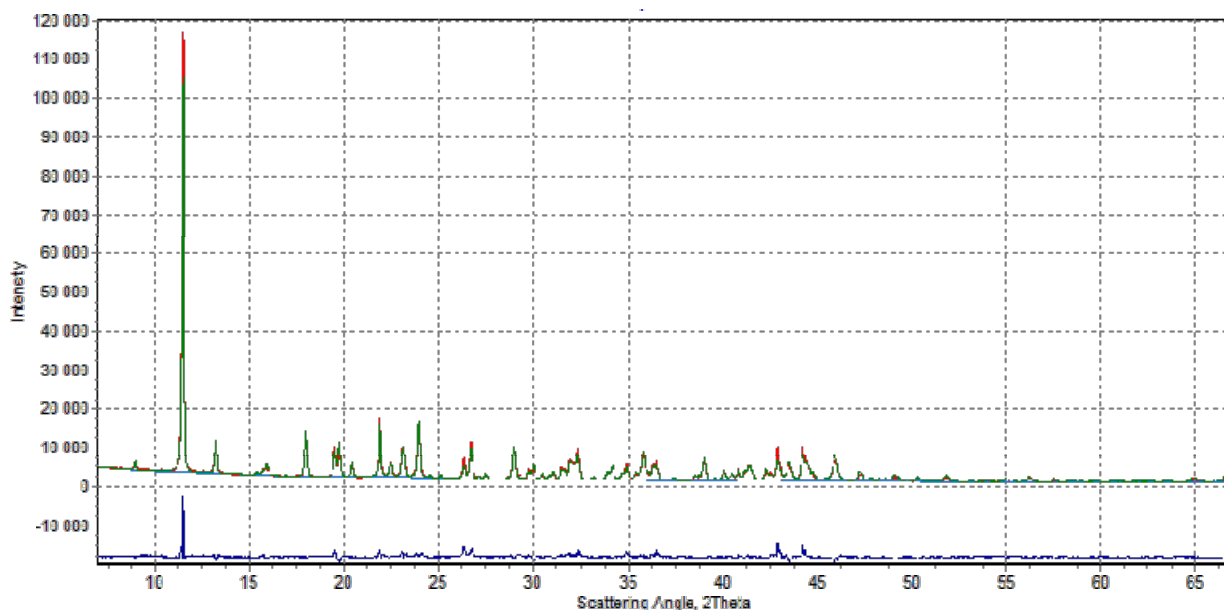


Рисунок 33 – Расчетная (зеленая) и экспериментальная (красная) дифрактограммы, и их разность (синяя) для найденной по МПГА структуры $\text{CaC}_4\text{H}_4\text{O}_6 \cdot 4\text{H}_2\text{O}$; R_{wp} -фактор 14,82 %

Таблица 11 – Сравнение атомных координат для истинной* структуры $\text{CaC}_4\text{H}_4\text{O}_6 \cdot 4\text{H}_2\text{O}$, и найденной при помощи МПГА ($R_{\text{wp}} = 14,82$)

Ат.	X^*	$X_{\text{МПГА}}$	$ \Delta $	Y^*	$Y_{\text{МПГА}}$	$ \Delta $	Z^*	$Z_{\text{МПГА}}$	$ \Delta $
Ca	0.33672	0.33956	0.00284	0.19675	0.18432	0.01244	0.26828	0.27481	0.00652
O1	0.66573	0.68263	0.01690	0.01142	0.01493	0.00351	0.90677	0.91580	0.00903
O2	0.82099	0.80960	0.01139	0.02823	0.01329	0.01494	0.58985	0.59619	0.00634
O3	0.92401	0.91360	0.01042	0.01022	0.03138	0.02117	0.15925	0.16723	0.00798
O4	0.62713	0.63307	0.00594	0.01731	0.02483	0.00752	0.34292	0.33745	0.00547
O5	0.05711	0.06744	0.01033	0.27591	0.28745	0.01154	0.63215	0.63038	0.00177
O6	0.71164	0.68763	0.02401	0.29766	0.28762	0.01004	0.98637	0.96059	0.02579
C1	0.06053	-0.00795	0.06849	0.14692	0.12875	0.01817	0.67903	0.66345	0.01558
C2	0.64663	0.58145	0.06518	0.17455	0.12467	0.04988	0.89655	0.86979	0.02676
C3	0.23595	0.16460	0.07135	0.14281	0.11287	0.02993	0.81865	0.79504	0.02361
C4	0.42796	0.39540	0.03256	0.15194	0.12936	0.02258	0.74649	0.73780	0.00870
Ow1	0.59365	0.61800	0.02435	0.32710	0.31107	0.01604	0.49262	0.51137	0.01875
Ow2	0.13754	0.11839	0.01915	0.35161	0.32290	0.02871	0.39298	0.37787	0.01511
Ow3	0.69829	0.70368	0.00540	0.37763	0.38124	0.00361	0.23809	0.22328	0.01482
Ow4	0.18158	0.16703	0.01455	0.34148	0.35054	0.00905	0.10155	0.09028	0.01128

Связывание нескольких атомов в жесткий фрагмент обеспечивает уменьшение числа степеней свободы структуры и, следовательно, упрощает процесс поиска её структурных параметров по МПГА. Но жесткое связывание группы атомов одновременно делает финальное решение МПГА лишь приближением к истинному решению. Это хорошо видно из предыдущей таблицы – погрешность в нахождении атомных координат достигает 0,07 доли ячейки. В этом случае актуальным становится доуточнение полученной структуры по МНК. В таблице 12 приведено сравнение уточнённого по методу производной разности (МПР) решения с истинными значениями атомных координат для структуры $\text{CaC}_4\text{H}_4\text{O}_6 \cdot 4\text{H}_2\text{O}$. Значение профильного R-фактора после уточнения снизилось до 12.19%. Положения атомов углерода продолжают заметно отличаться от истинных значений. Вероятно, это вызвано тем, что атомы углерода наиболее лёгкие в этой структуре и не учтены положения связанных с ними атомов водорода.

Таблица 12 – Сравнение атомных координат для истинной* структуры $\text{CaC}_4\text{H}_4\text{O}_6 \cdot 4\text{H}_2\text{O}$ и найденной при помощи МПГА и уточнённой по МПР ($R_{\text{wp}} = 12,19$)

Ат.	X*	X _{МПГА}	\Delta	Y*	Y _{МПГА}	\Delta	Z*	Z _{МПГА}	\Delta
Ca	0.33672	0.34202	0.00530	0.19675	0.18786	0.00889	0.26828	0.27273	0.00445
O1	0.66573	0.68805	0.02232	0.01142	0.02134	0.00992	0.90677	0.91435	0.00758
O2	0.82099	0.82711	0.00613	0.02823	0.02270	0.00552	0.58985	0.60081	0.01096
O3	0.92401	0.90956	0.01445	0.01022	0.02419	0.01397	0.15925	0.17088	0.01164
O4	0.62713	0.62914	0.00201	0.01731	0.01926	0.00195	0.34292	0.34080	0.00213
O5	0.05711	0.08331	0.02620	0.27591	0.29831	0.02240	0.63215	0.62778	0.00437
O6	0.71164	0.70525	0.00638	0.29766	0.29054	0.00712	0.98637	0.96037	0.02600
C1	0.06053	0.00594	0.05459	0.14692	0.15421	0.00728	0.67903	0.66548	0.01355
C2	0.64663	0.59331	0.05333	0.17455	0.14361	0.03094	0.89655	0.87954	0.01701
C3	0.23595	0.18537	0.05057	0.14281	0.13619	0.00662	0.81865	0.80736	0.01129
C4	0.42796	0.40765	0.02031	0.15194	0.13419	0.01774	0.74649	0.74057	0.00592
Ow1	0.59365	0.61108	0.01743	0.32710	0.31254	0.01457	0.49262	0.50901	0.01639
Ow2	0.13754	0.13327	0.00427	0.35161	0.33771	0.01390	0.39298	0.38333	0.00965
Ow3	0.69829	0.69324	0.00505	0.37763	0.37930	0.00167	0.23809	0.22135	0.01674
Ow4	0.18158	0.18848	0.00690	0.34148	0.35605	0.01457	0.10155	0.09869	0.00287

2.4.3 Выводы по результатам апробации МПГА

В результате апробации и тестирования на разных структурах была выработана обобщённая методика поиска кристаллической структуры по МПГА, которая сводится к следующему. Вначале пользователь загружает прецизионную дифрактограмму образца. Затем задаёт имеющуюся информацию о фазе, для которой определяется структура, а в случае необходимости, и об известных примесных фазах. Вводятся группы симметрии и параметры элементарной ячейки, а для примесных фаз и параметры их структуры. Затем список известных или предполагаемых атомов определяемой фазы и, при наличии, известных структурных фрагментов (молекул) и ограничений. Потом производится уточнение профильных параметров по процедуре декомпозиции Ле Бейла, которая не требует информации об атомном составе фаз и, поэтому, полученное значение R-фактора служит далее целевым значением. После этого пользователь отмечает структурные и прочие параметры, которые требуется найти по МПГА, задаёт настройки процесса ГА, и запускает процесс поиска. Поскольку МПГА – стохастический метод, рекомендуется задание нескольких последовательных пусков МПГА в цикле (3-5 пусков). Контроль результатов выполняется по графикам сходимости МПГА, близости R-фактора лучшей структурной модели с целевым значением, графикам соответствия расчетных рентгенограмм с экспериментальной и динамической визуализации структурной модели. В случае достижения удовлетворительных результатов выполняется доуточнение полученной структурной модели полнопрофильным методом, а затем проверка пользователем ее кристаллохимической корректности. Если же ни в одном из пусков удовлетворительных результатов не наблюдается, необходимо

сопоставление графиков сходимости МПГА на УЯ с графиками сходимости индивидуальных ГА на разных вычислительных ядрах для оценки возможных причин стагнации эволюционного процесса и соответствующая перенастройка параметров поиска кристаллической структуры в МПГА, описанных в таблице 4.

Результаты тестирования МПГА подтверждают, что генетические алгоритмы обеспечивают *ab initio* поиск кристаллической структуры химических соединений по порошковой дифрактограмме. Если искомая структура не слишком сложная (~30 степеней свободы для ПК и ~50 для кластера), дифрактограмма высокого качества, а химический состав, симметрия и параметры решетки верны, то при подходящей настройке параметров поиска МПГА структура может быть найдена в одном из нескольких пусков программы.

Для поиска более сложных структур алгоритм предстоит усовершенствовать. Для этого имеется существенный потенциал развития, в частности, за счет более эффективной организации параллельной работы на многоядерных ПК и, особенно, на суперкомпьютерных кластерах с большим числом ядер.

Повышение эффективности поиска структур происходит за счет миграции в мультипопуляционных алгоритмах и может быть объяснено следующим. Стохастические процессы ГА на разных ядрах попадают в разные многочисленные локальные минимумы на гиперповерхности R-фактора. Периодическое и умеренное «разбавление» плохо сходящихся популяций лучшими структурными моделями из других популяций приводит к «конкуренции локальных минимумов». Это обеспечивает выходы из них и улучшает общую сходимость эволюционного процесса в глобальный минимум. Другим статистически важным фактором ускорения сходимости является периодическое уточнение лучших структурных моделей полнопрофильным методом. Уточнение в ряде случаев обеспечивает установку отдельных атомов в их истинные позиции в структуре с понижением R-фактора (что можно наблюдать на графиках сходимости). Это, в свою очередь, обеспечивает преимущественный селекционный отбор этих уточненных моделей, распространение и накопление уточненных атомных позиций в некоторых структурных моделях следующих поколений данной популяции в результате скрещиваний, а также их распространение в другие популяции за счет миграции. В итоге, стохастический процесс накопления уточненных атомных позиций может привести к сходимости какой-либо из структурных моделей к истинной кристаллической структуре данного вещества.

Нами планируется дальнейшее развитие этого подхода для определения более сложных кристаллических структур.

2.5 Сравнение МПГА с современными методами определения кристаллических структур в прямом пространстве

2.5.1 Определение структуры Ba_2CrO_4

Было выполнено экспериментальное сравнение надежности поиска структуры по МПГА с последними доступными версиями известных программ FOX [137] и DASH [138], реализующих метод имитации отжига. В качестве тестовой была взята структура соединения Ba_2CrO_4 (пр. гр. $Pn21a$; $a=7.67 \text{ \AA}$, $b=5.89 \text{ \AA}$, $c=10.39 \text{ \AA}$). Для исключения влияния на поиск структуры качества моделирования профильных параметров, индивидуального в разных программах, была использована расчетная и зашумленная дифрактограмма Ba_2CrO_4 . Таким образом, во всех программах использовались идентичные профильные параметры, заданные при расчете дифрактограммы. Производился *ab initio* поиск 7 атомов в общих позициях (всего 21 атомная координата), целевой профильный R-фактор, соответствующий истинной структуре, был равен 0.79%. Результаты 10 пусков программ FOX, DASH и ГА приведены, соответственно, в таблицах 13–15. Поскольку целевые функции этих программ различны, для сравнения финальных структур в таблицах приведен профильный R-фактор МПР. Затем каждая финальная структура уточнялась методом полнопрофильного анализа МПР, полученный профильный R-фактор приведен рядом.

Таблица 13 – Результаты 10-ти пусков программы FOX при поиске тестовой структуры Ba_2CrO_4

№ пуска	Значение ЦФ: Cost FOX	Профильный R-фактор МПР финальной структуры	R-фактор после уточнения по МПР	Результат
1	10 212	3.21	0.83	Сошелся
2	10 263	3.27	0.79	Сошелся
3	10 204	3.10	0.79	Сошелся
4	10 146	3.55	0.79	Сошелся
5	16 112	5.46	0.79	Сошелся
6	10 156	4.04	0.79	Сошелся
7	10 357	3.34	0.82	Сошелся
8	103 610	14.50	6.41	Не сошелся
9	127 062	17.55	7.59	Не сошелся
10	10 281	3.03	0.79	Сошелся
Итого процент успешных пусков:				80%
Среднее время на пуск:				14 минут

Таблица 14 – Результаты 10-ти пусков программы DASH при поиске тестовой структуры Ba_2CrO_4

№ пуска	Значение ЦФ: Profile Chi Squared	Профильный R- фактор МПР финальной структуры	R-фактор после уточнения по МПР	Результат
1	19.9946	12.25	7.64	Не сошелся
2	12.8983	10.25	0.81	Сошелся
3	23.8734	13.46	6.84	Не сошелся
4	12.4780	10.35	0.79	Сошелся
5	12.3666	10.25	0.79	Сошелся
6	32.9379	17.10	8.52	Не сошелся
7	22.8771	15.08	8.33	Не сошелся
8	19.6824	12.30	7.64	Не сошелся
9	12.6913	10.42	0.79	Сошелся
10	12.3135	10.35	0.79	Сошелся
Итого процент успешных пусков:				50%
Среднее время на пуск:				5 минут

Таблица 15 – Результаты 10-ти пусков МПГА при поиске тестовой структуры Ba_2CrO_4

№ пуска ГА	Значение ЦФ финальной структуры: профильный R-фактор МПР	R-фактор после уточнения по МПР	Результат
1	2,59	0.79	Сошелся
2	2,63	0.79	Сошелся
3	2,60	0.79	Сошелся
4	2,59	0.79	Сошелся
5	2,59	0.79	Сошелся
6	2,66	0.79	Сошелся
7	2,59	0.79	Сошелся
8	2,59	0.79	Сошелся
9	2,60	0.79	Сошелся
10	2,59	0.79	Сошелся
Итого процент успешных пусков:			100%
Среднее время на пуск:			1,5 минуты

На рисунке 34 изображены типичные графики сходимости программ FOX, DASH и ГА для успешных пусков (последних в таблицах 13-15).

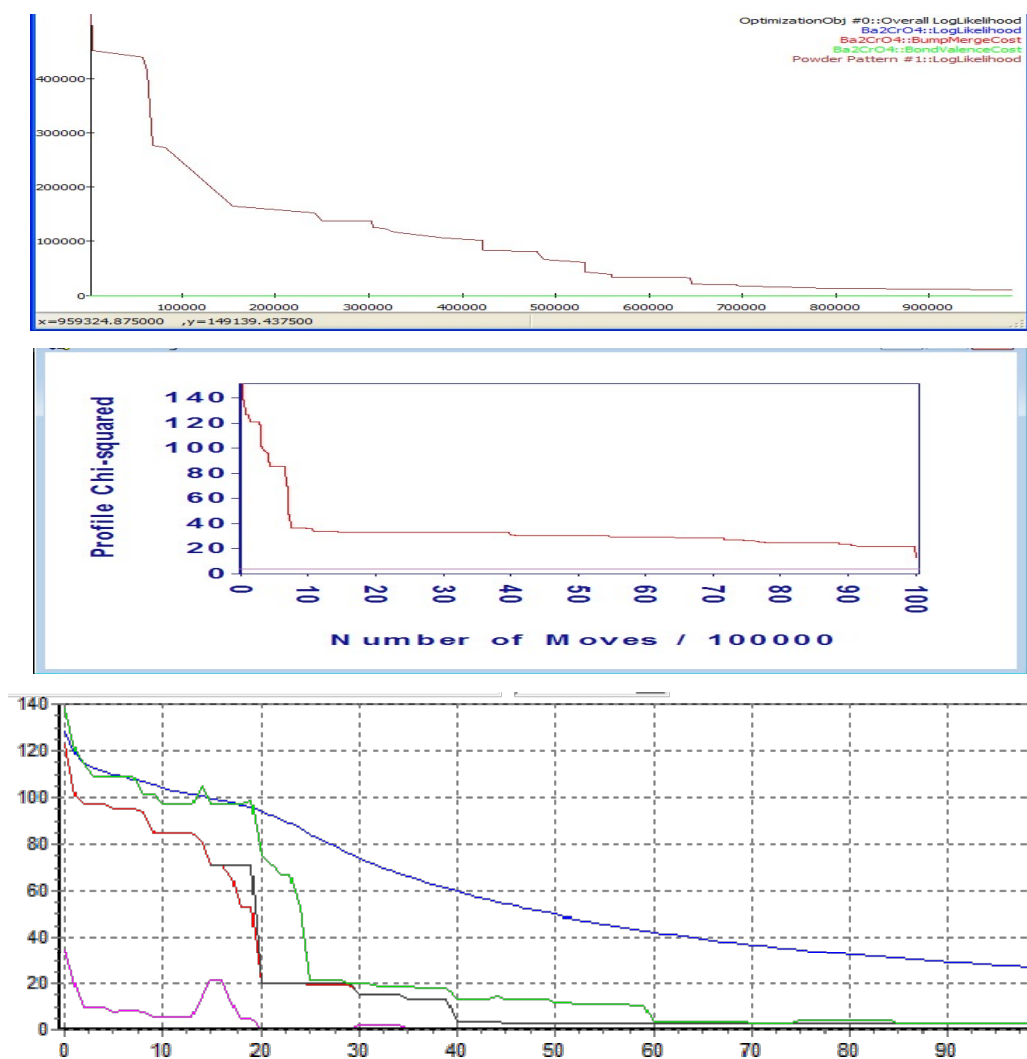


Рисунок 34 – Типичные графики сходимости для успешных пусков FOX (сверху), DASH (в середине) и ГА (снизу) при поиске тестовой структуры Ba_2CrO_4

Программа FOX находит адекватную структурную модель Ba_2CrO_4 после оценки примерно 10^6 пробных конфигураций структуры, R-фактор МПР для полученной по FOX модели равен 3.03 и после ее уточнения по МПР снизился до целевого значения 0.79, а уточненная структура совпала с истинной. Программе DASH на это понадобилось около $100 \cdot 10^6$ пробных конфигураций, однако ее целевая функция – проще, код программы хорошо оптимизирован, и поэтому ей для сходимости понадобилось меньше времени. R-фактор МПР для полученной по DASH модели равен 10.35 и после ее уточнения по МПР снизился до целевого значения 0.79, а уточненная структура совпала с истинной. Для ГА был задан размер популяции из 100 модельных структур, а истинная структура была найдена на 45-м поколении с R-фактором 2.6, уточненным по МПР до 0.79.

Таким образом, для нахождения 21 атомной координаты в независимой части ячейки было оценено всего около 4500 модельных структур. Столь малое число оцениваемых структурных моделей объясняется параллельным выполнением генетических операций скрещивания и мута-

ций над целой популяцией из 100 структур, и вероятностной селекцией по R-фактору лучших из них на следующие поколения.

2.5.2 Определение структуры $\text{Pd}(\text{NH}_3)_2(\text{NO}_2)_2$

Здесь МПГА, использованная в варианте локальной оптимизации структурных моделей по МПР, сравнивается с наиболее распространенной программой глобальной оптимизации FOX, описанной в разделах 1.3 и 1.6.3. Программа активно развивается более 12 лет, хорошо себя зарекомендовала и ее можно считать «точкой отсчета» при определении структуры. Например, из 12 решений структурных задач *SDPDRR-3*, описанного в разделе 1.3, почти половина получена с помощью FOX. Для тестирования использован поиск кристаллической структуры соединения $\text{Pd}(\text{NH}_3)_2(\text{NO}_2)_2$ [139], в которой атом Pd расположен в априори известной частной позиции (в центре инверсии), и экспериментальная рентгенограмма этого соединения, зарегистрированная на дифрактометре с шагом 0,02 по 2 тета в монохроматизированном излучении $\text{CuK}\alpha$.

Эта задача решалась с помощью FOX двумя различными путями – размещением одиночных атомов непосредственно и заданием жестко определенных структурных фрагментов. При задании независимых одиночных атомов (аналогично установкам в генетическом алгоритме) при поиске в разных режимах удалось корректно определить расположение атомов N и O (таблица 16), однако ни разу не удалось даже приблизительно локализовать атомы H. В большинстве запусков алгоритм поиска оставался в ложном минимуме с перестановкой атомов N и O, не обеспечив находение истинного решения (12 тестовых запусков из 20). Для получения данного решения в среднем требовалось перебрать и оценить $\sim 10^5$ тестовых структур. Резкий спад в начале процесса поиска ($\sim 10^4$ структур) иногда сменяется стагнацией в локальном минимуме (рисунок 35).

Таблица 16 – Структура $\text{Pd}(\text{NH}_3)_2(\text{NO}_2)_2$. Приведены эталонные значения координат, найденные при помощи FOX (с независимым заданием атомов) и их разность

Ат.	X^*	X_{FOX}	$ \Delta $	Y^*	Y_{FOX}	$ \Delta $	Z^*	Z_{FOX}	$ \Delta $
N1	0.2041	0.204921	0.000821	0.16096	0.150756	0.0102	0.80107	0.8237	0.02263
O2	0.3156	0.305154	0.010446	0.0324	0.040205	0.0078	0.70688	0.707127	0.000247
O3	0.2172	0.222088	0.004888	0.38686	0.378549	0.00831	0.77424	0.773245	0.000995
N2	0.22449	0.233889	0.009399	0.28738	0.27606	0.01132	0.25252	0.241128	0.011392

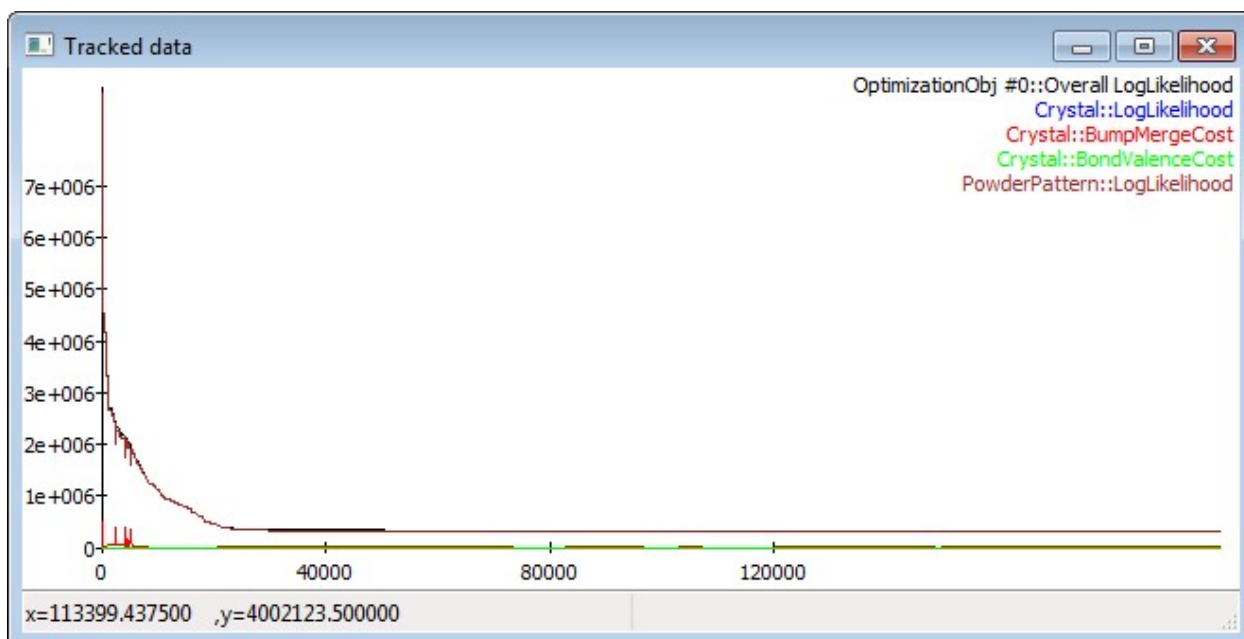


Рисунок 35 – Типичный график сходимости процесса решения структуры $\text{Pd}(\text{NH}_3)_2(\text{NO}_2)_2$ в FOX. По оси абсцисс – число сгенерированных тестовых структур, по оси ординат – значение критерия качества тестовой структуры (χ^2)

Сходимость к верному решению достигалась только при задании ограничений на межатомные расстояния для каждого типа атомов во избежание их сближения и «схлопывания» друг в друга (рисунок 36). Надежность и эффективность алгоритма в основном зависит от степени сужения пространства поиска по имеющейся дополнительной информации о длинах связей в исследуемой структуре. К примеру, только задание расстояния N-N больше 2 ангстрем позволило устойчиво избегать перестановок атомов N и O. Однако, на практике определение подобных значений может быть затруднительным. Важно учесть, что чрезмерно жесткие ограничения способны резко ухудшить надежность алгоритма, практически исключив возможность сходимости к верному решению. Заметим, что в ГА поиск всех атомов выполнялся по независимой части ячейки – без ограничений. Перестановки атомов N и O в ГА тоже иногда имели место, но заметно реже, а «схлопывания» друг в друга урегулировались штрафной функцией и динамическим усилением мутаций.

	Pd	N	O	H
Pd	5.000	1.500	2.500	2.000
N	1.500	2.000	1.000	0.650
O	2.500	1.000	1.500	1.500
H	2.000	0.650	1.500	1.000

Рисунок 36 – Ограничения на межатомные расстояния – функция anti-bump из FOX

Координаты атомов водорода поддавались локализации только с заданием 2-х жестко определенных структурных фрагментов: NO₂ и NH₃. Однако, фрагмент NH₃ не был точно ориентирован программой даже при заданных углах между атомами Н (таблица 17). Такой подход требует знания существенной априорной информации о структуре исследуемого вещества, что в практических приложениях не всегда выполнимо.

Таблица 17 – Сравнение атомных координат для эталонной структуры, и найденной в программе при помощи FOX (с использованием структурных фрагментов) для структуры Pd(NH₃)₂(NO₂)₂

Ат.	X*	X _{FOX}	\Delta	Y*	Y _{FOX}	\Delta	Z*	Z _{FOX}	\Delta
N1	0.2041	0.212208	0.008108	0.16096	0.163584	0.00262	0.80107	0.81277	0.0117
O2	0.3156	0.302501	0.013099	0.0324	0.0271	0.0053	0.70688	0.706738	0.00014
O3	0.2172	0.21987	0.00267	0.38686	0.387083	0.00022	0.77424	0.778595	0.00436
N2	0.22449	0.231467	0.006977	0.28738	0.290128	0.00275	0.25252	0.244531	0.00799
H1	0.37096	0.350977	0.019983	0.37277	0.259324	0.11345	0.23067	0.190784	0.03989
H2	0.11429	0.204062	0.089772	0.38565	0.43952	0.05387	0.24882	0.235553	0.01327
H3	0.28708	0.300508	0.013428	0.22297	0.264147	0.04118	0.37765	0.385378	0.00773

При помощи МПГА решение этой структуры включая приблизительные значения координат атомов водорода было найдено в 100% запусков при переборе $2 \cdot 10^4$ тестовых структур. Увеличение эффективности нахождения этой структуры по МПГА связано с возможностью задавать ограничения на количество межатомных связей (для этой структуры они приведены в таблице 18). В таблице 19 приведены значения атомных координат, найденных по МПГА.

Таблица 18 – Сравнение атомных координат для эталонной структуры, и найденной в программе при помощи FOX (с использованием структурных фрагментов) для структуры Pd(NH₃)₂(NO₂)₂

Тип атома 1	Тип атома 2	Необходимое количество связей атомов типа 1 с атомами типа 2
Pd	N	4
Pd	O	0
Pd	H	0
H	N	1
O	H	0
H	H	0

Таблица 19 – Сравнение атомных координат для эталонной структуры, и найденной в программе при помощи МПГА (без использования структурных фрагментов) для структуры $\text{Pd}(\text{NH}_3)_2(\text{NO}_2)_2$

Ат.	X*	X _{МПГА}	\Delta	Y*	Y _{МПГА}	\Delta	Z*	Z _{МПГА}	\Delta
N1	0.2041	0.20758	0.00348	0.16096	0.15606	0.0049	0.80107	0.82258	0.02151
O2	0.3156	0.30709	0.00851	0.0324	0.04341	0.01101	0.70688	0.70758	0.0007
O3	0.2172	0.22461	0.00741	0.38686	0.37908	0.00778	0.77424	0.77176	0.00248
N2	0.22449	0.22515	0.00066	0.28738	0.27494	0.01244	0.25252	0.23677	0.01575
H1	0.37096	0.38309	0.01213	0.37277	0.4425	0.06973	0.23067	0.22435	0.00632
H2	0.11429	0.13951	0.02522	0.38565	0.37175	0.0139	0.24882	0.22179	0.02703
H3	0.28708	0.27628	0.0108	0.22297	0.30169	0.07872	0.37765	0.35513	0.02252

На рисунке 37 приведен график соответствия экспериментальной рентгенограммы с расчетной для найденной по МПГА структуры $\text{Pd}(\text{NH}_3)_2(\text{NO}_2)_2$.

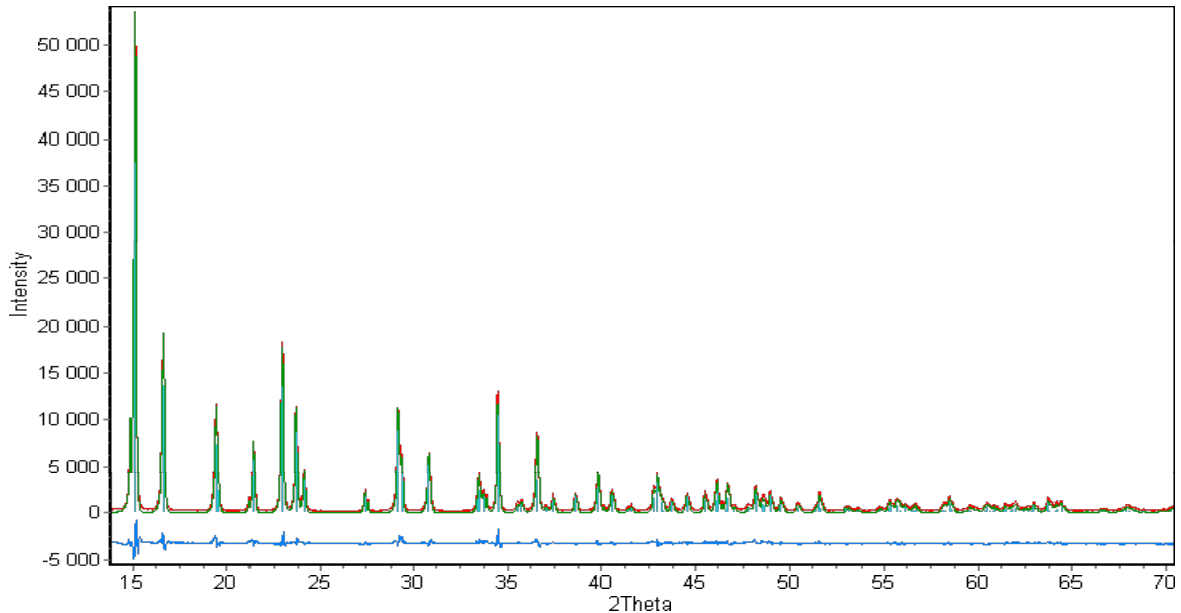


Рисунок 37 – График соответствия экспериментальной и расчетной рентгенограмм для найденной по МПГА структуры $\text{Pd}(\text{NH}_3)_2(\text{NO}_2)_2$

Таким образом, преимуществами МПГА (в варианте с локальным уточнением по МПР) перед FOX при решении данной задачи являются:

- 1) Локализация атомов водорода.
- 2) Точное определение всех атомных позиций после завершающего полнопрофильного уточнения структуры по МПР.
- 3) Менее жесткие требования к ограничениям на структуру.

Очевидно, что определение атомов водорода – это исключительный, хотя и показательный случай, полученный здесь за счет высокой чувствительности уточнения по МПР к легким

атомам и эволюционного процесса ГА к малым изменениям профильного R-фактора МПР при фиксации тяжелого атома Pd в частной позиции.

2.5.3 Определение структуры $\text{Er}_{10}\text{W}_2\text{O}_{21}$

Структура $\text{Er}_{10}\text{W}_2\text{O}_{21}$ (пр. гр. P b c n) с 54 степенями свободы атомных координат является наиболее сложной из тестовых структур и определяется по МПГА на 4-х ядерной ПК весьма редко (в 1-м пуске из 10). Поэтому она является подходящей для изучения сходимости МПГА на суперкомпьютере. При поиске по МПГА на суперкомпьютерном кластере СФУ с использованием 24-х вычислительных узлов (23 популяции размера 500, число поколений эволюции 500) эта структура определяется в 6 случаях из 10. Следует отметить что, например, более простая структура K_4SnO_4 с 27 степенями свободы атомных координат определяется на кластере при тех же ресурсах со 100% надежностью.

На рисунке 38 представлены типичные графики сходимости МПГА на УЯ при поиске тестовой структуры $\text{Er}_{10}\text{W}_2\text{O}_{21}$ на кластере, полученные в одном из успешных пусков. Вверху слева приведен график роста R_{wp} -фактора при последовательном исключении из структуры $\text{Er}_{10}\text{W}_2\text{O}_{21}$ легких, а затем тяжелых атомов. Из него видно, что фрагменту структуры без легких атомов соответствует R_{wp} -фактор менее 25%, а исключение одного из атомов Er сразу же приводит к возрастанию R_{wp} -фактора до 160%. Поэтому из графика сходимости R_{wp} в МПГА (красный график) следует, что положения всех тяжелых атомов нашлось весьма быстро – уже к 30-му поколению (R_{wp} ниже 50). Что вполне естественно, т.к. каждая установка тяжелого атома приводит к снижению R_{wp} -фактора на десятки процентов, а такие структурные модели отбираются селекцией для скрещиваний и переходят в следующее поколение популяции. Далее отыскивались положения только легких атомов. Это отражается на рисунке в виде многократного скачкообразного снижения значений R_{wp} -фактора – до значения 5% при попадании всех тяжелых атомов в истинные позиции в структуре. На рисунке видны скачки сначала красного графика (начавшиеся от значений R_{wp} -фактора около 200%), а затем черного графика – значения, найденные по МНК, рассчитываемый здесь при значениях пригодности ниже 50. Всё оставшееся время МПГА производил поиск положения лёгких атомов кислорода. Последовательное попадание легких атомов в истинные позиции отражается на рисунке также в виде многократного скачкообразного снижения значений R_{wp} -фактора, но уже всего на доли %.

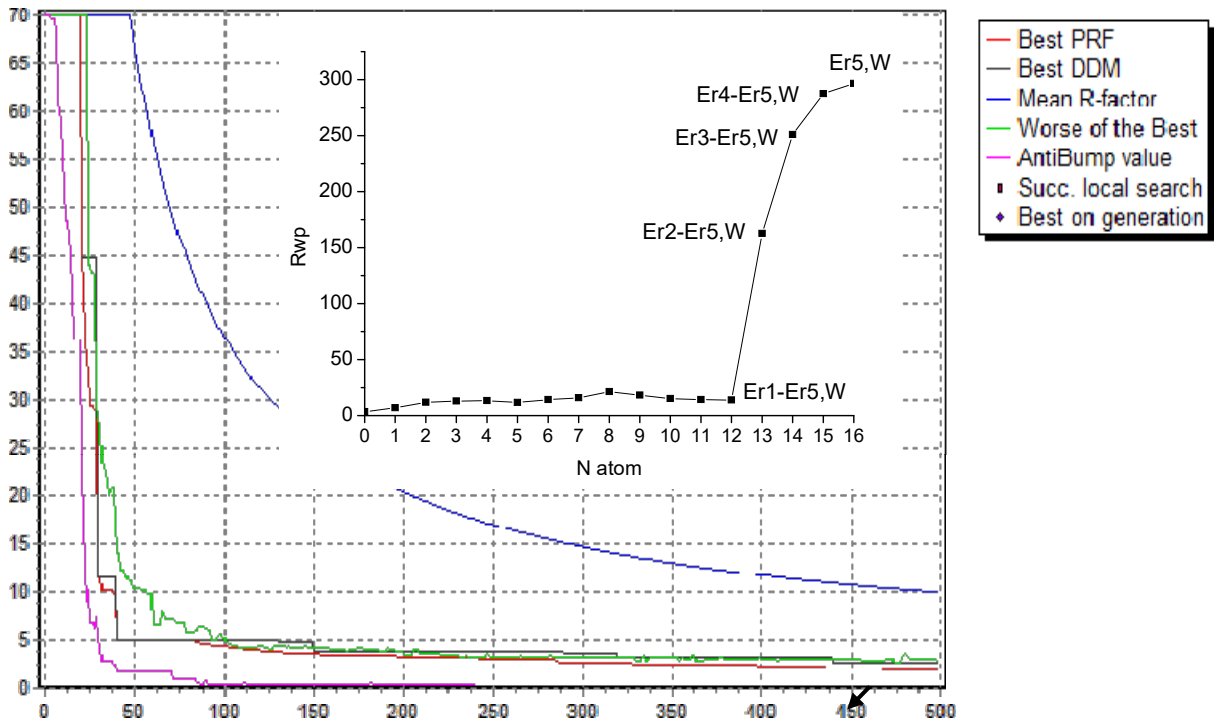


Рисунок 38 – Графики сходимости МПГА при поиске структуры $\text{Er}_{10}\text{W}_2\text{O}_{21}$; ось абсцисс – номер поколения эволюции, ось ординат – значения R_{wp} -фактора; вверху слева – график роста R_{wp} -фактора при последовательном исключении из структуры $\text{Er}_{10}\text{W}_2\text{O}_{21}$ легких, а затем тяжелых атомов

На рисунке 39 показано соответствие расчетной и экспериментальной дифрактограмм для финального решения структуры $\text{Er}_{10}\text{W}_2\text{O}_{21}$ по МПГА. В таблице 20 приведено сравнение полученного решения данной структуры с истинными значениями атомных координат.

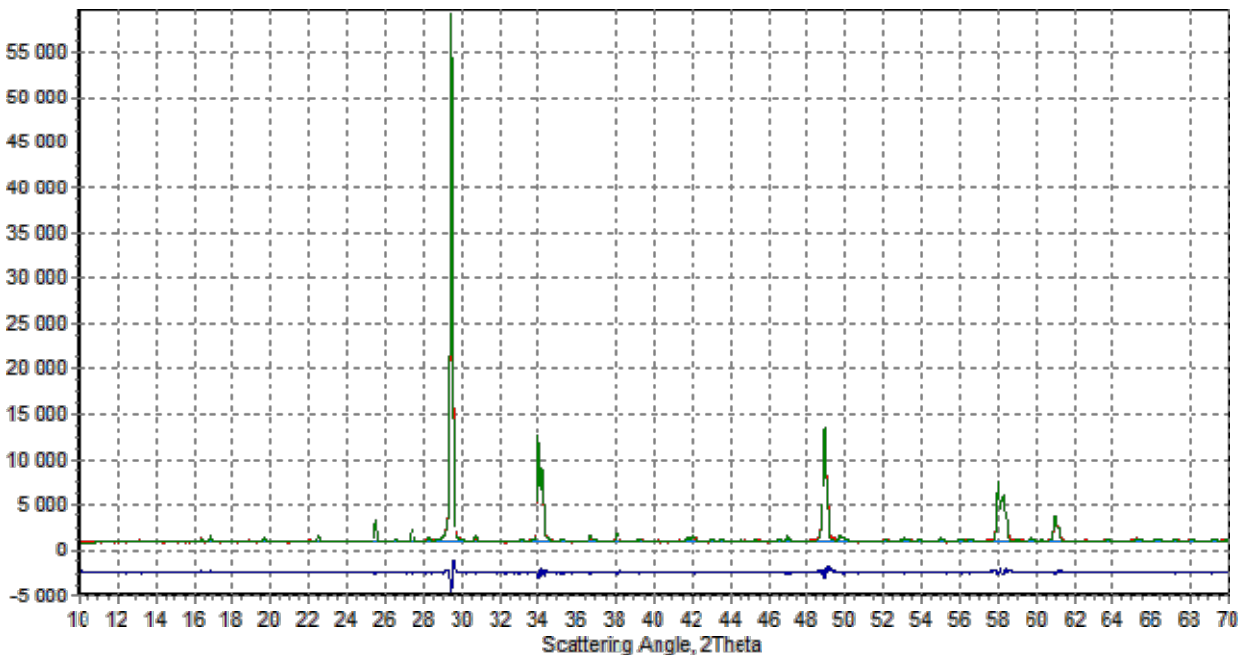


Рисунок 39 – Расчетная (зеленая) и экспериментальная (красная) дифрактограммы, и их разность (синяя) для найденной по МПГА структуры $\text{Er}_{10}\text{W}_2\text{O}_{21}$; R_{wp} -фактор 4.62%.

Таблица 20 – Сравнение атомных координат для истинной структуры $\text{Er}_{10}\text{W}_2\text{O}_{21}$, и найденной при помощи МПГА.

Ат.	X*	X _{МПГА}	\Delta	Y*	Y _{МПГА}	\Delta	Z*	Z _{МПГА}	\Delta
W1	0.08600	0.08630	0.00030	0.24900	0.24850	0.00050	0.87970	0.87825	0.00145
Er1	0.08060	0.08080	0.00020	0.00800	0.00716	0.00084	0.13640	0.13745	0.00105
Er2	0.27360	0.27431	0.00071	0.24980	0.24935	0.00045	0.12840	0.12866	0.00026
Er3	0.41290	0.41315	0.00025	0.01480	0.01572	0.00092	0.61190	0.61227	0.00037
Er4	0.10060	0.09920	0.00140	0.24740	0.24659	0.00081	0.34420	0.34403	0.00017
Er5	0.25510	0.25618	0.00108	0.00880	0.00873	0.00007	0.35880	0.35793	0.00087
O1	0.17200	0.16916	0.00284	0.08100	0.08049	0.00051	0.51300	0.51930	0.00630
O2	0.04900	0.04257	0.00643	0.10000	0.09112	0.00888	0.95700	0.96457	0.00757
O3	0.13900	0.13260	0.00640	0.11800	0.11285	0.00515	0.78600	0.78228	0.00372
O4	0.19500	0.19019	0.00480	0.09300	0.10721	0.01421	0.22500	0.21934	0.00566
O5	0.49100	0.49236	0.00136	0.15400	0.14295	0.01105	0.45800	0.45347	0.00453
O6	0.32400	0.32187	0.00213	0.21600	0.22477	0.00877	0.49800	0.49964	0.00164
O7	0.35500	0.34637	0.00863	0.12100	0.12348	0.00248	0.99500	0.98094	0.01406
O8	0.34800	0.35336	0.00536	0.11200	0.11017	0.00183	0.26200	0.27357	0.01157
O9	0.34100	0.34376	0.00276	0.15000	0.15235	0.00235	0.72800	0.73914	0.01114
O10	0.00000	0.00000	0.00000	0.21200	0.22288	0.01088	0.75000	0.75000	0.00000
O11	0.00000	0.00000	0.00000	0.12000	0.09247	0.02753	0.25000	0.25000	0.00000
O12	0.50000	0.50000	0.00000	0.12100	0.11626	0.00474	0.75000	0.75000	0.00000

Однако, в 10 независимых пусках FOX и DASH структура $\text{Er}_{10}\text{W}_2\text{O}_{21}$ не была определена *ab initio* (без задания структурных фрагментов) с точностью до позиций всех атомов кислорода, которые можно бы уточнить далее методом полнопрофильного анализа, ни разу. Вместе с тем, позиции тяжелых атомов определяются, как правило, достаточно хорошо. На рисунке 40 приведён графика сходимости для одного из пусков процесса поиска структуры $\text{Er}_{10}\text{W}_2\text{O}_{21}$ в программе FOX; на рисунке 41 приведены графики сходимости 10-и пусков в программе DASH.

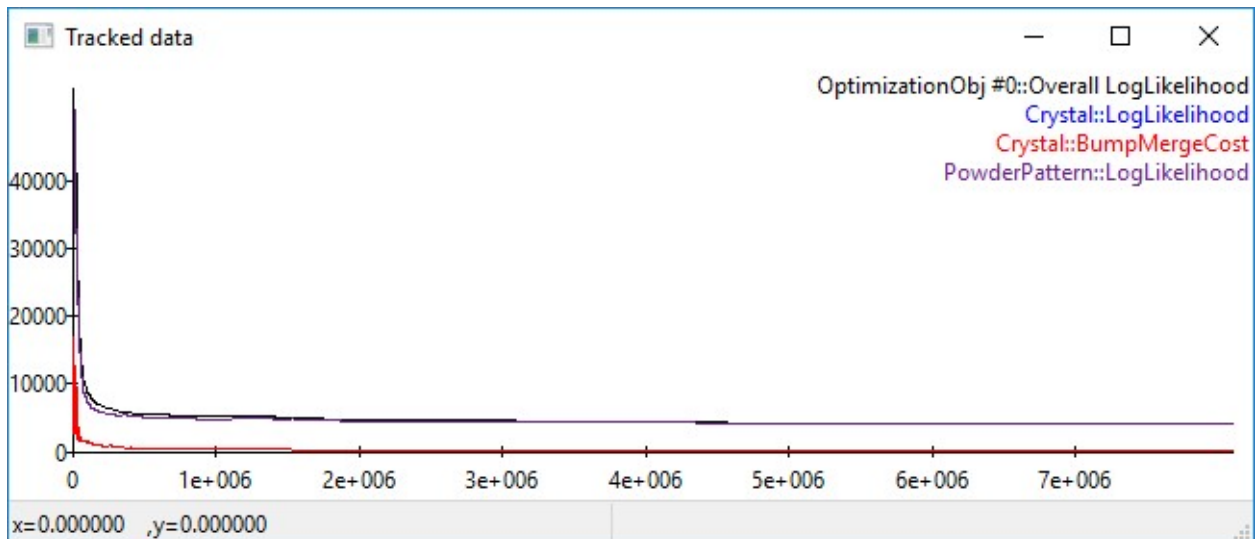


Рисунок 40 – Типичный график сходимости процесса решения структуры тестовой $\text{Er}_{10}\text{W}_2\text{O}_{21}$ в FOX. По оси абсцисс – число сгенерированных тестовых структур, по оси ординат – значение критерия качества тестовой структуры (χ^2).

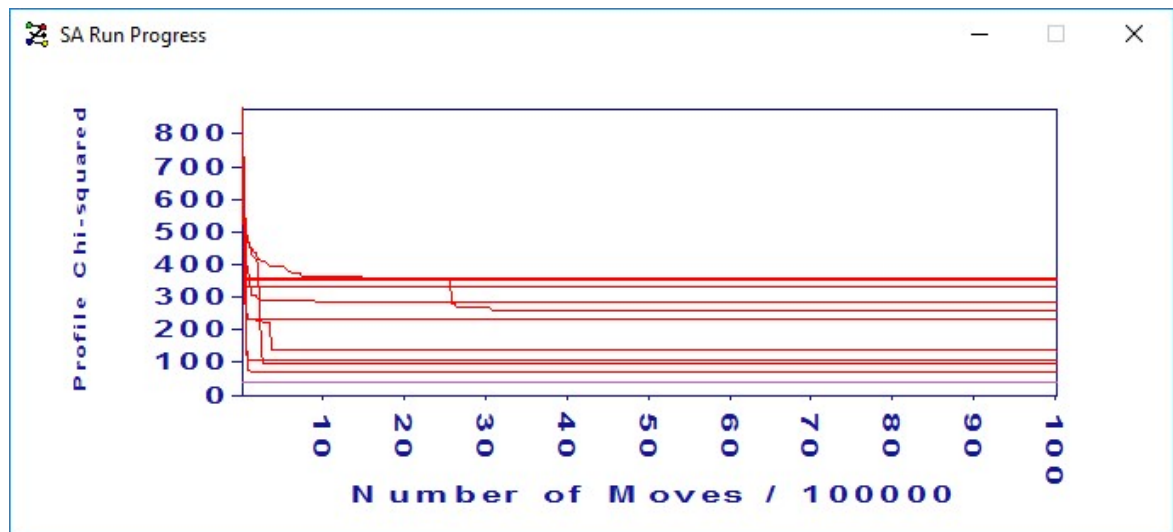


Рисунок 41 – Графики сходимости десяти пусков процесса поиска тестовой структуры $\text{Er}_{10}\text{W}_2\text{O}_{21}$

2.5.4 Выводы по результатам сравнения

В результате сравнительной экспериментальной оценки эффективности программ FOX и DASH с МПГА можно сделать вывод о том, что при *ab initio* поиске кристаллической структуры, когда не используются априори известные структурные фрагменты, МПГА демонстрирует определенные преимущества. Особенно интересным и перспективным является тестирование в процессе успешного поиска по МПГА относительно весьма малого числа структурных моделей (на 3-5 порядков меньше). Это позволяет выполнять при тестировании полнопрофильное уточнение большого числа различных промежуточных моделей, что повышает общую сходимости метода.

Глава 3. Определение неизвестных кристаллических структур при помощи МПГА

3.1 Описание объектов исследования

Объектами исследования являются кристаллические структуры комплексных соединений $[\text{Pd}(\text{CH}_3\text{NH}_2)_4][\text{PdBr}_4]$ и $[\text{Pt}(\text{NH}_3)_5\text{Cl}]\text{Br}_3$, синтезированные из стандартных химических реактивов чистотой не ниже “х.ч. – ч.д.а.” Синтез $[\text{Pd}(\text{CH}_3\text{NH}_2)_4][\text{PdBr}_4]$ был выполнен следующим образом. Концентрированный раствор H_2PdBr_4 , полученный растворением палладиевой черни в смеси бромистоводородной кислоты и брома, и упаренный до начала кристаллизации PdBr_2 , разбавили водой в 4 раза и охладили до 0°C . В полученный раствор добавили 10-15 %-ый водный раствор метиламина (CH_3NH_2), также охлажденный до 0°C , до достижения pH раствора 6-8. В результате образовался мелкокристаллический розовый осадок. Синтез $[\text{Pt}(\text{NH}_3)_5\text{Cl}]\text{Br}_3$ выполнен следующим образом. Комплексное соединение $[\text{Pt}(\text{NH}_3)_5\text{Cl}]\text{Cl}_3 \cdot \text{H}_2\text{O}$, полученное в соответствии с [140, с. 160-161], растворили в минимальном количестве воды при $30-40^\circ\text{C}$ и добавили концентрированный раствор бромида калия в мольном соотношении $\text{Pt} : \text{Br} = 1 : (3-4)$. Образовавшиеся осадки этих соединений отфильтровали, промыли водой и затем этанолом, и сушили на фильтре. Дифрактограммы получены с использованием $\text{CuK}\alpha$ излучения на дифрактометре X'Pert PRO с детектором PIXcel (PANalytical), снабженным графитовым монохроматором. Условия съемки: диапазон от 5 до 90° по шкале 2θ , шагом $0,026^\circ$, $\Delta t = 50$ с. В результате индирования дифрактограмм и определения пространственной группы симметрии при помощи программы EXPO [41] определены следующие кристаллографические характеристики этих соединений: $[\text{Pd}(\text{CH}_3\text{NH}_2)_4][\text{PdBr}_4]$ – пр. гр. $P4/mnc$, параметры ячейки $a=10.6866(7) \text{ \AA}$, $c=6.7262(3) \text{ \AA}$, $V=768.16(10) \text{ \AA}^3$, $Z=4$; $[\text{Pt}(\text{NH}_3)_5\text{Cl}]\text{Br}_3$ – пр. гр. $I41/a$, параметры ячейки $a=17.2587(5) \text{ \AA}$; $c=15.1164(3) \text{ \AA}$, $V=4502,61(10) \text{ \AA}^3$, $Z=16$. Соответствие всех позиций дифракционных рефлексов расчетным значениям, вычисленным по параметрам кристаллической решетки, подтвердило чистоту синтеза.

3.2 Описание методики исследования

Определение кристаллической структуры выполнялось с помощью МПГА на компьютере с процессором Intel i7-3770, имеющем 8 вычислительных ядер: 7 для процессов эволюционного поиска структуры в независимых популяциях и одно в качестве управляющего для регулярного накопления лучших структурных моделей из популяций и межпопуляционного обмена [141]. Был выбран вариант МПГА [142] с вещественным кодированием атомных координат и периодической локальной оптимизацией лучших структурных моделей по программе МНК метода

Ритвельда. Процесс эволюционного поиска структуры после запуска МПГА выполняется автоматически, контроль выполняется визуально по графикам сходимости, распределения атомов популяций в кристаллической ячейке и соответствия экспериментальной дифрактограммы с расчетными для лучших структурных моделей, получаемых на данном этапе эволюции. Для работы МПГА требуется предварительная настройка параметров поиска, от качества которой зависит вероятность сходимости структурных моделей к истинной структуре. Настройка была выполнена путем эмпирического выбора параметров в процессах поиска известных кристаллических структур соединений $[\text{Pd}(\text{NH}_3)_4](\text{C}_2\text{O}_4)$ и $[\text{Pt}(\text{NH}_3)_2(\text{C}_2\text{O}_4)]$, близких по составу к исследуемым. Выбранные настройки приведены в таблице 21.

Таблица 21 – Настройки параметров МПГА, выбранные для поиска структур

Параметр	Значение
Размер популяций (количество структурных моделей)	300
Регулярность запуска локального поиска по МНК, поколений	10
Циклов МНК при запуске локального поиска	5
Количество структурных моделей, которые подвергались локальному поиску при каждом запуске МНК	3
Регулярность отправки структурных моделей с управляющего ядра на рабочие, поколений	9
Количество отправляемых структурных моделей	2
Количество поколений для локального элитизма (временного сохранения в популяции лучшей структурной модели)	3
Размер турнира (количество структурных моделей для турнирного отбора при селекции)	3
Коэффициент скрещивания (доля структурных моделей, участвующих в вероятностной парной рекомбинации их атомов)	0,75
Коэффициент мутации атомных координат структурной модели	1,5

После сходимости МПГА к значениям R-фактора, близким к целевому значению, выполняется кристаллохимическая верификация найденной структуры, включающая оценку корректности межатомных расстояний, валентных углов и т.п., расстановку атомов водорода, уточнение дифрактограммы по методу Ритвельда, а также контроль корректности структуры при помощи on-line сервиса IUCr CheckCIF/PLATON [143].

3.3 Определение структуры соединения $[\text{Pt}(\text{NH}_3)_5\text{Cl}]\text{Br}_3$

В результате уточнения по методу Ле Бэйла определены профильные параметры дифрактограммы и целевое значение профильного R-фактора, равное 5.49%. По данным статистики распределения межатомных расстояний в структурах разных соединений аналогичного состава, рассчитанной по программе Diamond [144], были наложены ограничения на длины межатомных

расстояний, приведенные в таблице 22. Аналогично установлены ограничения на минимальные межатомные расстояния для других пар атомов.

Таблица 22 – Значения межатомных расстояний для атомов, координированных к Pt

Межатомные расстояния, Å:	Pt-Cl	Pt-N	Pt-Br	Pt-Pt
Минимально допустимое расстояние	2,15	1,8	3,95	4
Типичное расстояние для химического соединения	2,32	2,15	-	
Диапазон возможного отклонения от типичного меж- атомного расстояния	±0,05	±0,08	-	

При поиске структуры по МПГА искомыми параметрами являлось расположение в ячейке трех атомов Br и априори известного структурного фрагмента PtN₅Cl. На этом этапе атомы водорода не учитывались, а параметры изотропных тепловых колебаний атомов были взяты из структур близкого состава. Перечень параметров приведен в таблице 23.

Таблица 23 – Степени свободы структурных параметров [Pt(NH₃)₅Cl]Br₃ в поиске по МПГА

Характеристика структуры	Кол-во параметров
Координаты атома Br1	3
Координаты атома Br2	3
Координаты атома Br3	3
Координаты и углы, определяющего ориентацию фрагмента PtN ₅ Cl	7
Общее число степеней свободы структуры	16

На рисунке 42 представлены графики сходимости МПГА при поиске этой структуры. По оси абсцисс указаны номера поколений эволюции (нулевой номер – начальная случайная генерация всех популяций), по оси ординат – соответствующие значения функции пригодности (R-фактор плюс штраф за нарушение структурных ограничений) лучшей структурной модели из числа накопленных на управляющем ядре. Красная линия показывает значение пригодности лучшей из найденных структурных моделей. Чёрная линия – значение лучшей пригодности, полученной после уточнения лучшей структурной модели по МНК (запаздывает, если установлен режим выполнения МНК раз в несколько поколений). Фиолетовая линия – вклад штрафов в значениях функции пригодности. Синяя – среднее значение пригодности структурных решений, накопленных на управляющем ядре. Зелёная линия – худшее из лучших решений структуры на рабочих ядрах. Совокупность графиков позволяет анализировать и контролировать процесс сходимости.



Рисунок 42 – График сходимости при поиске структуры $[\text{Pt}(\text{NH}_3)_5\text{Cl}]\text{Br}_3$ по МПГА; по оси X - номер поколения; по оси Y – значение функции пригодности; пригодность лучшей структурной модели - красная линия, величина штрафов - фиолетовая.

Из рисунка 42 видно, что уже к 40-му поколению эволюции достигнуто значение пригодности 15% и структура в основном сформировалась. Однако, она еще недостаточно корректна, т.к. часть межатомных расстояний в ней выходят за ограничения (уровень штрафов около 3%). К сотому поколению штрафы исчезают и все межатомные расстояния становятся корректными, а пригодность совпадает со значениями R-фактора. Далее, R-фактор постепенно снижается с 13,5 до 12%, в основном, за счет уточнения ориентации фрагмента PtN_5Cl , и к 240-му поколению процесс сходимости завершается. При этом, лучшие структурные модели постепенно распространяются по популяциям на рабочих ядрах (минимизируется зелёная линия). Следует отметить, что практически все резкие спады пригодности происходят за счет локальных оптимизаций структурных моделей по МНК (красные линии на них совпадают с черными).

Полученная модель структуры уточнена по программе полнопрофильного анализа МПР вместе с параметрами тепловых колебаний и получен R-фактор 7,64 %, который оказался выше целевого на 2,15%. Модель найденной по МПГА и уточненной по МПР кристаллической структуры соединения $[\text{Pt}(\text{NH}_3)_5\text{Cl}]\text{Br}_3$ приведена в табл. 24, график сравнения экспериментальной дифрактограммы с уточненной по МПР – на рисунке 43.

Таблица 24 – Модель кристаллической структуры соединения $[\text{Pt}(\text{NH}_3)_5\text{Cl}]\text{Br}_3$, найденная по МПГА и уточненная по МПР (неводородные атомы)

Атом	X	Y	Z	Uiso	Occup.
Pt	0.24598(12)	0.26843(8)	0.50305(19)	0.0216(14)	1.0000
Br1	0.12448(26)	0.28273(31)	0.23106(49)	0.0497(23)	1.0000
Br2	0.28436(25)	0.52807(23)	0.48582(27)	0.0439(23)	1.0000
Br3	0.15066(33)	0.30736(33)	0.7663(5)	0.0956(38)	1.0000
Cl	0.1291(6)	0.32802(44)	0.4930(11)	0.0347(45)	1.0000
N1	0.3510(13)	0.2165(14)	0.5194(19)	0.0008(39)	1.0000
N2	0.2174(17)	0.1957(15)	0.4096(18)	0.0008(39)	1.0000
N3	0.2772(19)	0.3507(15)	0.6016(16)	0.0008(39)	1.0000
N4	0.2074(17)	0.1918(15)	0.5978(15)	0.0008(39)	1.0000
N5	0.2825(16)	0.3421(15)	0.4087(16)	0.0008(39)	1.0000

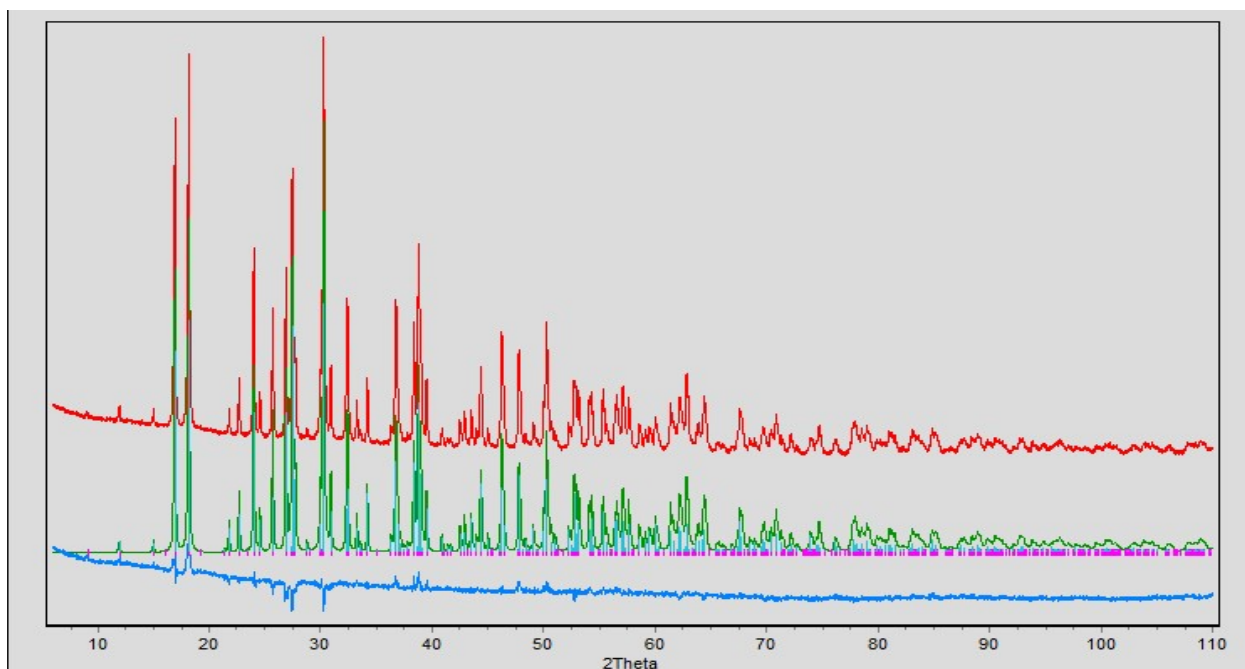


Рисунок 43 – График сравнения экспериментальной дифрактограммы $[\text{Pt}(\text{NH}_3)_5\text{Cl}]\text{Br}_3$ с уточнением по МПР; красная линия обозначает экспериментальную дифрактограмму, зелёная – рассчитанную, синяя - разность между ними; R-фактор 7.639 %

Далее, в структуру были добавлены положения атомов водорода, кристаллохимически связанные с азотами в группах NH_3 (взяты из БД ICSD для близких структур), и выполнено аналогичное повторное уточнение по МПР. При этом, водороды жестко связывались с азотами, так, что уточнялись позиции групп NH_3 . Модель уточненной по МПР кристаллической структуры соединения $[\text{Pt}(\text{NH}_3)_5\text{Cl}]\text{Br}_3$ с включением в нее атомов водорода приведена в табл. 25,

Таблица 25 – Модель кристаллической структуры соединения $[\text{Pt}(\text{NH}_3)_5\text{Cl}]\text{Br}_3$, найденная по МПГА и уточненная по МПР (с включением атомов водорода)

Атом	X	Y	Z	Uiso	Occup.
Pt	0.24599(11)	0.26851(8)	0.50289(17)	0.0207(15)	1.0000
Br1	0.12401(24)	0.28197(31)	0.23115(45)	0.0478(22)	1.0000
Br2	0.28422(24)	0.52821(21)	0.48550(26)	0.0410(22)	1.0000
Br3	0.15045(31)	0.30746(32)	0.76715(45)	0.0867(33)	1.0000
Cl	0.1280(6)	0.32726(41)	0.4925(10)	0.0283(41)	1.0000
N1	0.3517(13)	0.2154(13)	0.5133(23)	0.0120(43)	1.0000
N2	0.2208(19)	0.1944(14)	0.4062(17)	0.0120(43)	1.0000
N3	0.2779(19)	0.3500(13)	0.5967(16)	0.0120(43)	1.0000
N4	0.2083(18)	0.1939(14)	0.5979(14)	0.0120(43)	1.0000
N5	0.2851(15)	0.3418(13)	0.4111(15)	0.0120(43)	1.0000
H11	0.39	0.24	0.48	0.0380	1.0000
H12	0.36	0.21	0.57	0.0380	1.0000
H13	0.35	0.16	0.49	0.0380	1.0000
H21	0.26	0.16	0.39	0.0380	1.0000
H22	0.18	0.17	0.43	0.0380	1.0000
H23	0.21	0.22	0.35	0.0380	1.0000
H31	0.33	0.34	0.61	0.0380	1.0000
H32	0.24	0.35	0.65	0.0380	1.0000
H33	0.27	0.40	0.57	0.0380	1.0000
H41	0.16	0.21	0.61	0.0380	1.0000
H42	0.24	0.20	0.65	0.0380	1.0000
H43	0.21	0.14	0.58	0.0380	1.0000
H51	0.24	0.38	0.39	0.0380	1.0000
H52	0.33	0.37	0.44	0.0380	1.0000
H53	0.30	0.31	0.36	0.0380	1.0000

В результате ввода атомов водорода (в группах NH_3) и уточнения по МПР профильный R-фактор был снижен до 6,89 %, но также был заметно выше целевого – на 1,9%, и при этом атомы Br (особенно, Br3) имели слишком большие значения коэффициентов изотропных тепловых колебаний. На основании этого был сделан вывод о том, что в атомных позициях Br происходит статистическое замещение части его атомов на более лёгкие атомы Cl. Для учета этого в позиции 3-х атомов Br в структуре были добавлены 3 атома Cl (Cl1 – Cl3 с такими же координатами), а их коэффициенты заполнения позиций N были функционально связаны с коэффициентами заполнения позиций соответствующих атомов Br. На рисунке 44 представлены установки МПР для уточнения нестехиометрического варианта структуры $[\text{Pt}(\text{NH}_3)_5\text{Cl}]\text{Br}_{3-x}\text{Cl}_x$.

Name	Type	X	Y	Z	N	Biso	Dist	Aniso
Pt	Pt+4	0.245790	0.268408	0.502131	16.00000	1.586708	2.350	<input type="checkbox"/>
Br1	Br-1	0.125618	0.282629	0.231992	14.36683	2.400329	0.000	<input type="checkbox"/>
Br2	Br-1	0.283708	0.527841	0.485314	14.76441	2.770951	0.000	<input type="checkbox"/>
Br3	Br-1	0.149428	0.308202	0.766927	9.44091	2.356355	0.000	<input type="checkbox"/>
Cl1	Cl-1	0.125618	0.282629	0.231992	1.63317	2.400329	0.000	<input type="checkbox"/>
Cl2	Cl-1	0.283708	0.527841	0.485314	1.23559	2.770951	0.000	<input type="checkbox"/>
Cl3	Cl-1	0.149428	0.308202	0.766927	6.55909	2.356355	0.000	<input type="checkbox"/>
Cl4	Cl-1	0.126060	0.325580	0.493591	16.00000	2.700477	0.000	<input type="checkbox"/>
N1	N	0.353691	0.216381	0.519378	16.00000	1.223119	0.000	<input type="checkbox"/>
N2	N	0.213213	0.190200	0.406011	16.00000	1.223119	0.000	<input type="checkbox"/>
N3	N	0.279471	0.350369	0.594746	16.00000	1.223119	0.000	<input type="checkbox"/>
N4	N	0.200655	0.195169	0.596677	16.00000	1.223119	0.000	<input type="checkbox"/>
N5	N	0.287456	0.342279	0.408421	16.00000	1.223119	0.000	<input type="checkbox"/>
H11	H	0.392396	0.243787	0.487459	16.00000	3.000000	0.000	<input type="checkbox"/>
H12	H	0.366796	0.215387	0.580359	16.00000	3.000000	0.000	<input type="checkbox"/>
H13	H	0.350396	0.164787	0.497559	16.00000	3.000000	0.000	<input type="checkbox"/>
H21	H	0.254513	0.155100	0.394411	16.00000	3.000000	0.000	<input type="checkbox"/>
H22	H	0.169713	0.162300	0.427511	16.00000	3.000000	0.000	<input type="checkbox"/>
H23	H	0.199513	0.216300	0.352911	16.00000	3.000000	0.000	<input type="checkbox"/>
H31	H	0.331271	0.340869	0.612846	16.00000	3.000000	0.000	<input type="checkbox"/>
H32	H	0.245971	0.346569	0.644346	16.00000	3.000000	0.000	<input type="checkbox"/>
H33	H	0.275971	0.400969	0.570146	16.00000	3.000000	0.000	<input type="checkbox"/>
H41	H	0.147655	0.208069	0.604377	16.00000	3.000000	0.000	<input type="checkbox"/>
H42	H	0.227655	0.202969	0.650777	16.00000	3.000000	0.000	<input type="checkbox"/>
H43	H	0.204955	0.142369	0.579577	16.00000	3.000000	0.000	<input type="checkbox"/>
H51	H	0.247356	0.375879	0.388921	16.00000	3.000000	0.000	<input type="checkbox"/>
H52	H	0.328056	0.371879	0.434021	16.00000	3.000000	0.000	<input type="checkbox"/>
H53	H	0.307056	0.313779	0.359521	16.00000	3.000000	0.000	<input type="checkbox"/>

Рисунок 44 – Установки МПР для уточнения нестехиометрической структуры $[\text{Pt}(\text{NH}_3)_5\text{Cl}]\text{Br}_3 \cdot x\text{Cl}_x$; уточняемые параметры отмечены кнопками: независимые – зелеными, функционально связанные – коричневыми (цифра на кнопке – номер связи); в окнах слева – результаты уточнения отмеченных параметров

Финальный вариант уточнения по МПР имеет профильный R-фактор 5,64 %, весьма хорошо соответствующий целевому (5.49 %). На рисунке 45 приведено сравнение экспериментальной дифрактограммы с полученной в результате финального уточнения. Координаты атомов, их изотропные тепловые параметры и коэффициенты заполнения кристаллографических позиций приведены в таблице 26. Химическая формула соединения, рассчитанная по данным таблицы 26, имеет вид $[\text{Pt}(\text{NH}_3)_5\text{Cl}]\text{Br}_{2.4}\text{Cl}_{0.6}$.

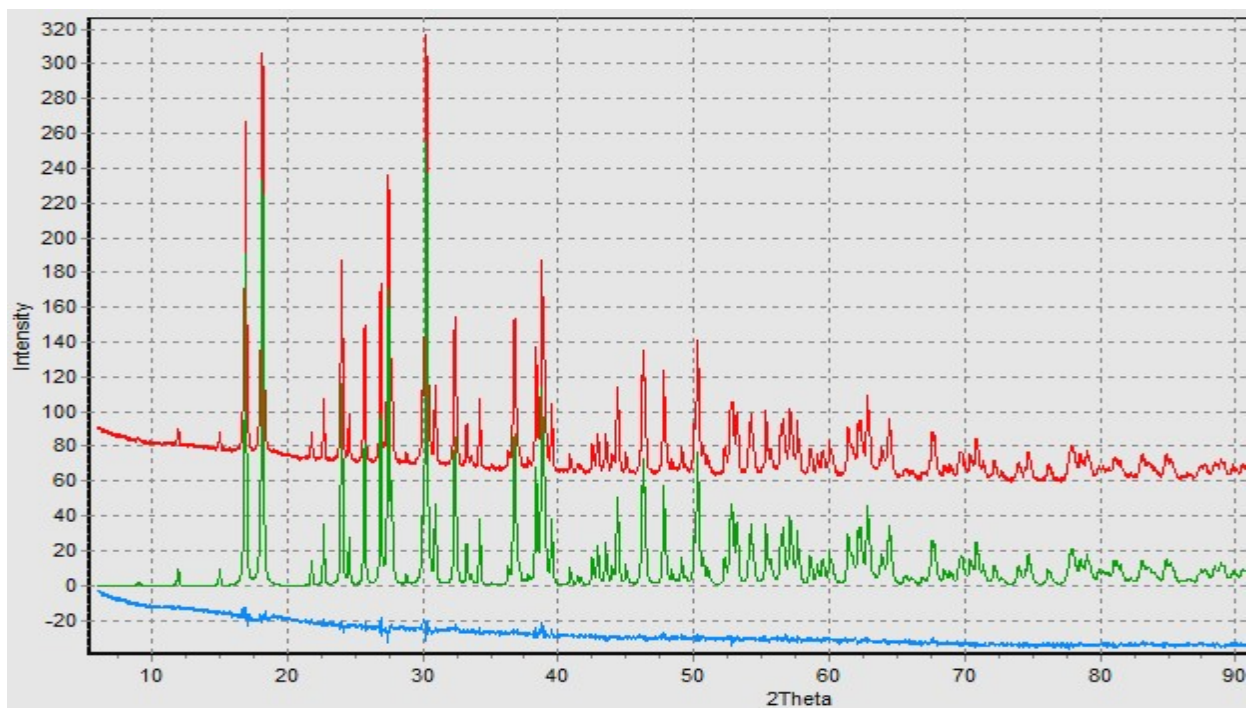


Рисунок 45 – График сравнения экспериментальной дифрактограммы $[\text{Pt}(\text{NH}_3)_5\text{Cl}]\text{Br}_3$ с уточнением по DDM; красная линия обозначает экспериментальную дифрактограмму, зелёная – рассчитанную, синяя - разность между ними; R-фактор 5.637 %.

Таблица 26 – Кристаллическая структура соединения $[\text{Pt}(\text{NH}_3)_5\text{Cl}]\text{Br}_{2.4}\text{Cl}_{0.6}$

Атом	X	Y	Z	Uiso	Occup.
Pt	0.24579(10)	0.26841(7)	0.50209(15)	0.015(6)	1.0
Br1	0.12571(23)	0.28267(24)	0.23195(38)	0.026(6)	0.896(12)
Cl1	0.12571(23)	0.28267(24)	0.23195(38)	0.026(6)	0.104(12)
Br2	0.28375(21)	0.52788(18)	0.48527(24)	0.030(6)	0.921(9)
Cl2	0.28375(21)	0.52788(18)	0.48527(24)	0.030(6)	0.079(9)
Br3	0.14937(29)	0.30820(26)	0.76687(42)	0.025(6)	0.587(12)
Cl3	0.14937(29)	0.30820(26)	0.76687(42)	0.025(6)	0.413(12)
Cl4	0.1262(5)	0.32560(38)	0.4936(9)	0.030(6)	1.0
N1	0.3534(10)	0.2167(11)	0.5192(16)	0.011(7)	1.0
N2	0.2130(17)	0.1903(12)	0.4063(12)	0.011(7)	1.0
N3	0.2792(15)	0.3501(13)	0.5946(13)	0.011(7)	1.0
N4	0.2008(15)	0.1954(13)	0.5960(12)	0.011(7)	1.0
N5	0.2872(14)	0.3424(12)	0.4083(13)	0.011(7)	1.0
H11	0.39(1)	0.24(1)	0.49(1)	0.038	1.0
H12	0.37	0.22	0.58	0.038	1.0
H13	0.35	0.17	0.49	0.038	1.0
H21	0.25	0.16	0.34	0.038	1.0
H22	0.17	0.16	0.43	0.038	1.0
H23	0.20	0.22	0.35	0.038	1.0
H31	0.33	0.34	0.61	0.038	1.0
H32	0.25	0.35	0.64	0.038	1.0
H33	0.28	0.40	0.57	0.038	1.0
H41	0.15	0.21	0.60	0.038	1.0
H42	0.23	0.20	0.65	0.038	1.0
H43	0.21	0.14	0.58	0.038	1.0
H51	0.25	0.38	0.39	0.038	1.0
H52	0.33	0.37	0.43	0.038	1.0
H53	0.30	0.31	0.36	0.0380	1.0

Схема расположения атомов кристаллической структуры соединения $[\text{Pt}(\text{NH}_3)_5\text{Cl}]\text{Br}_{2.4}\text{Cl}_{0.6}$ изображена на рисунке 46.

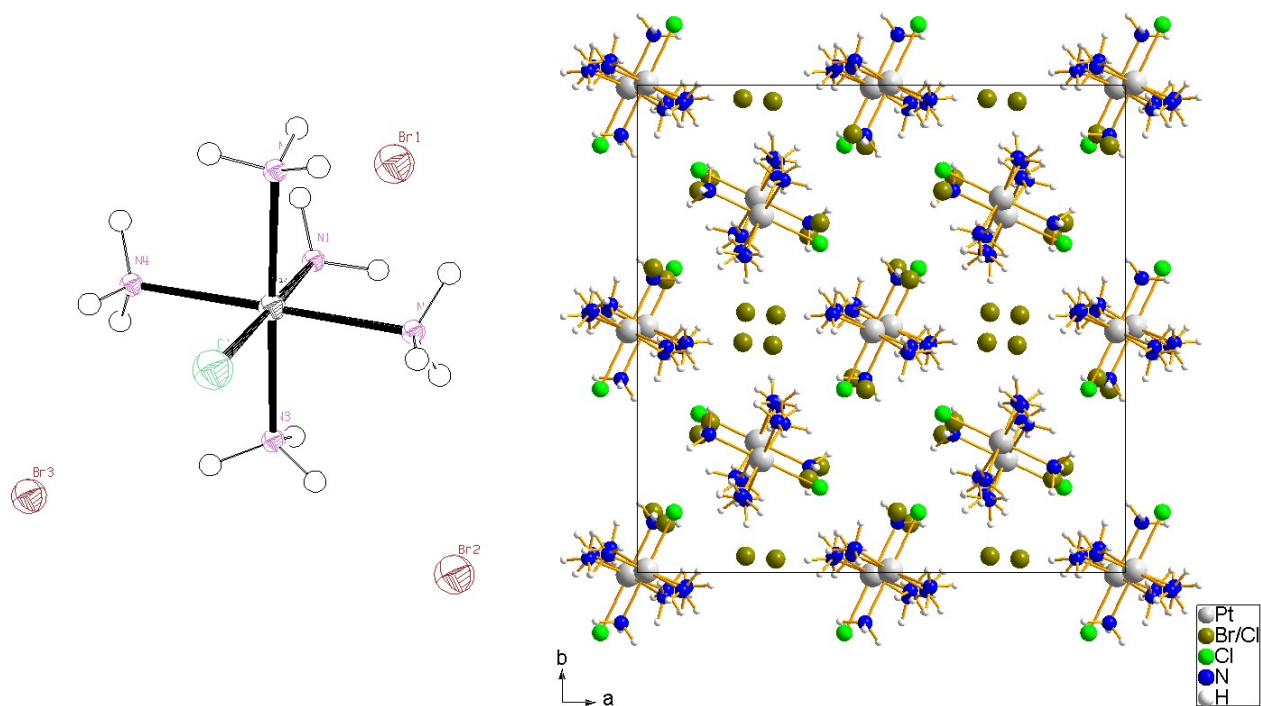


Рисунок 46 – Кристаллическая структура соединения $[\text{Pt}(\text{NH}_3)_5\text{Cl}]\text{Br}_{2.4}\text{Cl}_{0.6}$: слева – комплексный катион $[\text{Pt}(\text{NH}_3)_5\text{Cl}]^{3+}$ в окружении трех независимых анионов Br^- ; справа – проекция структуры на плоскость ab кристаллической ячейки

Кристаллическая структура построена из комплексных катионов $[\text{Pt}(\text{NH}_3)_5\text{Cl}]^{3+}$ и трех кристаллографически независимых анионов Br^- . Все атомы располагаются в общих позициях. Комплексный катион из-за присутствия в нем атома хлора имеет вид неправильного октаэдра с длинами связей: $\text{Pt}-\text{Cl}4 = 2,291 \text{ \AA}$; $\text{Pt}-\text{N}1 = 2,077 \text{ \AA}$; $\text{Pt}-\text{N}2 = 2,058 \text{ \AA}$; $\text{Pt}-\text{N}3 = 2,067 \text{ \AA}$; $\text{Pt}-\text{N}4 = 2,051 \text{ \AA}$; $\text{Pt}-\text{N}5 = 2,038 \text{ \AA}$. Атомы $\text{N}1$ и $\text{Cl}4$ лежат в вершинах октаэдра, остальные четыре азота в его основании. Углы $\text{N}-\text{Pt}-\text{N}$ в основании октаэдра отклоняются от прямых и меняются от $86,9^\circ$ до $93,2^\circ$. Угол между атомами в вершинах октаэдра $\text{N}1-\text{Pt}-\text{Cl}4$ составил 176° . Полученные длины связи хорошо соответствуют межатомным расстояниям в близких структурах. Хорошее соответствие длин межатомных связей статистике их распределения в структурах аналогичного состава и низкое значение профильного R-фактора свидетельствуют об адекватности найденной кристаллической структуры этого соединения. Контроль с помощью on-line сервиса IUCr CheckCIF/PLATON подтвердил корректность структуры.

3.4 Определение структуры соединения $[Pd(CH_3NH_2)_4][PdBr_4]$

В результате уточнения по методу Ле Бэйла определены профильные параметры дифрактограммы и целевое значение R-фактора, равное 6,88%. По данным статистики распределения межатомных расстояний введены ограничения, приведенные на рисунке 47.

Power 1	Power 2	Min. dist.
Pd	Pd	3
Pd	N	1.85
Pd	C	3
Pd	Br	2.2
N	N	2.5
N	C	1.4
N	Br	3.5
C	C	3
C	Br	3.5
Br	Br	3

Рисунок 47 – Задание минимальных межатомных расстояний в $[Pd(CH_3NH_2)_4][PdBr_4]$

Исходя из пр. гр. ($P4/mnc$) и ячейки ($V=768.16 \text{ \AA}^3$, $Z=4$) было сделано предположение, подтвержденное в результате применения метода Паттерсона, что тяжелые атомы палладия лежат на грани перпендикулярной к оси c , причём один из них в начале координат, другой – в центре грани, а лёгкие неводородные атомы также расположены на этой грани. С учётом этого число степеней свободы атомных координат при поиске структуры составило 6. Перечень структурных параметров для поиска данной структуры по МПГА приведен в таблице 27.

Таблица 27 – Степени свободы структурных параметров $[Pd(CH_3NH_2)_4][PdBr_4]$ при поиске структуры по МПГА

Характеристика структуры	Кол-во параметров
Координаты X и Y атома C1	2
Координаты X и Y атома N1	2
Координаты X и Y атома Br1	2
Общее число степеней свободы структуры	6

На рисунке 48 представлены графики сходимости МПГА при поиске этой структуры. Из рисунка 48 видно, что уже к 10 поколению эволюции достигнуто значение пригодности 13% и структура в основном сформировалась. Далее происходит уточнение позиций атомов C и N, и к 23-му поколению процесс сходимости практически завершается.

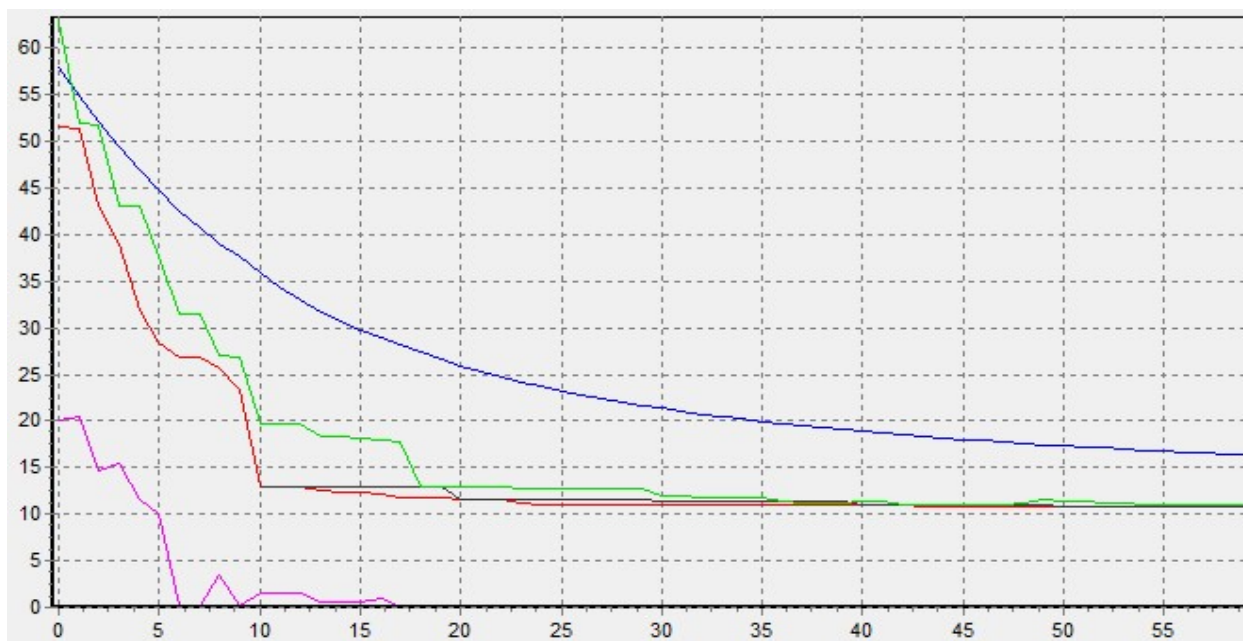


Рисунок 48 – График сходимости при поиске структуры $[Pd(CH_3NH_2)_4][PdBr_4]$ по МПГА; по оси X - номер поколения; по оси Y – значение функции пригодности; пригодность лучшей структурной модели - красная линия, величина штрафов - фиолетовая.

Полученная модель структуры была уточнена по программе МПР, проведено совместное уточнение всех профильных и структурных параметров, а также параметра текстуры для ее эллипсоидной модели (в соответствии с сингонией соединения). Всего было уточнено 19 параметров (10 профильных, 3 параметра текстуры и 6 координат атомов), в итоге был получен R-фактор 7,67%. После этого, в структуру были добавлены положения атомов водорода, кристаллохимически связанные с азотами и углеродами в группах CH_3 и NH_2 (из БД ICSD), а позиции групп уточнены по МПР. Финальный вариант решения имеет профильный R-фактор 7,49%, достаточно близкий к целевому. На рисунке 49 приведено сравнение экспериментальной дифрактограммы с полученной в результате уточнения.

Найденные по МПГА и уточненные по DDM структурные данные (координаты атомов x, y, z и коэффициенты заполнения атомных позиций N) для независимой части ячейки $[Pd(CH_3NH_2)_4][PdBr_4]$ приведены в таблице 28. Схема расположения атомов структуры приведена на рисунке 50.

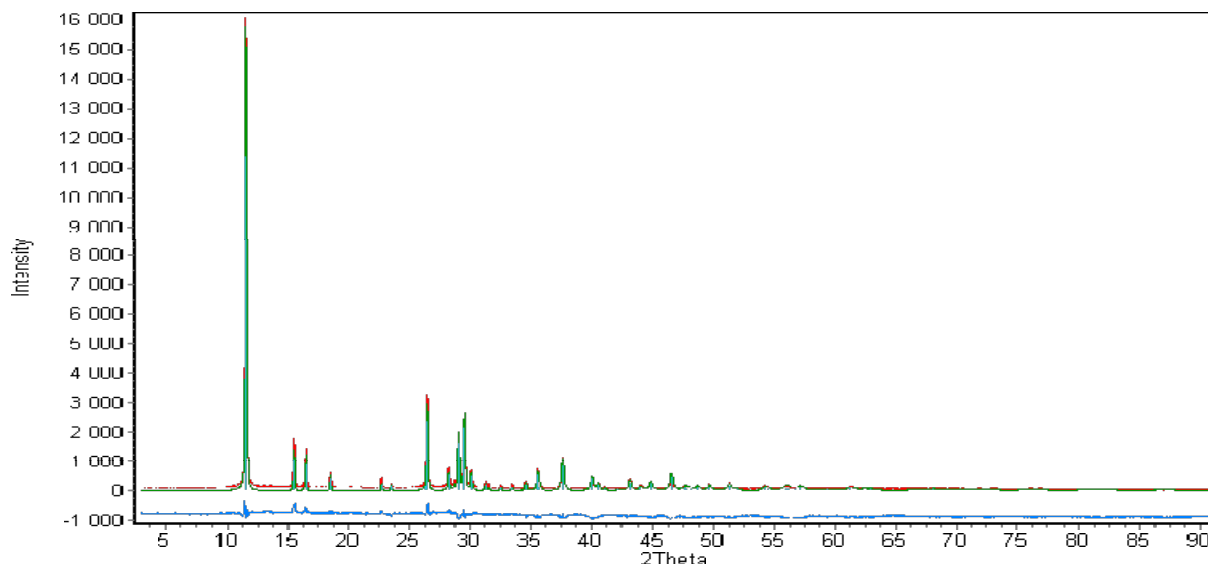


Рисунок 49 – Экспериментальная и расчетная дифрактограммы $[Pd(CH_3NH_2)_4][PdBr_4]$ после уточнения по DDM, внизу их разница; $R_{wp} = 7,49\%$

Таблица 28 – Кристаллическая структура $[Pd(CH_3NH_2)_4][PdBr_4]$

Атом	x	y	z	Uiso	Occup.
<i>Pd1</i>	0	0	0	0.036(6)	2
<i>Pd2</i>	0.5	0.5	0	0.032(8)	2
<i>N1</i>	0.152(7)	0.117(8)	0	0.02	8
<i>H1</i>	0.15	0.17	0.10	0.02	16
<i>Cl</i>	0.267(12)	0.044(8)	0	0.02	8
<i>H2</i>	0.27	-0.01	0.12	0.02	16
<i>H3</i>	0.33	0.10	0	0.02	8
<i>Br1</i>	0.2997(8)	0.4007(9)	0	0.0273(33)	8

Кристаллическая структура содержит две плоские квадратные части: $[Pd(CH_3NH_2)_4]$ и $[PdBr_4]$, которые располагаются параллельно друг относительно друга. Группа метиламин координируется к палладию через азот с расстоянием около 2,0542(1) Å (угол *Pd-N-C* 109,798(2)). Расстояние *N-C* составляет 1,4561(1) Å; *Pd-Br*, соответственно, 2,3894(1) Å. Локализованы все атомы водорода. Длина связи *C...H* составляет 0,96 Å, углы в группе *CH₃* 111,02° и 109,94°. Длина связи *N...H* составляет 0,87 Å, угол, соответственно, 106,2°. Полученные длины связи хорошо соответствуют межатомным расстояниям в близких структурах (таблица 29).

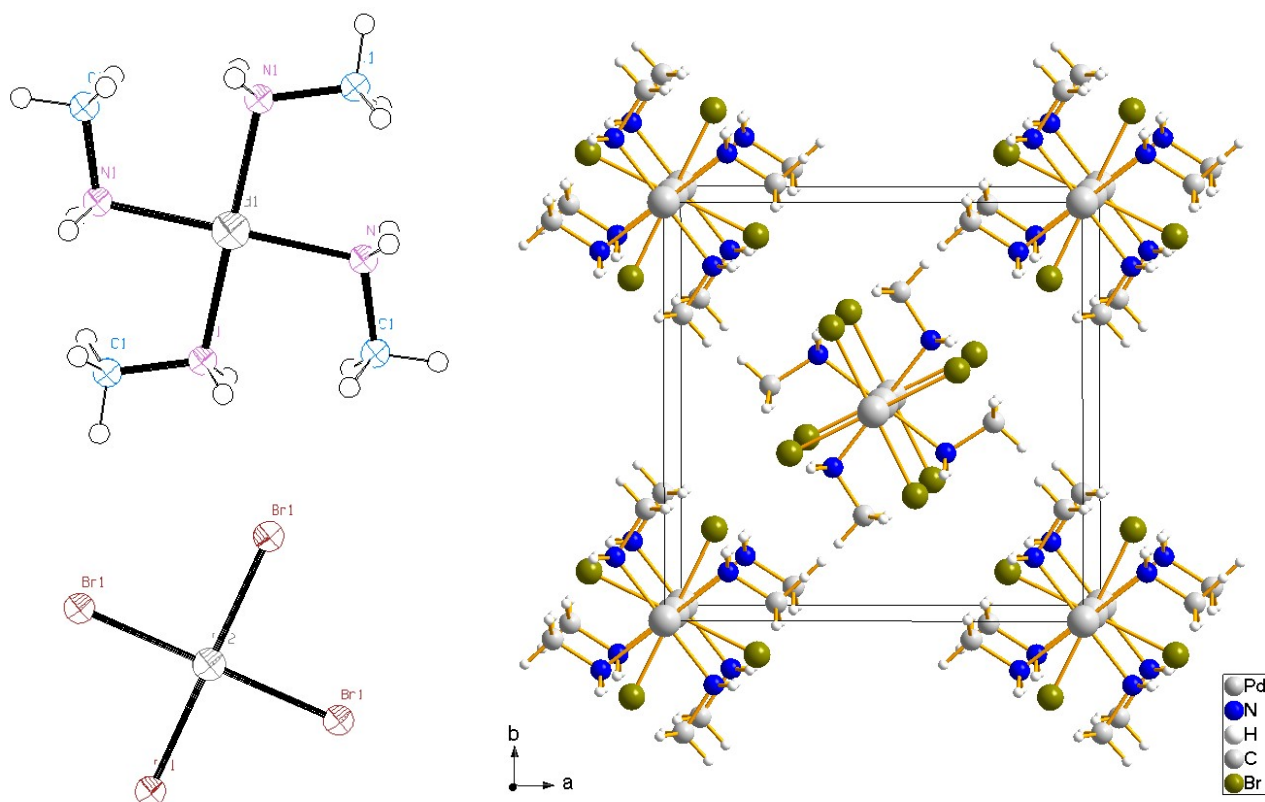


Рисунок 50 – Кристаллическая структура $[Pd(CH_3NH_2)_4][PdBr_4]$:
слева – плоско-параллельно фрагменты $[Pd(CH_3NH_2)_4]$ и $[PdBr_4]$;
справа – проекция структуры на плоскость ab кристаллической ячейки

Таблица 29 – Сравнение длин связей в $[Pd(CH_3NH_2)_4][PdBr_4]$ с их длинами в других структурах, взятых из базы структурных данных *ICSD*, Å.

Формула\№ карточки <i>ICSD</i>	<i>Pd-Br</i>	<i>Pd-N</i>	<i>N-C</i>	<i>C-H</i>	<i>N-H</i>
$[Pd(CH_3NH_2)_4][PdBr_4]$	2,389	2,054	1,456	0,961...0,963	0,865
$[Cr(NH_3)_5Cl][PdBr_4]$ / №158877	2,413				
$(NH_4)_2(PdBr_4)$ / №202799	2,437				
$[(CH_3)_2NH_2]_2[PdBr_4]$ / №110463	2,434		1,431		
$Pd(NH_3)_2Br_2$ / №280377	2,246; 2,440	2,037			
$(NH_3)_4Pd(N_3)_4Pd$ / №410384		2,056			0,912
$(NH_2(CH_3))(BH_3)$ / №249797			1,475	0,959...0,969	0,882
$N(CH_3)_4CuZn(CN)_4$ / №249948				0,979	

Соответствие длин межатомных связей, в т.ч. для водородных атомов, и низкое значение профильного R-фактора свидетельствуют об адекватности найденной кристаллической структуры соединения $[Pd(CH_3NH_2)_4][PdBr_4]$. Контроль с помощью on-line сервиса IUCr CheckCIF/PLATON подтвердил корректность структуры.

Выводы и основные результаты работы

1. Разработан оригинальный мультипопуляционный параллельный генетический алгоритм для эволюционного поиска и полнопрофильного уточнения моделей кристаллической структуры по дифрактограмме в прямом пространстве. Разработано математическое обеспечение МПГА, реализующее мультипопуляционный генетический алгоритм с использованием параллельных вычислений на многоядерных ПК и суперкомпьютерных кластерах.

2. В результате определения 11-ти известных кристаллических структур разной сложности с помощью МПГА изучены характеристики сходимости и надежности мультипопуляционного алгоритма. Показано, что увеличение числа вычислительных ядер позволяет определять более сложные структуры, надежность определения структур на 4-х ядерном ПК в 2–3 раза выше, чем в однопопуляционном ГА, а на суперкомпьютерном кластере – в 2–4 раза выше, чем на 4-х ядерном ПК. Так наиболее сложная из тестовых структур $\text{Er}_{10}\text{W}_2\text{O}_{21}$ (пр. гр. $Pbcn$) с 54 степенями свободы атомных координат однопопуляционным ГА не определена ни разу в 20-ти независимых пусках программы на ПК, по МПГА на 4-х ядерной ПК (коэволюция 3-х популяций) была определена в 10%, а на кластере с использованием 24-х ядер (коэволюция 23-х популяций) – в 40% пусков.

3. Экспериментальное сравнение МПГА с 2 современными программами глобальной оптимизации в прямом пространстве FOX и DASH (метод имитации отжига) показало, что при *ab initio* поиске тестовых структур (когда не используются априори известные структурные фрагменты), МПГА демонстрирует несколько лучшие показатели эффективности. В частности, тестовая структура $\text{Er}_{10}\text{W}_2\text{O}_{21}$ в 10 пусках FOX и DASH не была определена *ab initio* ни разу.

4. С помощью мультипопуляционного генетического алгоритма определены новые кристаллические структуры комплексных соединений $[\text{Pt}(\text{NH}_3)_5\text{Cl}]\text{Br}_{2.4}\text{Cl}_{0.6}$ и $[\text{Pd}(\text{CH}_3\text{NH}_2)_4][\text{PdBr}_4]$.

СПИСОК ИСТОЧНИКОВ

1. Pecharsky, V. Fundamentals of Powder Diffraction and Structural Characterization of Materials, Second Edition / V. Pecharsky, P. Zavalij. – Springer, 2009. – 741 p.
2. David, W. I. F. Structure Determination from Powder Diffraction Data / W.I.F. David, K. Shankland, L.B. McCusker, C. Bärlocher. – Oxford University Press, 2006. – 337 p.
3. Le Bail, A. Third structure determination by powder diffractometry round robin (SDPDRR-3) / A. Le Bail, L. M. D. Cranswick // Powder Diffr. – 2009. – Vol. 24, №3. – P. 254-262.
4. David, W. I. F. Structure determination from powder diffraction data / W. I. F. David, K. Shankland // Acta Cryst. – 2008. – Vol. A64. – P. 52-64.
5. Shirley, R. Overview of powder-indexing program algorithms (history and strengths and weaknesses) / R. Shirley // IUCr Comput. Commission Newsletter. – 2003. – №2. – P. 48-54.
6. Runge, C. Die Bestimmung eines Kristallsystems durch Röntgenstrahlen / C. Runge // Phys. Zeit. – 1917. – №18. – P. 509-515.
7. De Wolff, P. M. A simplified criterion for the reliability of a powder pattern indexing / P. M. de Wolff // J. Appl. Cryst. – 1968. – №1. – P. 108-113.
8. Smith, G. S. F_N : A criterion for rating powder diffraction patterns and evaluating the reliability of powder-pattern indexing / G. S. Smith, R. L. Snyder // J. Appl. Cryst. – 1979. – №12. – P. 60-65.
9. De Wolff, P. M. Reliability of unit cells derived from powder diffraction patterns / P. M. de Wolff // Acta Cryst. – 1961. – №14. – P. 579-582.
10. Visser, J. W. A Fully Automatic Program for Finding the Unit Cell from Powder Data / J. W. Visser // J. Appl. Cryst. – 1969. – №2. – P. 89-95.
11. Mighell, A. D. Geometrical Ambiguities in the Indexing of Powder Patterns / A. D. Mighell, A. Santoro // J. Appl. Cryst. – 1975. – №8. – P. 372-374.
12. Werner, P.-E. TREOR: A Semi-exhaustive Trial-and-error Powder Indexing Program for All Symmetries / P.-E. Werner, L. Eriksson, M. Westdahl // J. Appl. Cryst. – 1985. – №18. – P. 367-370.
13. Boultif, A. Indexing of powder diffraction patterns for low symmetry lattices by the successive dichotomy method / A. Boultif, D. Louer // J. Appl. Cryst. – 1991. – №24. – P. 987-993.
14. R. Shirley, The Crysfire 2002 System for Automatic Powder Indexing: User's Manual / R. Shirley. – England : Lattice Press, 2002. – 37 p.
15. Ito, T. A general powder x-ray photography / T. Ito // Nature. – 1949. – №164. – P. 755-756.
16. Ito, T. X-ray studies on polymorphism // T. Ito. – Tokyo : Japan Maruzen Co. Ltd., 1950. – 231 p.

17. De Wolff, P. M. On the determination of unit-cell dimensions from powder diffraction patterns / P. M. de Wolff // *Acta Cryst.* – 1957. – №10. – P. 590-595.
18. De Wolff, P. M. Detection of simultaneous zone relations among powder diffraction lines / P. M. de Wolff // *Acta Cryst.* – 1958. – №11. – P. 664-665.
19. Block, S. Accuracy in Powder Diffraction: proceedings of a Symposium / ed. S. Block, C. R. Hubbard. – Washington, D.C. : U.S. Dept. of Commerce, National Bureau of Standards, 1980. – 572 p.
20. Louer, D. Methode d'essais et erreurs pour l'indexation automatique des diagrammes de poudre / D. Louer, M. Louer // *J Appl Cryst.* – 1972. – №5. – P. 271-275.
21. Louer, D. Indexation automatique des diagrammes de poudre par dichotomies successives / D. Louer, R. Vergas // *J Appl Cryst.* – 1982. – №15. – P. 542-545.
22. Bergmann, J. Renewed interest in powder diffraction data indexing / J. Bergmann, A. Le Bail, R. Shirley, V. Zlokazov // *Z. Kristallogr.* – 2004. – №219. – P. 783-790.
23. Altomare, A. Advances in powder diffraction pattern indexing: N-TREOR09 / A. Altomare, C. Giacovazzo, A. Guagliardi, A. G. G. Moliterni, R. Rizzi, P.-E. Werner // *J. Appl. Cryst.* – 2009. – №42. – P. 768-775.
24. Boultif, A. Powder pattern indexing and the dichotomy algorithm / A. Boultif, D. Louer // *Z. Kristallogr. Suppl.* – 2007. – №26. – P. 191-196.
25. Kariuki, B. M. A new approach for indexing powder diffraction data based on whole-profile fitting and global optimization using a genetic algorithm / B. M. Kariuki, S. A. Belmonte, M. I. McMahon, R. L. Johnston, K. D. M. Harris, R. J. Nelmes // *J. Synchrotron Radiat.* – 1999. – №6. – P. 87-92.
26. Coelho, A.A. Indexing of powder diffraction patterns by iterative use of singular value decomposition / A. A. Coelho // *J. Appl. Cryst.* – 2003. – №36. – P. 86-95.
27. Neumann, M. X-Cell: a novel indexing algorithm for routine tasks and difficult cases / M. Neumann // *J. Appl. Cryst.* – 2003. – №36. – P. 356-365.
28. Le Bail, A. Monte carlo indexing with mcmaille / A. Le Bail // *Powder Diffr.* – 2004. – №19. – P. 249-254.
29. Goldberg, D. E. Genetic Algorithms in Search, Optimization, and Machine Learning / D. E. Goldberg. – Boston, MA, USA : Addison-Wesley Longman Publishing Co., Inc., 1989. – 412 p.
30. Harris, K. D. M. The Genetic Algorithm: Foundations and Applications in Structure Solution from Powder Diffraction Data / K. D. M. Harris, R. L. Johnston, B. M. Kariuki // *Acta Crystallogr., Sect. A.* – 1998. – №54. – P. 632-645.
31. Tam, K.Y. GAMATCH – a genetic algorithm-based program for indexing crystal faces / K. Y. Tam, R. G. Compton // *J. Appl. Cryst.* – 1995. – №28. – P. 640-645.

32. Paszkowicz, W. Application of the Smooth Genetic Algorithm for Indexing Powder Patterns Tests for the Orthorhombic System / W. Paszkowicz // *Mat. Sci. Forum.* – 1996. – №228-231. – P. 19-24.
33. Le Bail, A. Ab-initio structure determination of LiSbWO_6 by X-ray powder diffraction / A. Le Bail, H. Duroy, J.L. Fourquet // *Mater. Res. Bull.* – 1998. – №23. – P. 447-452.
34. Coelho, A.A. TOPAS software [Электронный ресурс] URL: <http://www.topas-academic.net>
35. Nash, J. C. Compact numerical methods for computers: linear algebra and function minimisation – 2nd ed / J. C. Nash – Bristol, England : Adam Hilger, 1990. – 278 p.
36. Accelrys Material Studio software [Электронный ресурс] URL: <http://accelrys.com/products/collaborative-science/biovia-materials-studio/>
37. Markvardsen, A. J. A probabilistic approach to space-group determination from powder diffraction data / A. J. Markvardsen, W. I. F. David, J. C. Johnson, K. Shankland // *Acta Crystallogr., Sect. A.* – 2001. – №57. – P. 47-54.
38. David, W. I. F. DASH User Guide and Tutorials / W. I. F. David, K. Shankland, J. Cole, S. Maginn, W. D. S. Motherwell, R. Taylor. – Cambridge, UK : Cambridge Crystallographic Data Centre, 2017. – 231 p. [Электронный ресурс] URL: <https://www.ccdc.cam.ac.uk/solutions/csd-materials/components/dash/>
39. Altomare, A. Space-group determination from powder diffraction data: a probabilistic approach / A. Altomare, R. Caliendo, M. Camalli, C. Cuocci, I. da Silva, C. Giacovazzo, A. G. G. Moliterni, R. Spagna // *J. Appl. Cryst.* – 2004. – №37. – P. 957-966.
40. Altomare, A. Space group determination: improvements in EXPO2004 / A. Altomare, M. Camalli, C. Cuocci, I. da Silva, C. Giacovazzo, A. G. G. Moliterni and R. Rizzi // *J. Appl. Cryst.* – 2005. – №38. – P. 760-767.
41. Altomare, A. Automatic structure determination from powder data with EXPO2004 / A. Altomare, R. Caliendo, M. Camalli, C. Cuocci, C. Giacovazzo, A. G. G. Moliterni and R. Rizzi // *J. Appl. Cryst.* – 2004. – №37. – P. 1025-1028.
42. Winson, A. J. C. The probability distribution of X-ray intensities / A. J. C. Winson // *Acta Crystallogr.* – 1949. – №2. – P. 318-321.
43. Giacovazzo, C. Direct Phasing in Crystallography: Fundamentals and Applications / C. Giacovazzo – Oxford, UK : Oxford University Press, 1998. – 767 p.
44. Powder Diffraction: Theory and Practice / Ed. R. E. Dinnebier, S. J. L. Billinge. – Cambridge, UK : Royal Society of Chemistry, 2008. – 582 p.
45. Shoemaker, D. P. The crystal structure of L-threonine / D. P. Shoemaker, J. Donohue, V. Schomaker, R. B. Corey // *J. Am. Chem. Soc.* – 1950. – №72. – P. 2328-2349.

46. Jacobson, R. A. The crystal and molecular structure of cellobiose / R. A. Jacobson, J. A. Wunderlich, W. N. Lipscomb // *Acta Crystallogr.* – 1961. – №14. – P. 598-607.
47. Moras, D. Crystallographic computing 5: from chemistry to biology / ed. D. Moras, A. D. Podjarny, J. C Thierry. – Oxford, UK : Oxford University Press, 1991. – 480 p.
48. Rius, J. Determination of crystal structures with large known fragments directly from measured X-ray powder diffraction intensities / J. Rius, C Miravittles // *J. Appl. Cryst.* – 1988. – №21. – P. 224-227.
49. David, W. I. F. The probabilistic determination of intensities of completely overlapping reflections in powder diffraction patterns / W. I. F. David // *J. Appl. Cryst.* – 1987. – №20. – P. 316-319.
50. Estermann, M. A. Solving crystal structures with the symmetry minimum function / M. A. Estermann // *Nucl. Instrum. Methods Phys. Res. A.* – 1995. – №354. – P. 126-133.
51. Altomare, A. Solving Crystal Structures from Powder Data. IV. The Use of the Patterson Information for Estimating the $|F|$'s / A. Altomare, J. Foadi, C. Giacovazzo, A. G. G. Moliterni, M. C. Burla, G. Polidori // *J. Appl. Cryst.* – 1998. – №31. – P. 74-77.
52. Altomare, A. EXPO: a program for full powder pattern decomposition and crystal structure solution / A. Altomare, M. C. Burla, M. Camalli, B. Carrozzini, G. Cascarano, C. Giacovazzo, A. Guagliardi, A. G. G. Moliterni, G. Polidori, R. Rizzi // *J. Appl. Cryst.* – 1999. – №32. – P. 339-340.
53. Harker, D. Phases of Fourier coefficients directly from crystal diffraction data / D. Harker, J. S. Kasper // *Acta Crystallogr.* – 1948. – №1. – P. 70-75.
54. Altomare, A. Internal Report Istituto di Cristallografia (C.N.R), Via Amendola 122/0, 70126 Bari, Italy / A. Altomare, R. Caliandro, C. Cuocci, I. da Silva, C. Giacovazzo, A. G. G. Moliterni, R. Rizzi // Bari, Italy : Istituto di Cristallografia, 2004.
55. Pawley, G. S. Unit-cell refinement from powder diffraction scans / G. S. Pawley // *J. Appl. Cryst.* – 1981. – №14. – P. 357-361.
56. Altomare, A. A systematic procedure for the decomposition of a powder diffraction pattern / A. Altomare, R. Caliandro, C. Cuocci, C. Giacovazzo, A. G. G. Moliterni, R. Rizzi // *J. Appl. Cryst.* – 2003. – №36. – P. 906-913.
57. Altomare, A. The use of error-correcting codes for the decomposition of a powder diffraction pattern / A. Altomare, R. Caliandro, C. Cuocci, I. da Silva, C. Giacovazzo, A. G. G. Moliterni, R. Rizzi // *J. Appl. Cryst.* – 2004. – №37. – P. 204-209.
58. Accelrys Cerius 2 software [Электронный ресурс] URL:
<http://sites.udel.edu/martingroup/equipment/accelrys-molecular-modeling-software/>

59. PerkinElmer ChemOffice Professional [Электронный ресурс] URL:
http://www.cambridgesoft.com/Ensemble_for_Chemistry/ChemOffice/
60. Certara SYBYL-X Suite software [Электронный ресурс] URL:
<https://www.certara.com/software/molecular-modeling-and-simulation/sybyl-x-suite/>
61. Allen F. H. 3D search and research using the Cambridge Structural Database / F. H. Allen, O. Kennard // *Chem. Des. Autom. News.* – 1993. – №8. – P. 31-37.
62. Favre-Nicolin, V. FOX, free objects for crystallography: a modular approach to ab initio structure determination from powder diffraction / V. Favre-Nicolin, R. Cerny // *J. Appl. Cryst.* – 2002. – №35. – P. 734-743.
63. Andreev, Y. G. A General Monte Carlo Approach to Structure Solution from Powder Diffraction Data: Application to Poly(ethylene oxide)₃:LiN(SO₃CF₃)₂ / Y. G. Andreev, P. Lightfoot, P. G. Bruce // *J. Appl. Cryst.* – 1997. – №30. – P. 294-305.
64. Hirshfeld, F. L. Symmetry in the generation of trial structures / F. L. Hirshfeld // *Acta Crystallogr., Sect. A.* – 1968. – №24. – P. 301-311.
65. Lattman, B. E. Optimal sampling of the rotation function / B. E. Lattman // *Acta Crystallogr. Sect. B.* – 1972. – №28. – P. 1065-1068.
66. Chernyshev, V. V. A grid search procedure of positioning a known molecule in an unknown crystal structure with the use of powder diffraction data / V. V. Chernyshev, H. Schenk // *Z. Kristallogr.* – 1998. – №213. – P. 1-3.
67. Hammond, R. B. X-Form metal-free phthalocyanine: crystal structure determination using a combination of high-resolution X-ray powder diffraction and molecular modelling techniques / R. B. Hammond, K. J. Roberts, R. Docherty, M. Edmondson, R. Gairns // *J. Chem. Soc, Perkin Trans.* – 1996. – №2. – P. 1527-1528.
68. Masciocchi, N. The Multiphase Nature of the Cu(pz) and Ag(pz) (Hpz = Pyrazole) Systems: Selective Syntheses and Ab-Initio X-ray Powder Diffraction Structural Characterization of Copper(I) and Silver(I) Pyrazolates / N. Masciocchi, M. Moret, P. Cairati, A. Sironi, G. A. Ardizzoia, G. La Monica // *J. Am. Chem. Soc.* – 1994. – №116. – P. 7668-7676.
69. Chernishev, V. V. Potentialities of Grid Search Method / V. V. Chernishev // *Acta Crystallogr., Sect. A.* – 2000. – №56. – P. s132.
70. Metropolis, N. Equation of State Calculations by Fast Computing Machines / N. Metropolis, A. Rosenbluth, M. Roscnbluth, A. Teller, E. Teller // *J. Chem. Phys.* – 1953. – №21. – P. 1087-1092.
71. Solovyov, L. A. Application of a Simulated Annealing Approach in Powder Crystal Structure Analysis / L. A. Solovyov, S. D. Kirik // *Materials Science Forum.* – 1993. – №133-136. – P. 195-200.

72. Harris, K. D. M. Crystal Structure Determination from Powder Diffraction Data by Monte Carlo Methods / K. D. M. Harris, M. Tremayne, P. Lightfoot, P. G. Bruce // *J. Am. Chem. Soc.* – 1994. – №116. – P. 3543-3547.
73. Tremayne, M. Crystal Structure Solution from Neutron Powder Diffraction Data by a new Monte Carlo Approach Incorporating Restrained Relaxation of the Molecular Geometry / M. Tremayne, B. M. Kariuki, K. D. M. Harris, K. Shankland, K. S. Knight // *J. Appl. Cryst.* – 1997. – №30. – P. 968-974.
74. Tremayne, M. Structure determination of a complex organic solid from X-ray powder diffraction data by a generalized Monte Carlo method: The crystal structure of red fluorescein / M. Tremayne, B. M. Kariuki, K. D. M. Harris // *Angew. Chemie – Int. Ed. Engl.* – 1997. – №36. – P. 770-772.
75. Lanning, O. J. Definition of a 'guiding function' in global optimization: a hybrid approach combining energy and R-factor in structure solution from powder diffraction data / O. J. Lanning, S. Habershon, K. D. M. Harris, R. L. Johnston, B. M. Kariuki, E. Tedesco, G. W. Turner // *Chem. Phys. Lett.* – 2000. – №317. – P. 296-303.
76. Brodski, V. A Monte Carlo approach to crystal structure determination from powder diffraction data / V. Brodski, R. Peshar, H. Schenk // *J. Appl. Cryst.* – 2003. – №36. – P. 239-243.
77. Born, M. On the calculation of absolute crystal properties using Bohr atomic models / M. Born and A. Lande // *Sitzungsber. Preuss. Akad. Wiss. Berlin.* – 1918. – №45. – P. 1048-1068.
78. Rappe, A. K. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations / A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard III, W. M. Skiff // *J. Am. Chem. Soc.* – 1992. – №114. – P. 10024-10035.
79. Coelho, A. A. Whole-profile structure solution from powder diffraction data using simulated annealing / A. A. Coelho // *J. Appl. Cryst.* – 2000. – №33. – P. 899-908.
80. Born, M. Zur Gittertheorie der Ionenkristalle / M. Born and J. E. Mayer // *Z. Phys.* – 1932. – №75. – P. 1-18.
81. Putz, H. Combined method for ab initio structure solution from powder diffraction data / H. Putz, J. C. Schon, M. Jansen // *J. Appl. Cryst.* – 1999. – №32. – P. 864-870.
82. Bergerhoff, G. The inorganic crystal structure data base / G. Bergerhoff, R. Hundt, R. Sievers, I. D. Brown // *J. Chem. Inf. Comput. Sci.* – 1983. – №23. – P. 66-69.
83. Hsu, H.-P. Energy landscape paving for X-ray structure determination of organic molecules / H.-P. Hsu, S. C. Lin, U. H. E. Hansmann // *Acta Crystallogr., Sect. A.* – 2002. – №58. – P. 259-264.
84. Kirkpatrick, S. Optimization by simulated annealing / S. Kirkpatrick, C. D. Gelatt Jr., M. P. Vecchi // *Science.* – 1983. – №220. – P. 671-680.

85. Kirkpatrick, S. Optimization by simulated annealing: Quantitative studies / S. Kirkpatrick // *J. Stat. Phys.* – 1984. – №34. – P. 975-986.
86. Geman, S. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images / S. Geman, D. Geman // *IEEE Trans. Pattern Anal. Mach. Intell.* – 1984. – №6. – P. 721-741.
87. Andreev, Y. G. Solving crystal structures of molecular solids without single crystals: a simulated annealing approach / Y. G. Andreev, P. G. Brace // *J. Chem. Soc, Dalton Trans.* – 1998. – №24. – P. 4071-4080.
88. David, W. I. F. Routine Determination of Molecular Crystal Structures from Powder Diffraction Data / W. I. F. David, K. Shankland, N. Shankland // *Chem. Commun.* – 1998. – №8. – P. 931-932.
89. Markvardsen, A. J. A maximum-likelihood method for global-optimization-based structure determination from powder diffraction data / A. J. Markvardsen, W. I. F. David, K. Shankland // *Acta Crystallogr., Sect. A.* – 2002. – №58. – P. 316-326.
90. Shankland, K. Molecular, crystallographic and algorithmic factors in structure determination from powder diffraction data by simulated annealing / K. Shankland, L. McBride, W. I. F. David, N. Shankland, G. Steele // *J. Appl. Cryst.* – 2002. – №35. – P. 443-454.
91. Leach, A. R. *Molecular Modelling: Principles and Applications* / A. R. Leach – London, UK : Addison Wesley Longman Ltd., 1996. – 744 p.
92. Engel, G. E. PowderSolve – a complete package for crystal structure solution from powder diffraction patterns / G. E. Engel, S. Wilke, O. König, K. D. M. Harris, F. J. J. Leusen // *J. Appl. Cryst.* – 1999. – №32. – P. 1169-1179.
93. Le Bail, A. ESPOIR: A Program for Solving Structures by Monte Carlo Analysis of Powder Diffraction Data / A. Le Bail // *Mater. Sci. Forum.* – 2001. – №378-381. – P. 65-70.
94. Kariuki, B. M. The application of a genetic algorithm for solving crystal structures from powder diffraction data / B. M. Kariuki, H. Serrano-González, R. L. Johnston, K. D. M. Harris // *Chem. Phys. Lett.* – 1997. – №280. – P. 189-195.
95. Shankland, K. Crystal Structure Determination from Powder Diffraction Data by the Application of a Genetic Algorithm / K. Shankland, W. I. F. David, T. Csoka // *Z. Kristallogr.* – 1997. – №212. – P. 550-552.
96. Turner, G. W. Implementation of Lamarckian concepts in a Genetic Algorithm for structure solution from powder diffraction data / G. W. Turner, E. Tedesco, K. D. M. Harris, R. L. Johnston, B. M. Kariuki // *Chem. Phys. Lett.* – 2000. – №321. – P. 183-190.
97. Storn, R. Differential Evolution - a Simple and Efficient Heuristic for Global Optimization over Continuous Spaces / R. Storn, K. V. Price // *J. Global Optimization.* – 1997. – №11. – P. 341-359.

98. Tremayne, M. Structures of three substituted arenesulfonamides from X-ray powder diffraction data using the differential evolution technique / M. Tremayne, C. C. Seaton, C. Glidewell // *Acta Crystallogr., Sect. B.* – 2002. – №58. – P. 823-834.
99. Brenner, S. The application of structure envelopes in structure determination from powder diffraction data / S. Brenner, L. B. McCusker, C. Baerlocher // *J. Appl. Cryst.* – 2002. – №35. – P. 243-252.
100. Altomare, A. Completion of crystal structures from powder data: the use of the coordination polyhedral / A. Altomare, C. Giacovazzo, A. Guagliardi, A. G. G. Moliterni, R. Rizzi, *J. Appl. Cryst.* – 2000. – №33. – P1305-1310.
101. Altomare, A. Completion of crystal structure by powder diffraction data: a new method for locating atoms with polyhedral coordination / A. Altomare, C. Cuocci, C. Giacovazzo, A. G. G. Moliterni and R. Rizzi, *J. Appl. Cryst.* – 2002. – №35. – P. 422-429.
102. Tanahashi, Y. Ab initio structure determination of monoclinic 2,2-dihydroxymethylbutanoic acid from synchrotron radiation powder diffraction data: combined use of direct methods and the Monte Carlo method / Y. Tanahashi, H. Nakamura, S. Yamazaki, Y. Kojima, H. Saito, T. Ida and H. Toraya, *Acta Crystallogr., Sect. B.* – 2001. – №57. – P. 184-189.
103. Altomare, A. Solution of organic crystal structures from powder diffraction by combining simulated annealing and direct methods / A. Altomare, R. Caliandro, C Giacovazzo, A. G. G. Moliterni, R. Rizzi, *J. Appl. Cryst.* – 2003. – №36. – P. 230-238.
104. Harris, K. D. M. Contemporary Advances in the Use of Powder X-Ray Diffraction for Structure Determination / K. D. M. Harris, M. Tremayne, B. M. Kariuki // *Angew. Chem., Int. Ed.* – 2001. – №40. – P. 1626-1651.
105. Feng, Z. J. GEST: a program for structure determination from powder diffraction data using a genetic algorithm / Z. J. Feng, C. Dong // *J. Appl. Cryst.* – 2007. – №40. – P. 583-588.
106. Michalewicz, Z. Genetic Algorithms + Data Structures = Evolution Programs / Z. Michalewicz. – New York: Springer-Verlag, 1996. – 387 p.
107. MAUD software [Электронный ресурс] URL: <http://maud.radiographema.eu>
108. Johnston, J. C. A hybrid Monte Carlo method for crystal structure determination from powder diffraction data / J. C. Johnston, W. I. F. David, A. J. Markvardsen, K. Shankland // *Acta Crystallogr., Sect. A.* – 2002. – №58. – P. 441-447.
109. Brenner, S. Using a structure envelope to facilitate structure solution from powder diffraction data / S. Brenner, L. B. McCusker, C. Baerlocher // *J. Appl. Cryst.* – 1997. – №30. – P. 1167-1172.
110. Brenner, S. Structure Envelopes and their Application in Structure Determination from Powder Diffraction Data : PhD Thesis / Simon Brenner. – Zürich, 1999. – 93 p.

111. Turing, A.M. Computing machinery and intelligence / A.M. Turing // *Mind*. – 1950. – vol. 236, #59. – P. 433-460.
112. Емельянов, В. В. Теория и практика эволюционного моделирования / В.В. Емельянов, В.М. Курейчик, В.В. Курейчик. – М.: ФИЗМАТЛИТ, 2003. – 432 с.
113. Семенкин, Е. С. Оптимизация технических систем: учеб. пособие / Е.С. Семенкин, О.Э. Семенкина, С.П. Коробейников. – Красноярск: СИБУП, 1996. – 167 с.
114. Koza, J. R. Genetic Programming: On Programming Computer by Means of Natural Selection and Genetics / J. R. Koza. – Cambridge: The MIT Press, 1992. – 835 p.
115. Koza, J. R. Genetic Programming: A Paradigm for Genetically Breeding Populations of Computer Programs to Solve Problems – Technical report / J. R. Koza. – Stanford : Stanford University Press, 1990. – 127 p.
116. Rechenberg, I. Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der Biologischen Information / I. Rechenberg. – Freiburg : Fromman, 1973. – 170 p.
117. Schwefel, H-P. Evolution and optimum seeking / H.-P. Schwefel. – New York, John Wiley & Sons, 1995. – 444 p.
118. Holland, J. H. Adaptation in natural and artificial systems / J. H. Holland. – MI, USA : University of Michigan Press, 1975. – 228 p.
119. Beasley, D. An Overview of Genetic Algorithms: Part 2, Research topics / D. Beasley, D. R. Bull, R. R. Martin // *University Computing*. – 1993. – №15. – P. 170-181.
120. Решетников, Е. Дискретная математика: алгоритмы – кодирование целых чисел [Электронный ресурс] URL: <http://rain.ifmo.ru/cat/view.php/theory/coding/integer-2005>
121. Семенкин, Е. С. Применение генетического алгоритма с модифицированным оператором множественной рекомбинации при автоматизированном формировании интеллектуальных информационных технологий / Е.С. Семенкин, М.Е. Семенкина // *Вестник Сибирского государственного аэрокосмического университета имени академика М.Ф. Решетнева*. – 2007. – Вып. 3 (16). – С. 27-33.
122. Eiben, A. E. Genetic Algorithms with multi-parent recombination / A. E. Eiben, P-E. Raue, Zs. Ruttkay // *Parallel Problem Solving from Nature III*. – 1994. – P. 78-87.
123. De Jong, K. A. An Analysis of the Interacting Role of Population Size and Crossover in Genetic Algorithms / K. A. De Jong, W. M. Spears // *Parallel Problem Solving from Nature*. – 1990. – P. 38-47.
124. Lobo, F. G. Lost gems of EC: The Equilibrium Genetic Algorithm and the Role of Crossover / F. G. Lobo // *ACM SIGEVOlution*. – 2007. – №2. – P. 14-15.
125. Kursawe F. Breeding ES - first results / F. Kursawe // *Seminar "Evolutionary algorithms and their applications"*. – 1996. – P. 121-132.

126. Rietveld, H. M. A profile refinement method for nuclear and magnetic structures / H. M. Rietveld // *J. Appl. Cryst.* – 1969. – №2. – P. 65-71.
127. Cordero, B. Covalent radii revisited / B. Cordero, V. Gómez, A. E. Platero-Prats, M. Revés, J. Echeverría, E. Cremades, F. Barragán, Santiago Alvarez // *Dalton Trans.* – 2008. – №21. – P. 2832-2838.
128. Якимов, Я. И. Эволюционный метод моделирования кристаллической структуры вещества по данным порошковой дифракции / Я. И. Якимов, С. Д. Кирик, Е. С. Семенкин, Л. А. Соловьев, И. С. Якимов // *Журнал Сибирского федерального университета. Серия «Химия».* – 2013. – Том 6, №2. – С. 180-191.
129. Xia, Z. Crystal and local structure refinement in $\text{Ca}_2\text{Al}_3\text{O}_6\text{F}$ explored by X-ray diffraction and Raman spectroscopy / Z. Xia, M. S. Molokeev, A. S. Oreshonkov, V. V. Atuchin, R.-S. Liue, C. Dong // *Phys. Chem. Chem. Phys.* – 2014. – №16. – P. 5952-5957.
130. Egorysheva, A.V. New complex bismuth oxides in the $\text{Bi}_2\text{O}_3\text{--NiO--Sb}_2\text{O}_5$ system and their properties / A.V. Egorysheva, O.G. Ellert, Y.V. Zubavichus, O.M. Gajtko, N.N. Efimov, R.D. Svetogorov, V.Yu. Murzin // *Journal of Solid State Chemistry.* – 2015. – №225. – P. 97-104.
131. Cantú-Paz, E. Efficient parallel genetic algorithms: theory and practice / E. Cantú-Paz, D. E. Goldberg // *Computer Methods in Applied Mechanics and Engineering.* – 2000. – №186. – P. 221-238.
132. Habershon, S. Development of a multipopulation parallel genetic algorithm for structure solution from powder diffraction data / S. Habershon, K. D. M. Harris, R. L. Johnston // *J. Comput. Chem.* – 2003. – №24. – P. 1766-1774.
133. Griffin, T. A. N. GDASH: a grid-enabled program for structure solution from powder diffraction data / T. A. N. Griffin, K. Shankland, J. van de Streek, J. Cole // *J. Appl. Cryst.* – 2009. – №42. – P. 356-359.
134. Cerny, R. Direct space methods of structure determination from powder diffraction: principles, guidelines, perspectives / Radovan Cerny, Vincent Favre-Nicolin // *Z. Kristallogr.* – 2007. – №222. – P. 105-113.
135. Harris, K. D. M. Powder Diffraction Crystallography of Molecular Solids / K. D. M. Harris // *Top. Curr. Chem.* – 2012. – №315. – P. 133-177.
136. Meredig, B. A hybrid computational–experimental approach for automated crystal structure solution / B. Meredig, C. Wolverton // *Nature Materials.* – 2013. – №12. – P. 123-127.
137. FOX software [Электронный ресурс] URL: <http://fox.vincefn.net>
138. CCDC DASH software [Электронный ресурс] URL: <https://www.ccdc.cam.ac.uk/solutions/csd-materials/components/dash/>

139. Блохин, А. И. Кристаллическая структура транс-[Pd(NH₃)₂(NO₂)₂] по данным порошковой рентгенографии / А.И. Блохин, Л.А. Соловьев, М.Л. Блохина, С.Д. Кирик, И.С. Якимов // Коорд. химия. – 1996. – Том 22, № 3. – С. 198-202.
140. Черняев, И. И. Синтез комплексных соединений металлов платиновой группы. Справочник / Отв. ред. И. И. Черняев. – М.: Наука, 1964. – 340 с.
141. Залого, А. Н. Мультипопуляционный генетический алгоритм моделирования кристаллической структуры вещества из рентгенодифракционных данных / А. Н. Залого, С. В. Бураков, Е. С. Семенкин, И.С. Якимов // Журнал Сибирского федерального университета. Серия «Химия». – 2014. – Том 7, № 4. – С. 573-581.
142. Zaloga, A. N. Research on convergence of multipopulation binary- and real-coded genetic algorithms for solution of crystal structure from X-Ray powder diffraction data / A. N. Zaloga, S. V. Burakov, E. S. Semenkin, I. S. Yakimov. // Crystal Research and Technology. – 2015. – Vol. 50, Issue 9-10. – P. 724-728.
143. IUCr checkCIF tool [Электронный ресурс] URL: <http://checkcif.iucr.org>
144. Pennington, W.T. DIAMOND – Visual Crystal Structure Information System / W. T. Pennington // J. Appl. Cryst. – 1999. – №32 – P. 1028-1029.

ПРИЛОЖЕНИЕ А.

МПГА с использованием синхротронных дифракционных данных

Для демонстрации универсальности работы МПГА с различными типами дифракционных данных опишем процесс поиска структуры $\text{Bi}_3\text{FeO}_{11}\text{Sb}_2$ [1] при помощи МПГА. Его дифрактограмма была получена для однофазного порошкового материала с помощью синхротронного излучения на основе рентгеновской дифракции. Измерения производились на станции «Структурное Материаловедение» [2] Курчатовского источника синхротронного излучения (РНЦ «Курчатовский институт», Москва) с использованием 2D детектора FujiFilm ImagingPlate на длине волны $\lambda = 0.68886 \text{ \AA}$. Для калибровки угловой шкалы был использован поликристаллический кремниевый порошок (NIST SRM 640c). Дополнительные параметры дифрактограммы приведены в таблице 1. Уточнение структуры этого вещества авторы проводили методом Ритвельда с использованием программы Jana2006 [3].

Таблица 1 – Параметры представленного образца, описывающего кристаллическую структуру $\text{Bi}_3\text{FeO}_{11}\text{Sb}_2$

Структурный тип	KSbO3
Пр. группа симметрии	Pn-3 (#201)
Сторона ячейки a	9.51521(2) \AA
Длина волны излучения при съёмке	0.68886 \AA
Диапазон измерения 2θ	5–50 °

Для поиска использовались настройки МПГА, приведённые на рисунке 1. Предполагаемый атомный состав независимой части ячейки был известен: 2 атома Bi, атом Fe, атом Sb и три атома O. Было также известно, что тяжёлые атомы заполняют свои позиции не полностью, поэтому вместе с атомными координатами производился поиск их коэффициентов заполнения атомных позиций. Коэффициенты тепловых колебаний Biso для атомов Bi были заданы равными 1,5, для остальных атомов – равными 0,5. Список искомых по МПГА параметров приведён в таблице 2. В общей сложности производился поиск значений 25 структурных параметров.

Параметры работы ГА: Rwp Chi² Количество рабочих ядер

Бит на каждую координату	0
Размер популяции	300
Кол-во поколений	200
Кол-во индивидов для лок. поиска по МНК	5
Порог фитнеса для запуска МНК	50
Регулярность запуска МНК, поколений	10
Кол-во циклов МНК при лок. поиске	4
Диапазон поколений для обмена	4
Сколько индивидов отправлять при обмене	2
Кол-во поколений локального элитизма	3
Записывать в лог всех индивидов	0

Номер ядра	Размер турнира	К. скрещивания	К. мутации
1	3	0,75	1
2	3	0,75	1,3
3	3	0,75	1,6

Параметры поиска структуры:

Минимальная дистанция по умолчанию	1.0
Коеффициент штрафа при сближении < min_dist	0.5
Исп. динамическую заполняемость ат. позиций	0
Исп. схлопывание на элементах симметрии	1

Рисунок 1 – Настройки МПГА для поиска структуры $\text{Bi}_3\text{FeO}_{11}\text{Sb}_2$ по синхротронным данным

Таблица 2 – Список параметров для поиска структуры $\text{Bi}_3\text{FeO}_{11}\text{Sb}_2$

Атом	Диапазон поиска параметра			
	X	Y	Z	Occur.
Bi1	[0,0..0,25]	[0,0..0,25]	[0,0..1,0]	[0,05..1,0]
Bi2	[0,0..0,25]	[0,0..0,25]	[0,0..1,0]	[0,05..1,0]
Fe	[0,0..0,25]	[0,0..0,25]	[0,0..1,0]	[0,05..1,0]
Sb	[0,0..0,25]	[0,0..0,25]	[0,0..1,0]	[0,05..1,0]
O1	[0,0..0,25]	[0,0..0,25]	[0,0..1,0]	-
O2	[0,0..0,25]	[0,0..0,25]	[0,0..1,0]	-
O3	[0,0..0,25]	[0,0..0,25]	[0,0..1,0]	-

Перед началом процесса поиска был проведён поиск профильных параметров по процедуре Ле Бейла. Затем был произведён тестовый пуск МПГА, чтобы убедиться, что все параметры заданы корректно, затем была запущена серия из 10 запусков, и в качестве результата был принят запуск с наименьшим значением функции пригодности. В таблице 3 приведены результаты 10 запусков МПГА для поиска этой структуры. Значение целевой функции на заключительном этапе поиска численно совпадает с профильным R-фактором, потому что вклад штрафа сокращается до нуля. Профильный R_{wp} -фактор в МПГА вычисляется по методике библиотеки ObjCryst++ [4], встроенной в МПГА.

Таблица 3 – Результаты серии из десяти запусков МПГА для поиска кристаллической структуры $\text{Bi}_3\text{FeO}_{11}\text{Sb}_2$

Номер запуска	Значение целевой функции (R_{wp} -фактор)	Затраченное время, мин
1	4.86	8:31
2	5.59	8:19
3	3.97	8:27
4	4.06	8:34
5	4.05	8:38
6	13.08	8:17
7	6.08	8:08
8	5.13	8:39
9	3.83	7:52
10	3.51	8:41

График сходимости 10-ого запуска МПГА приведён на рисунке 2. Его можно интерпретировать так. К 30-му поколению МПГА удалось локализовать тяжелые атомы. К 70-му поколению МПГА успешно разместил кислороды вокруг тяжелых атомов, функция штрафа сократилась до нуля. Далее МПГА в основном уточнял коэффициенты заполнения тяжелыми атомами их позиций.

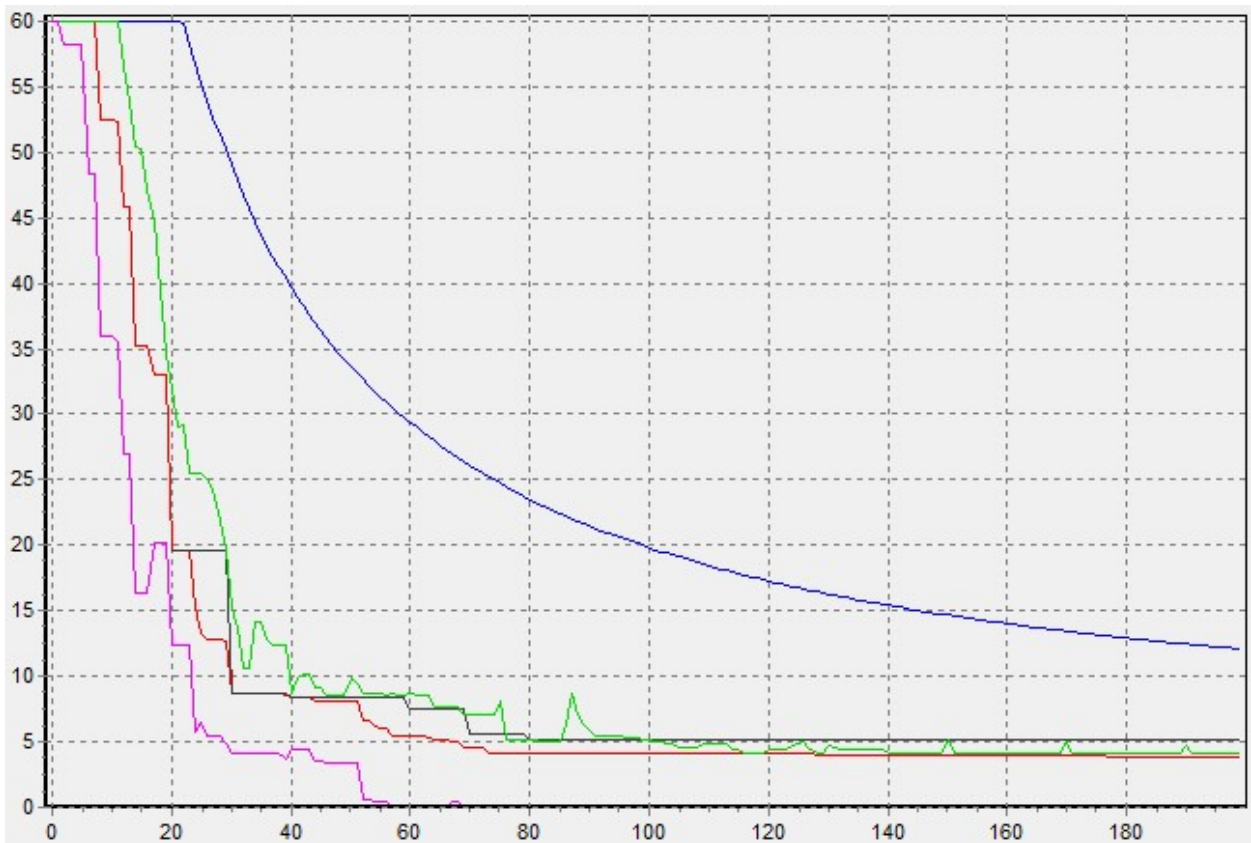


Рисунок 2 – График сходимости при поиске структуры $\text{Bi}_3\text{FeO}_{11}\text{Sb}_2$.

Ось абсцисс – номер поколения, ось ординат – значение функции пригодности.

На рисунке 3 приведено сравнение экспериментальной дифрактограммы с расчётной, полученной для структуры, найденной по МПГА. В таблице 4 приведено сравнение параметров структуры, найденных по МПГА и полученных её авторами [1]. На рисунке 4 приведено аналогичное сравнение в графическом виде.

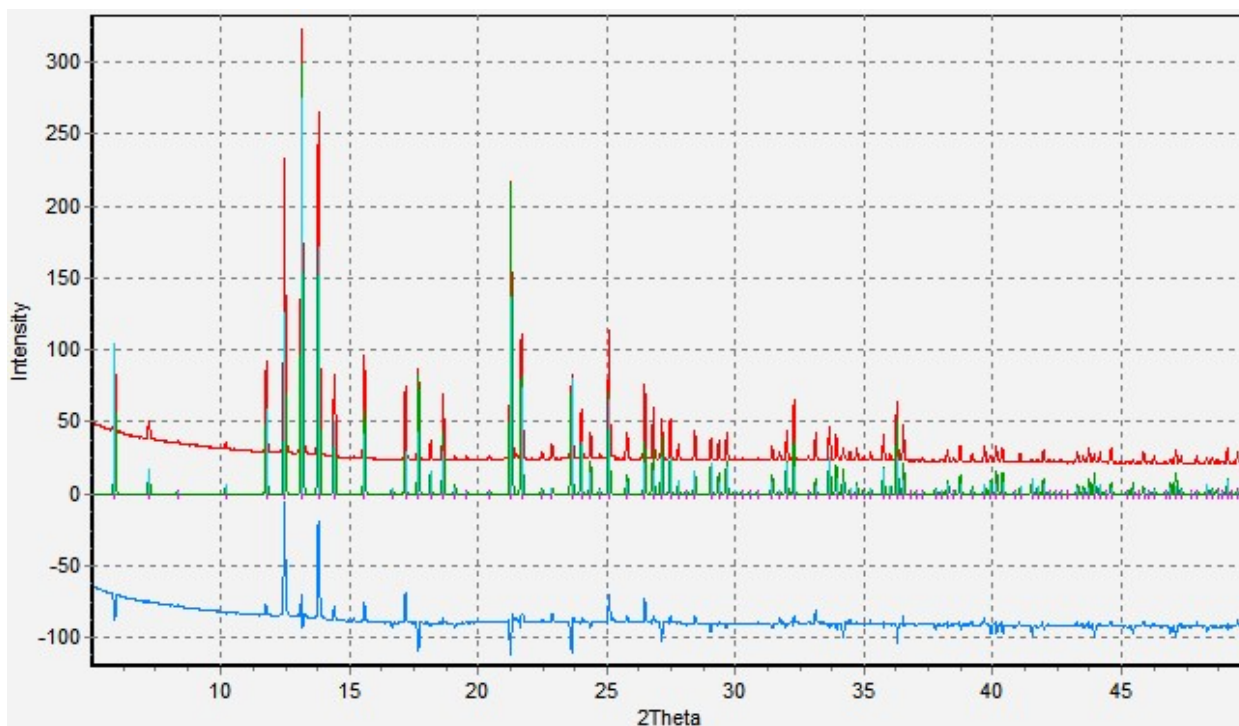


Рисунок 3 – Сравнение экспериментальной дифрактограммы $\text{Bi}_3\text{FeO}_{11}\text{Sb}_2$ (красная) с расчётной (зелёная), для найденной по МПГА структуры ($R_{\text{wp}} = 3.51$; $R\text{-DDM} = 13.71$). В нижней части графика приведена линия разности (синяя).

Таблица 4 – Соответствие параметров найденной по МПГА структуры ($R_{\text{wp}} = 3.51$; $R\text{-DDM} = 13.71$) и параметров, которые описаны её авторами (*).

Атом	$X^{\text{МПГА}}$	X^*	$Y^{\text{МПГА}}$	Y^*	$Z^{\text{МПГА}}$	Z^*	Осс. ^{МПГА}	Осс.*
Bi1	0.000997	0	0.002175	0	0.001739	0	1.0	1.0
Bi2	0.122773	0.0949(5)	0.116047	0.1316(39)	0.380561	0.3684(37)	0.3333	0.3333
Fe	0.093862	0.093(31)	0.25	0.25	0.75	0.75	0.4975	0.3333
Sb	0.092177	0.093(31)	0.25	0.25	0.75	0.75	0.6315	0.6667
O1	0.138908	0.1446(38)	0.138852	0.1446(38)	0.138910	0.1446(38)	1	1
O2	0.25	0.25	0.25	0.25	0.602171	0.6113(43)	1	1
O3	0.087173	0.042(5)	0.033(8)	0.033487	0.751511	0.7529(22)	1	1

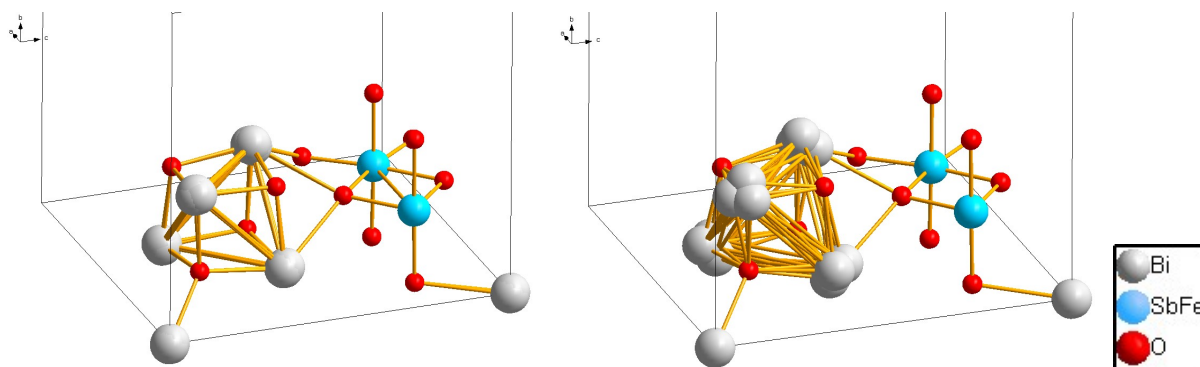


Рисунок 4 – Графическое изображения ссоответствия положений атомов в найденной по МПГА структуре (слева) и в описанной её авторами (справа).

Затем результат поиска по МПГА был подвергнут доуточнению всех параметров при помощи программы полнопрофильного анализа DDM. Решение, полученное по МПГА при его загрузке в программу DDM показало значение профильного R-DDM фактора 9.54 %. После ручного доуточнения профильных и структурных параметров профильный R-DDM фактор составил 6.615 %. В таблице 5 приведено сравнение параметров уточнённой по DDM структуры, и полученных её авторами [1]. На рисунке 5 приведено сравнение экспериментальной дифрактограммы с расчётной, полученной по DDM.

Структура, описанная авторами в статье [1] имеет значения $R_{wp} = 2,76$ и $R-DDM = 6,70$.

Таблица 5 – Соответствие параметров найденной по МПГА и уточнённой по DDM структуры ($R_{wp} = 2,88$; $R-DDM = 6,62$) и параметров, которые описаны её авторами (*).

Атом	$X^{МПГА}$	X^*	$Y^{МПГА}$	Y^*	$Z^{МПГА}$	Z^*	Осс. ^{МПГА}	Осс.*
Bi1	0	0	0	0	0	0	1	1
Bi2	0.133(5)	0.0949(5)	0.132(5)	0.1316(39)	0.4037(6)	0.3684(37)	0.3333	0.3333
Fe	0.0942(9)	0.093(31)	0.25	0.25	0.75	0.75	0.278(27)	0.3333
Sb	0.0942(9)	0.093(31)	0.25	0.25	0.75	0.75	0.722(27)	0.6667
O1	0.1484(31)	0.1446(38)	0.1484(31)	0.1446(38)	0.1484(31)	0.1446(38)	1	1
O2	0.25	0.25	0.25	0.25	0.595(5)	0.6113(43)	1	1
O3	0.0827(36)	0.042(5)	0.0394(36)	0.0891(41)	0.7536(24)	0.7529(22)	1	1



Рисунок 5 – Сравнение экспериментальной дифрактограммы $\text{Bi}_3\text{FeO}_{11}\text{Sb}_2$ (красная) с расчётной (зелёная), для найденной по МПГА и уточнённой по DDM структуры ($R_{\text{wp}} = 2,88$; $R\text{-DDM} = 6,62$). В нижней части графика приведена линия разности (синяя).

Список источников

1. Egorysheva, A.V. The $\text{Bi}_2\text{O}_3\text{--Fe}_2\text{O}_3\text{--Sb}_2\text{O}_5$ system phase diagram refinement, $\text{Bi}_3\text{FeSb}_2\text{O}_{11}$ structure peculiarities and magnetic properties / A.V. Egorysheva, O.G. Ellert, O.M. Gajtko, N.N. Efimova, R.D. Svetogorov, Y.V. Zubavichus, A.V. Grigorieva // *Journal of Solid State Chemistry*. – 2015. – №225. – P. 278-284.
2. Chernyshov, A. A. Structural Materials Science end-station at the Kurchatov Synchrotron Radiation Source: Recent instrumentation upgrades and experimental results / A.A. Chernyshov, A.A. Veligzhanin, Y.V. Zubavichus // *Nucl. Instrum. Methods Phys. Res.* – 2009. – №A603. – P. 95-98.
3. Petricek, V. Crystallographic Computing System JANA2006: General features / V. Petricek, M. Dusek, L. Palatinus // *Z. Kristallogr.* – 2014. – №229. – P. 345-352.
4. Favre-Nicolin, V. FOX, free objects for crystallography: a modular approach to ab initio structure determination from powder diffraction / V. Favre-Nicolin, R. Cerny // *J. Appl. Cryst.* – 2002. – №35. – P. 734-743.

ПРИЛОЖЕНИЕ Б.

Карты распределения атомов в популяциях в процессе сходимости ГА

Для лучшего понимания процессов сходимости ГА в популяциях на ядрах МПГА разработан специальный новый визуализатор. Он позволяет визуализировать проекции атомов на указанные базовые плоскости элементарной ячейки (напр., *ac*) сразу для всех структурных моделей популяции на указанном поколении эволюции. Возможно задание какой-либо известной структуры для сравнения, тогда фиксированные позиции ее атомов отмечаются цветными крестами. Одновременно, изображаются графики сравнения расчетной дифрактограммы для лучшей из этих моделей с экспериментальной. Таким образом, при on-line изучении сходимости визуализатор дает возможность листать и сравнивать страницы с распределениями атомов в разных популяциях на разных эволюционных поколениях, сопоставлять их с графиками сходимости на данных ядрах и т.д.

В частности, это средство очень удобно для динамического изучения сходимости на тестовых структурах. В качестве иллюстрации ниже рассмотрены крайние примеры быстрой и медленной сходимости МПГА при поиске тестовой структуры K_4SnO_4 .

Параметры тестовой кристаллической структуры K_4SnO_4 : $a=6,48\text{\AA}$, $b=6,51\text{\AA}$, $c=9,70\text{\AA}$, $\alpha=71,82^\circ$, $\beta=99,89^\circ$, $\gamma=113,13^\circ$; P-1; Z=2; 27 степеней свободы атомных координат. Используемые для её определения установка МПГА приведены на рисунке 1. Все популяции МПГА на ядрах состоят из 50 структурных моделей. На рисунках приведены распределения атомов в популяциях на наиболее показательных, в смысле сходимости, поколениях эволюции, а цветными крестами отмечены истинные позиции атомов. Можно оценить динамику сближения и, далее, слияния атомов моделей в их истинных позициях.

Параметры работы ГА:		<input checked="" type="radio"/> Rwp <input type="radio"/> Chi ²		Количество рабочих ядер	3		
Бит на каждую координату	0	Номер ядра	Размер турнира	К. скрещивания	К. мутации		
Размер популяции	50	1	3	0,75	1		
Кол-во поколений	50	2	3	0,75	1,1		
Кол-во индивидов для лок. поиска по МНК	5	3	3	0,75	1,2		
Порог фитнеса для запуска МНК	150						
Регулярность запуска МНК, поколений	3						
Кол-во циклов МНК при лок. поиске	5						
Диапазон поколений для обмена	4						
Сколько индивидов отправлять при обмене	2						
Кол-во поколений локального элитизма	3						
Записывать в лог всех индивидов	1						
Параметры поиска структуры:							
Минимальная дистанция по-умолчанию	1.6						
Кoeffициент штрафа при сближении < min_dist	0.2						
Исп. динамическую заполняемость ат. позиций	0						
Исп. схлопывание на элементах симметрии	1						

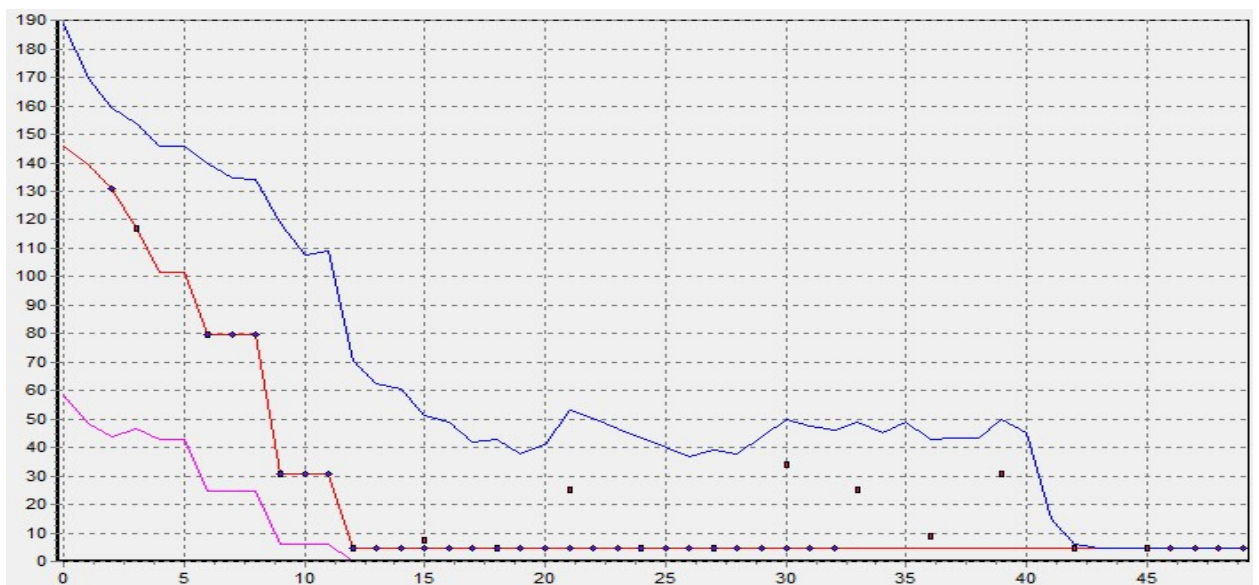
Рисунок 1 – Установки МПГА при определении кристаллической структуры K_4SnO_4 .

Б.1 Пример быстрой сходимости

На рисунке 2 приведены графики сходимости МПГА при определении структуры K_4SnO_4 . Представлен пример быстрой сходимости – истинная структура была найдена на 12-ом поколении работы МПГА. На рисунках 3-11 приведены карты распределения атомов на основных этапах определения структуры во время этого пуска МПГА.



а



б

Рисунок 2 – Графики сходимости лучшей структурной модели K_4SnO по МПГА: а) на управляющем ядре; б) на лучшем из вычислительных ядер (ядро №1).

Ось абсцисс – номер поколения, ось ординат – пригодность: $Y = R_{wp} + \text{штраф}$; Y – красная линия (черная – с уточнением по МНК); $Y_{\text{сред.}}$ – синяя линия; штраф за некорректные межатомные расстояния – фиолетовая линия. Структура определена на 12-м поколении эволюции с

$$Y = R_{wp} < 5\% \quad (\text{штраф} = 0)$$

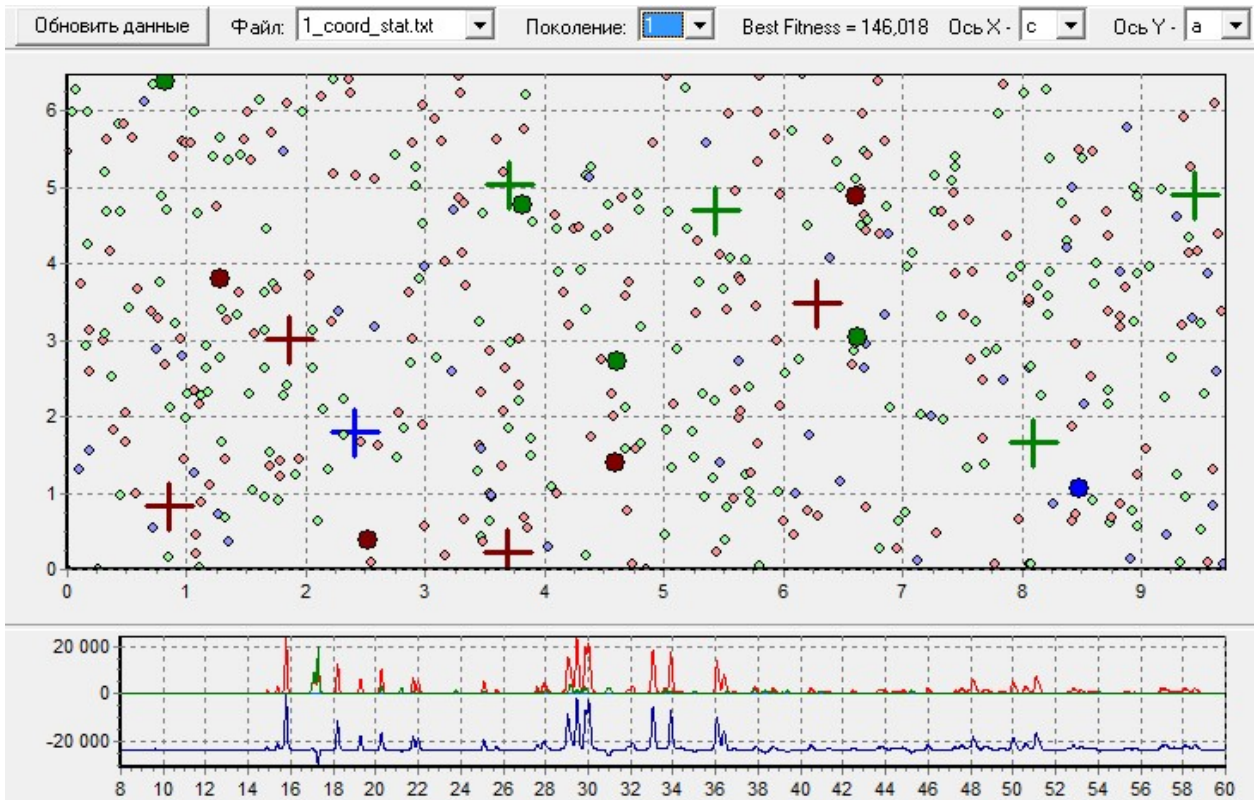


Рисунок 3 – Случайно сгенерированное 1-е поколение структурных моделей (50 штук): цвета атомов: Sn – синий, К – зеленый, О – красный; кресты – их истинные позиции в структуре; большие кружки – атомы лучшей модели; кружки с красной обводкой – атомы в истинных позициях, совпавшие по (x,y,z);

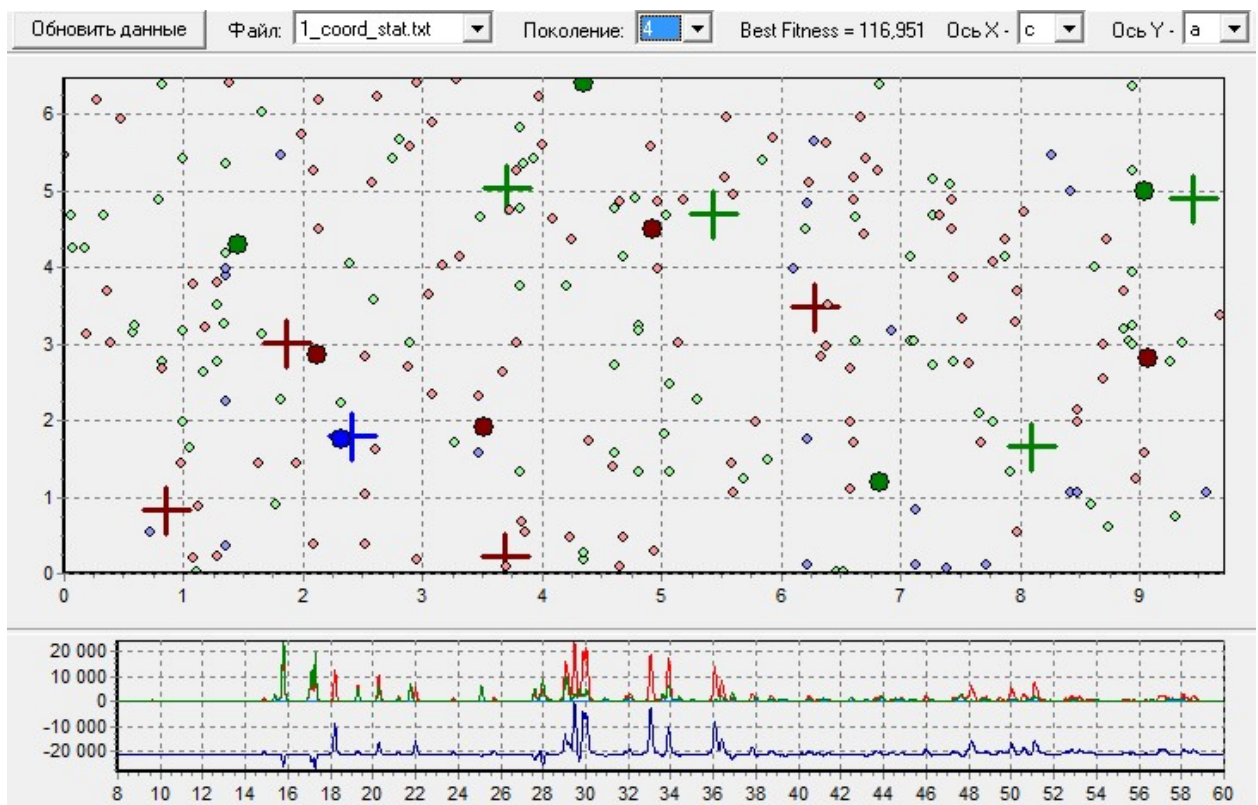


Рисунок 4 – 4-е поколение структурных моделей; R_{wp} лучшей структурной модели = 116,95%, $S_n(x,y)$ – в истинной позиции в структуре

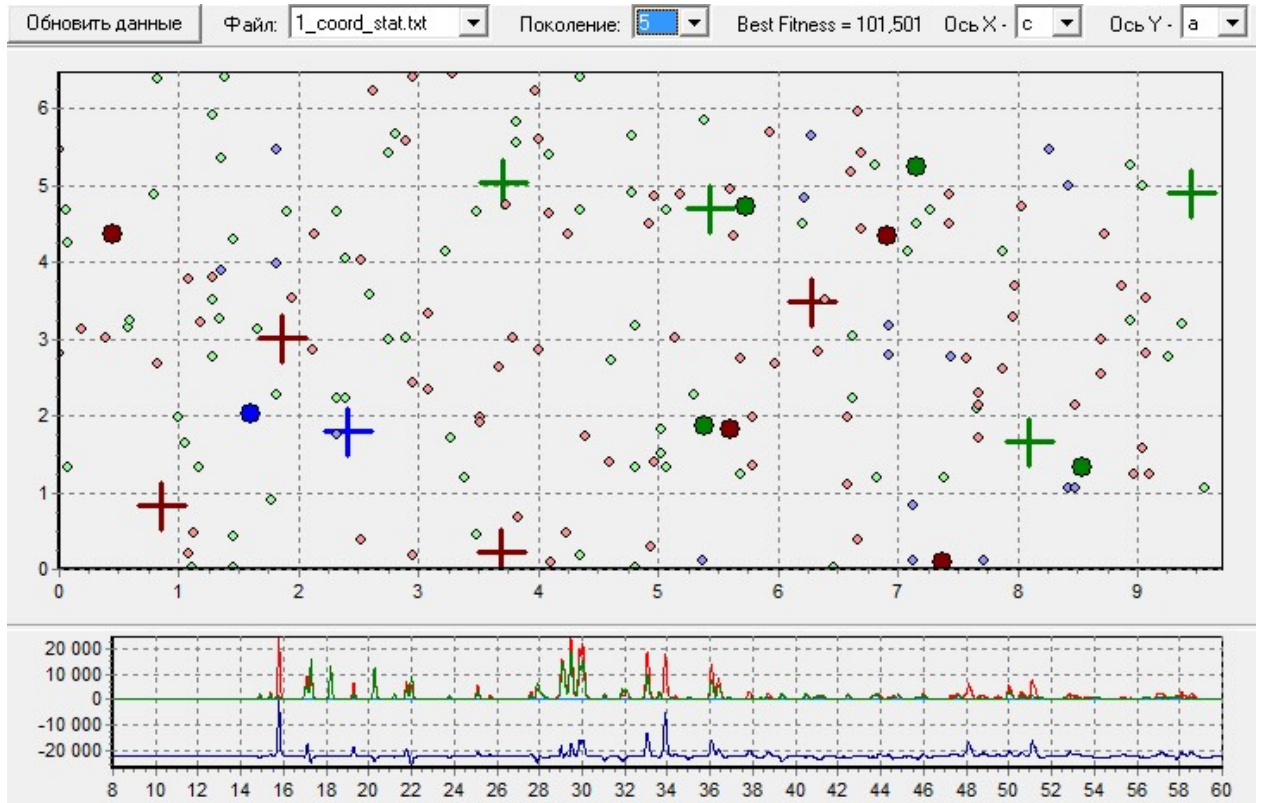


Рисунок 5 – 5-е поколение структурных моделей; R_{wp} лучшей структурной модели = 101,5%, это более удачная, в смысле R-фактора, модель

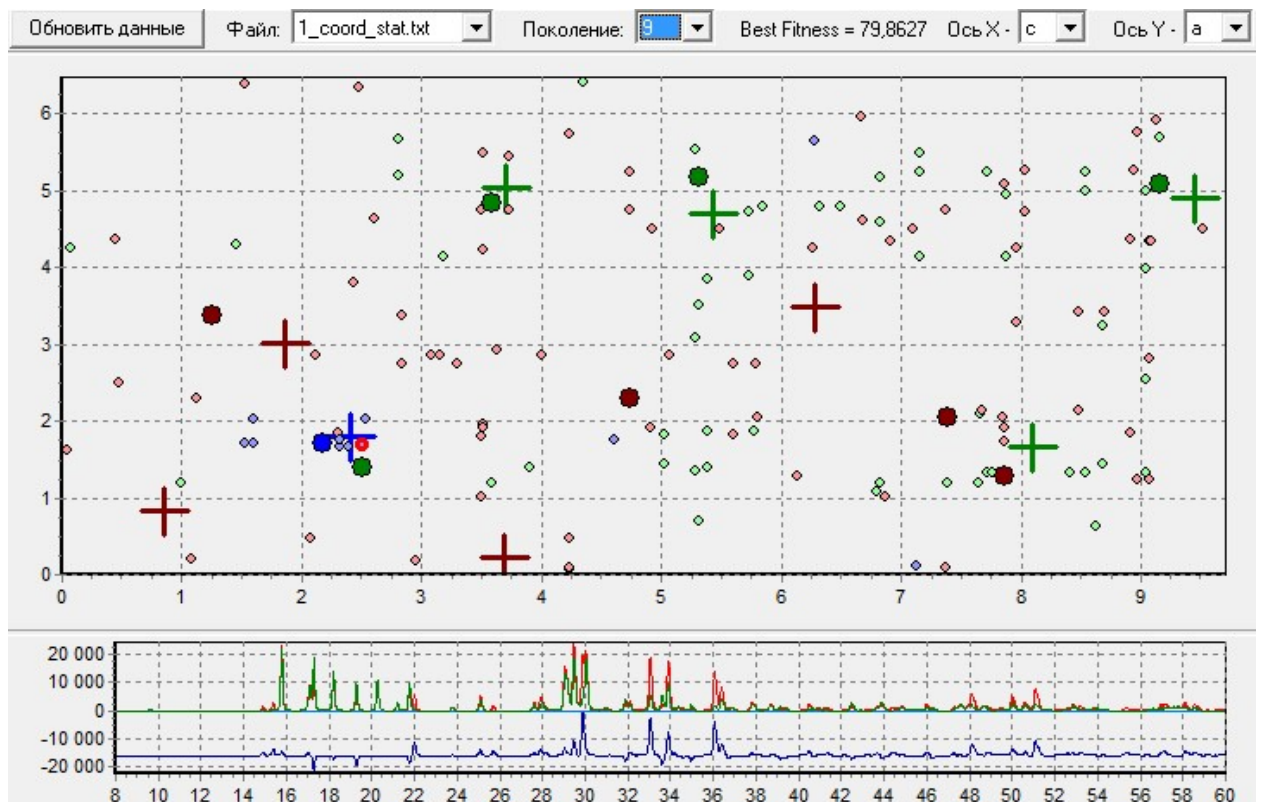


Рисунок 6 – 9-е поколение структурных моделей; R_{wp} лучшей структурной модели = 79,86%, в ней Sn(x,y) и 3 K(x,y) – вблизи истинных позиций

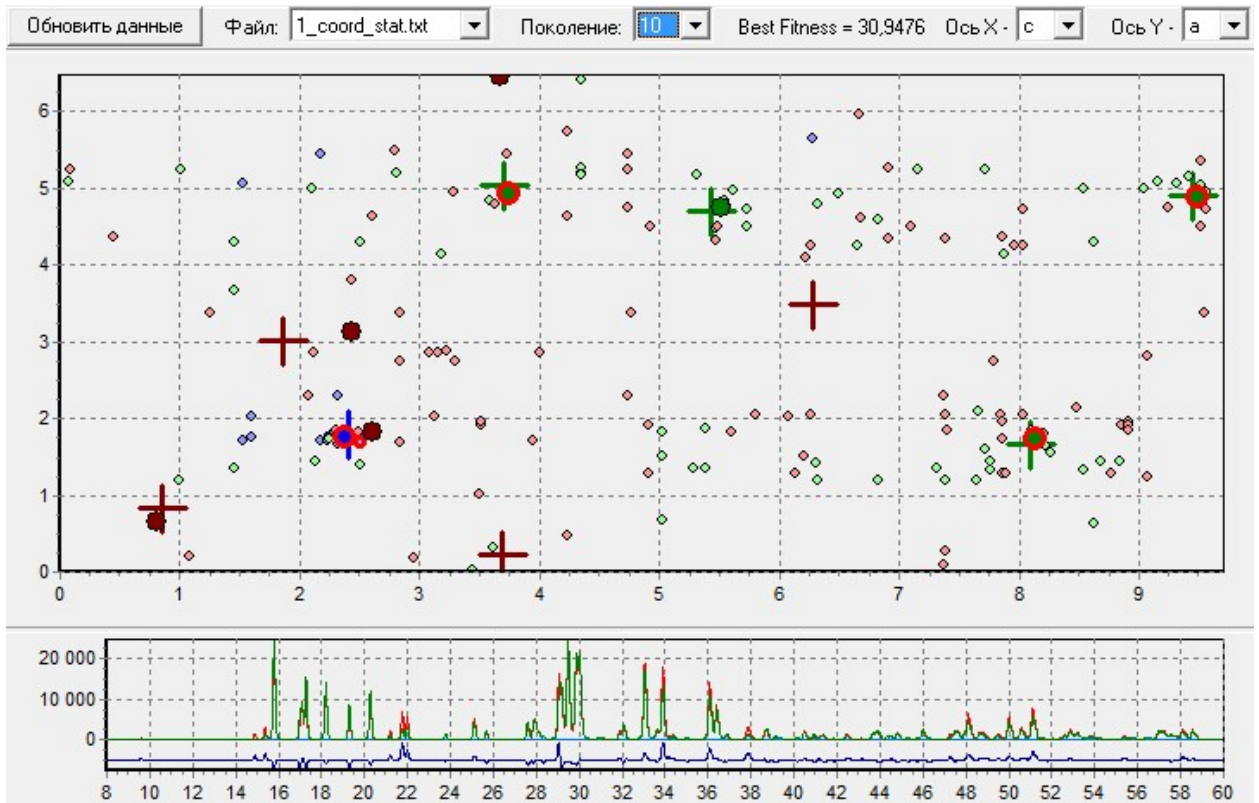


Рисунок 7 – 10-е поколение структурных моделей; R_{wp} лучшей структурной модели = 30,95%, в ней Sn(x,y,z) и 3 K(x,y,z) – в истинных позициях (красные обводы)

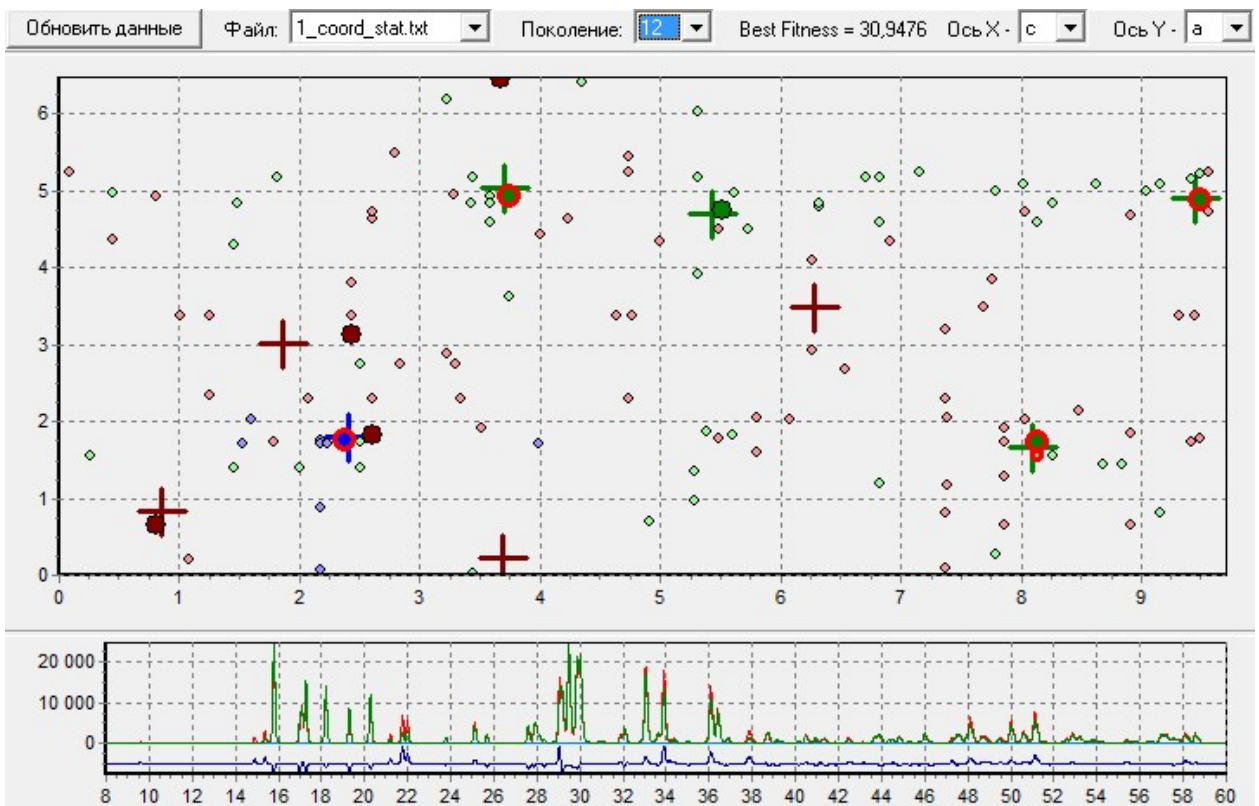


Рисунок 8 – 12-е поколение структурных моделей; R_{wp} лучшей структурной модели = 30,95%; больше моделей с атомами вблизи истинных позиций

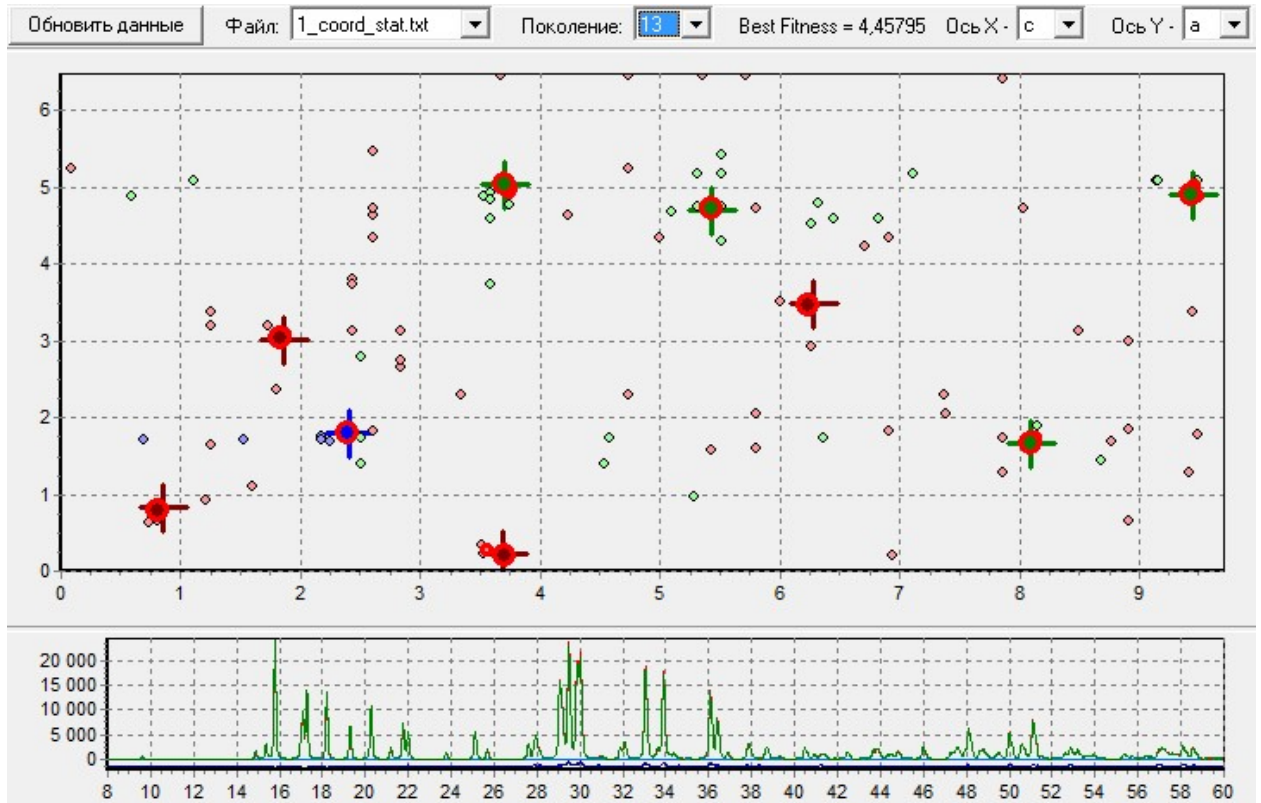


Рисунок 9 – 13-е поколение структурных моделей; R_{wp} лучшей структурной модели = 4,45%, в ней все остальные атомы одновременно попали в истинные позиции (за счет удачного скрещивания + МНК)

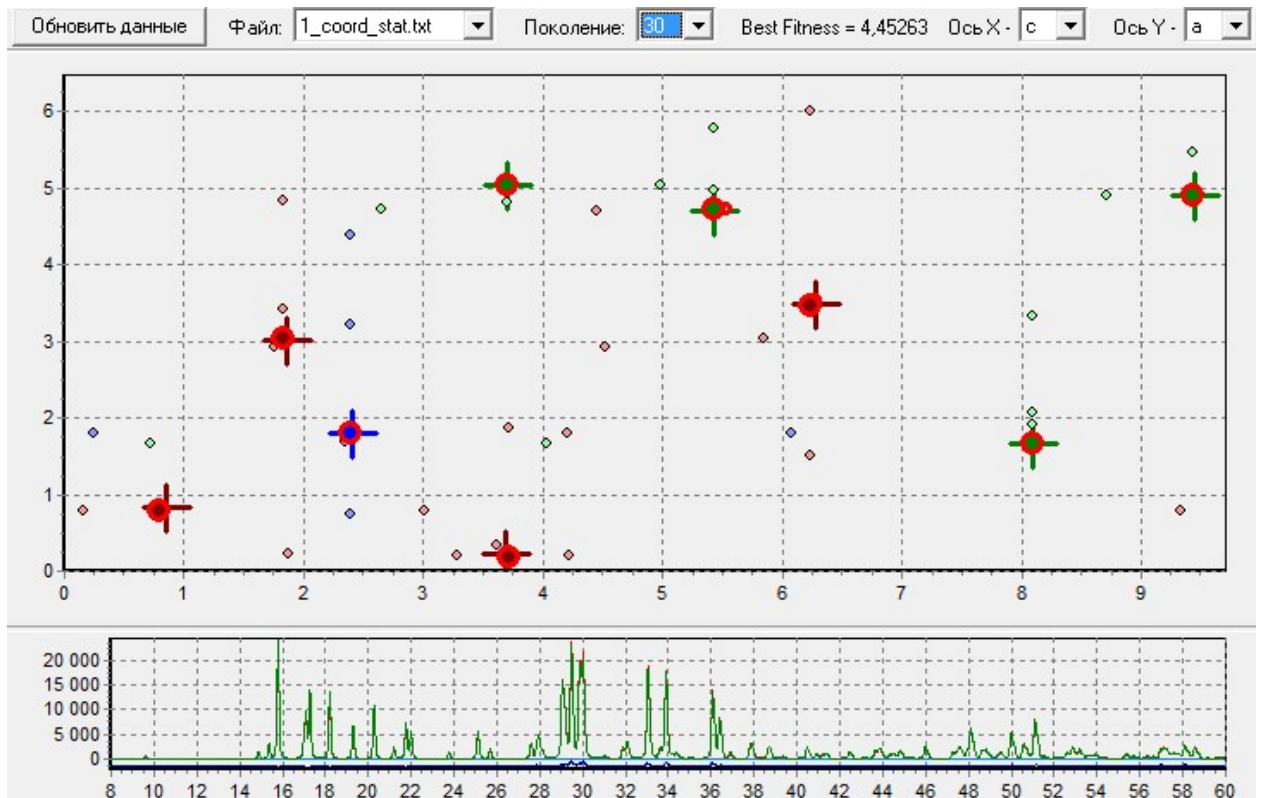


Рисунок 10 – 30-е поколение структурных моделей; R_{wp} лучшей структурной модели = 4,45%, большинство атомов из 50-ти моделей в истинных позициях (их малые кружки перекрыты большими кружками лучшей модели)

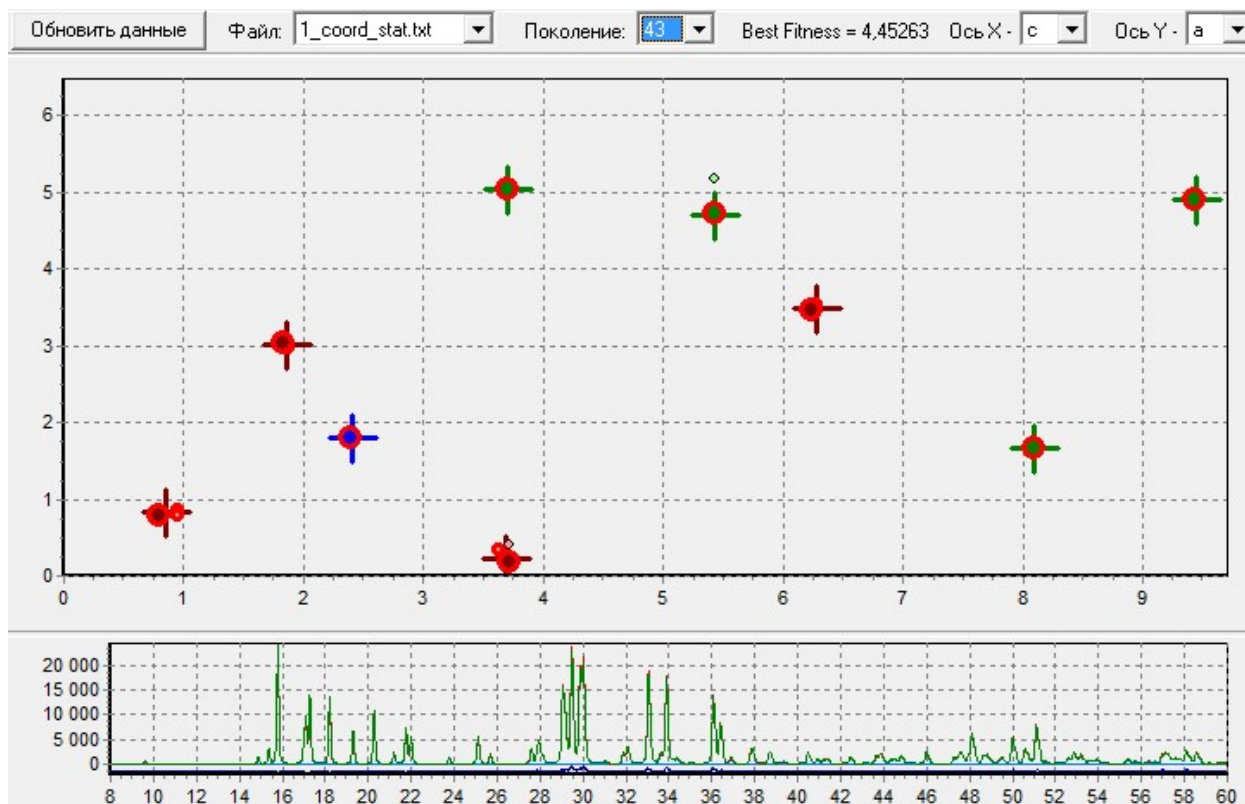
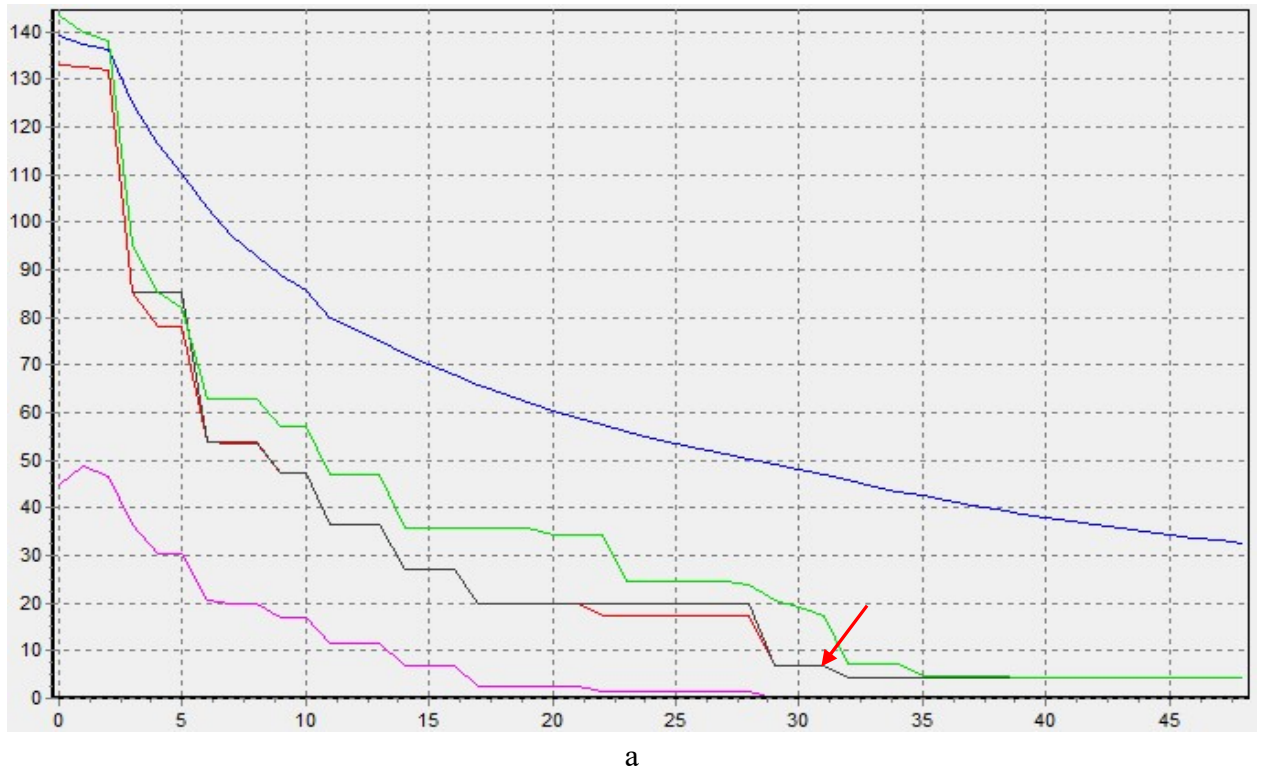


Рисунок 11 – 43-е поколение структурных моделей; R_{wp} лучшей структурной модели = 4,45%, практически все атомы из 50-ти моделей в истинных позициях

Б.2 Пример медленной сходимости

На рисунке 12 приведены графики сходимости МПГА при определении структуры K_4SnO_4 . В данном случае представлен пример более медленной сходимости – истинная структура была найдена на 32-ом поколении работы МПГА. На рисунках 13-20 приведены карты распределения атомов на основных этапах определения структуры во время этого пуска МПГА.



а



б

Рисунок 12 – Графики сходимости лучшей структурной модели K_4SnO по МПГА: а) на управляющем ядре; б) на лучшем из вычислительных ядер (ядро №3).

Ось абсцисс – номер поколения, ось ординат – пригодность: $Y = R_{wp} + \text{штраф}$; Y – красная линия (черная – с уточнением по МНК); $Y_{\text{сред.}}$ – синяя линия; штраф за некорректные межатомные расстояния – фиолетовая линия. Структура определена на 12-м поколении эволюции с $Y = R_{wp} < 5\%$ (штраф = 0)

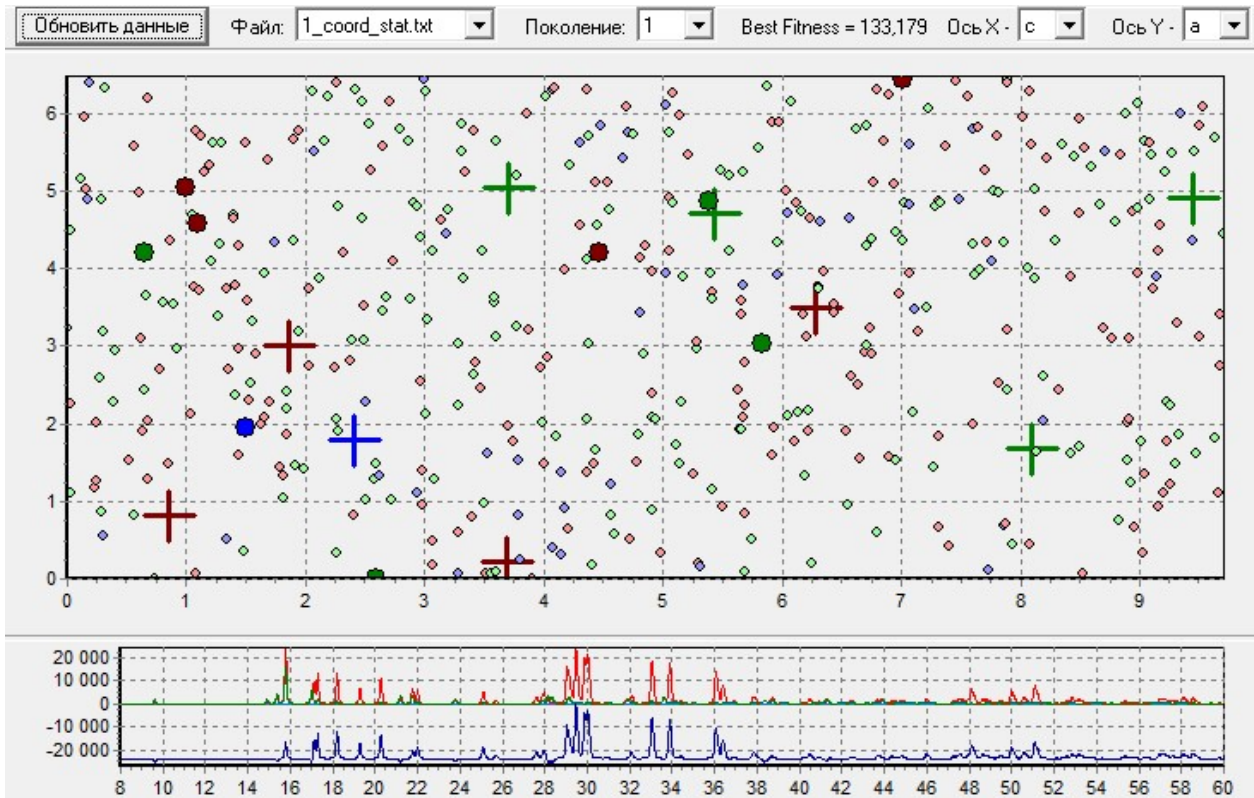


Рисунок 13 – Случайно сгенерированное 1-е поколение структурных моделей (50 штук): цвета атомов: Sn – синий, К – зеленый, О – красный; кресты – их истинные позиции в структуре; большие кружки – атомы лучшей модели; кружки с красной обводкой – атомы в истинных позициях, совпавшие по (x,y,z);

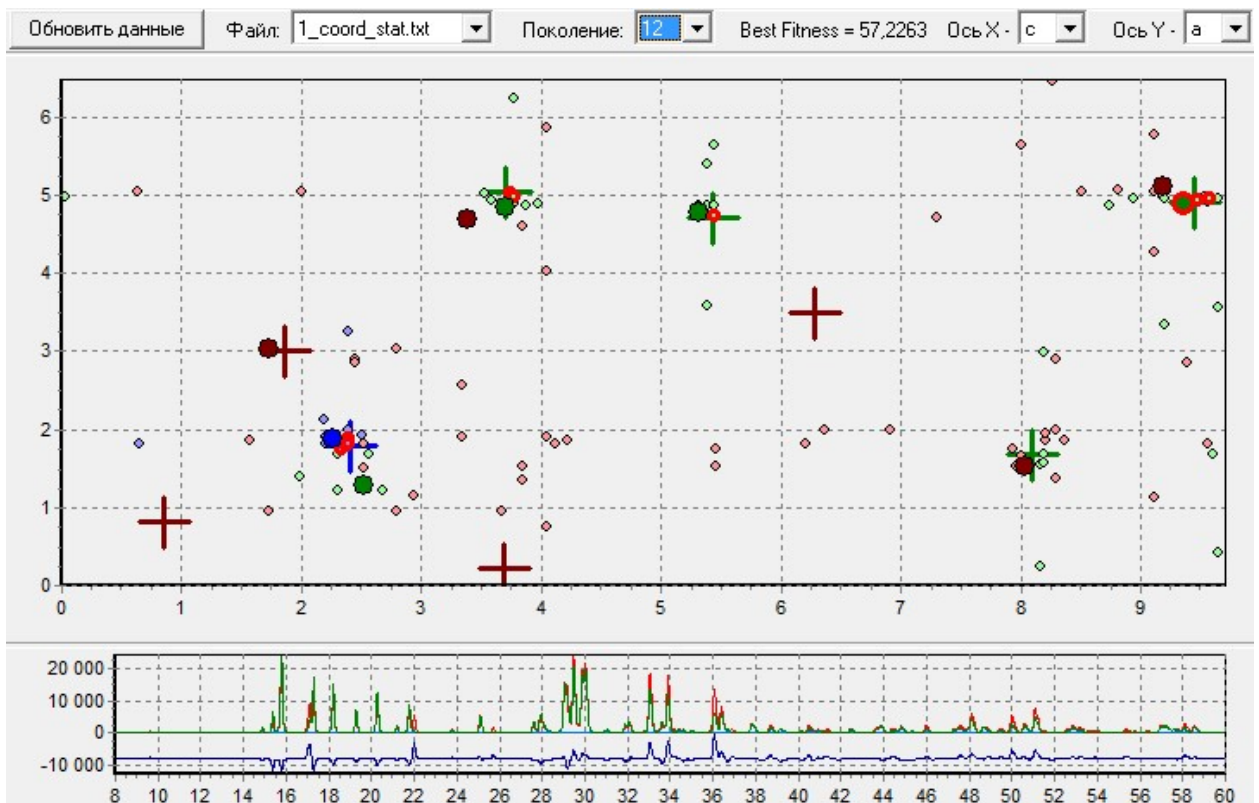


Рисунок 14 – 12-е поколение структурных моделей; R_{wp} лучшей структурной модели = 57,22%, в ней Sn, 2 атома К и 1 О в истинных позициях (x,y) и 1 атом К - в (x,y,z)

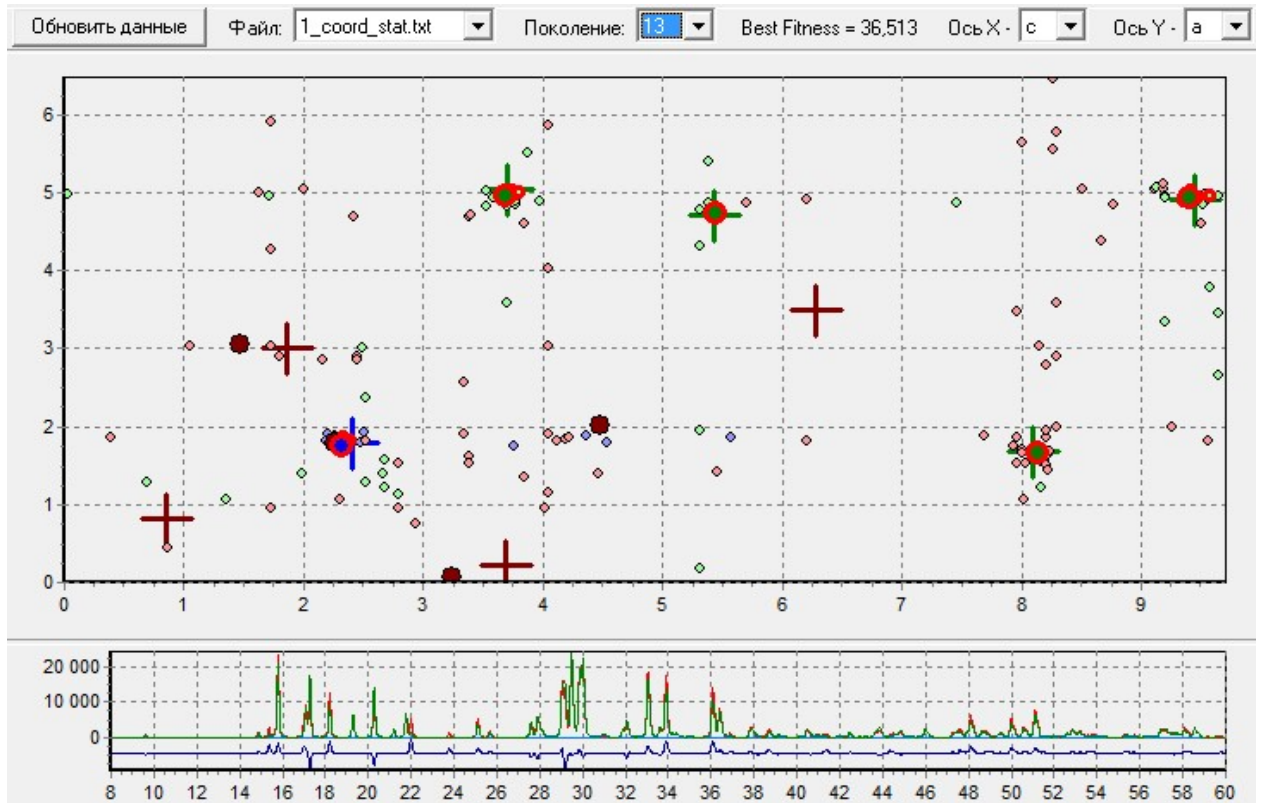


Рисунок 15 – 13-е поколение структурных моделей; R_{wp} лучшей структурной модели = 36,51%, в ней Sn и 4 атома К в истинной позиции (x,y,z)

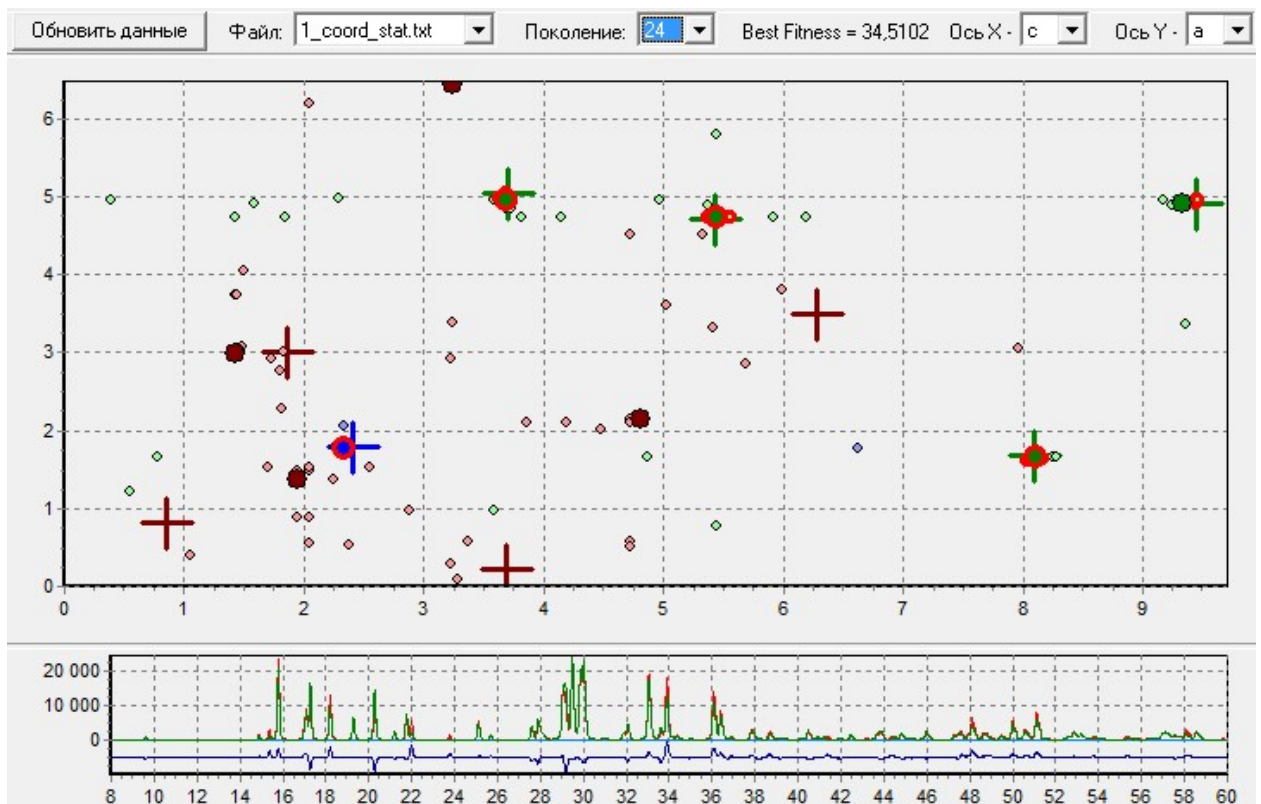


Рисунок 16 – 24-е поколение структурных моделей; R_{wp} лучшей структурной модели = 34,5%, в ней также Sn и 4 атома К в истинных позициях (x,y,z); часть атомов О вблизи истинных позиций

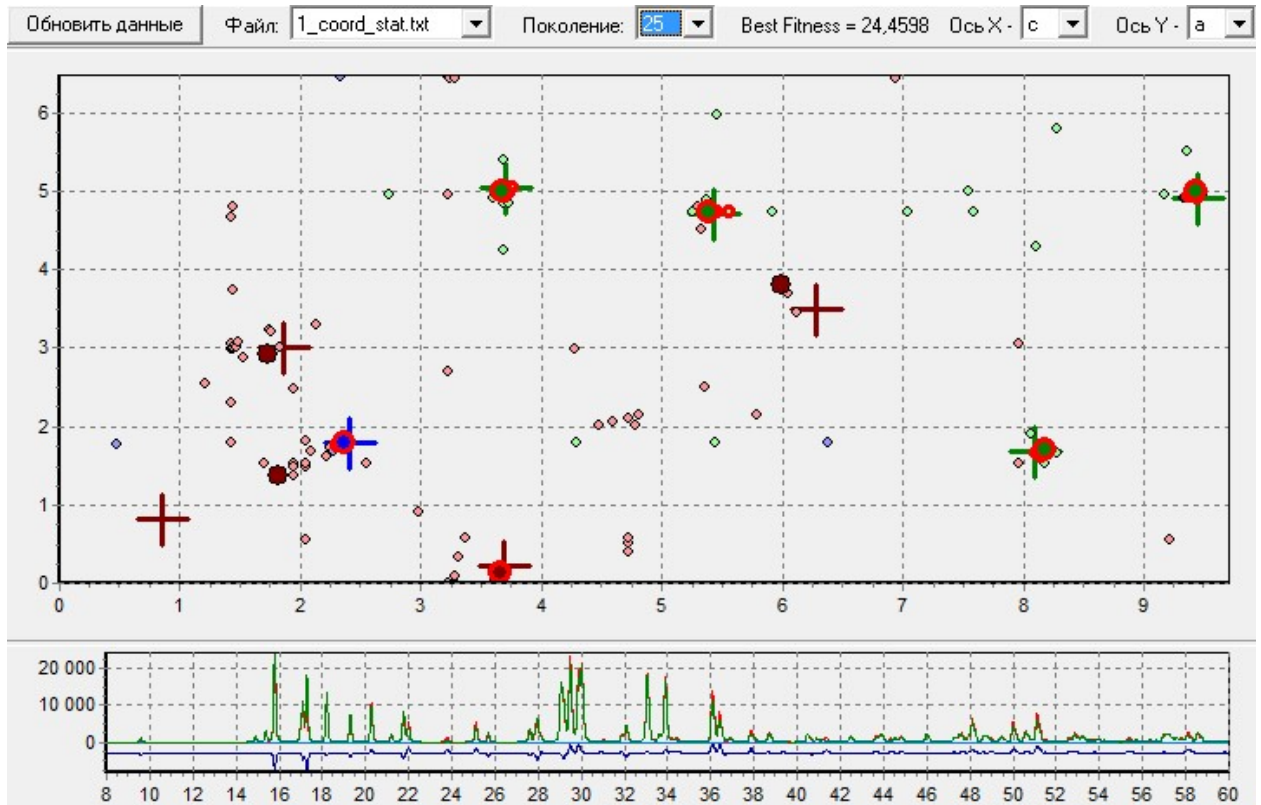


Рисунок 17 – 25-е поколение структурных моделей; R_{wp} лучшей структурной модели = 24,45%, в ней Sn, 4 атома К и 1 О в истинных позициях (x,y,z); часть атомов О вблизи истинных позиций

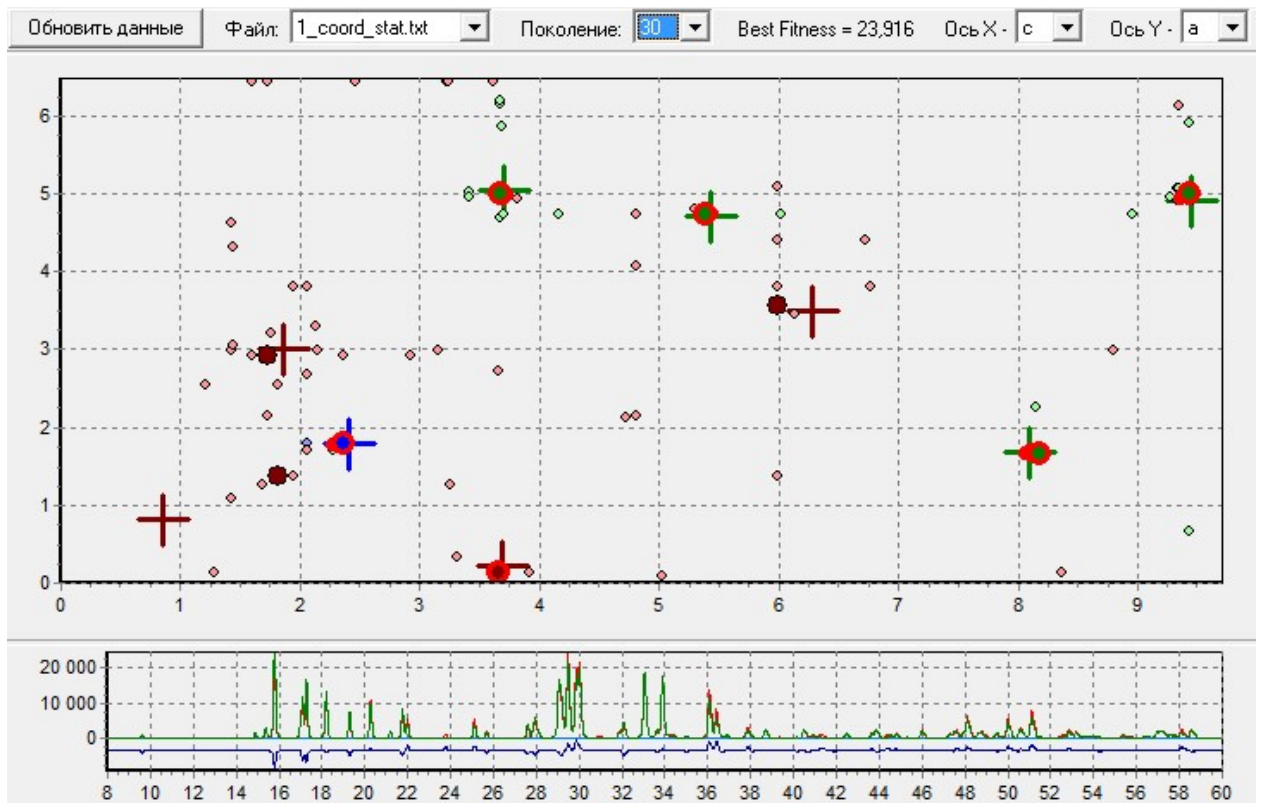


Рисунок 18 – 30-е поколение структурных моделей; R_{wp} лучшей структурной модели = 23,91%, в ней Sn, 4 атома К и 1 О в истинных позициях (x,y,z); большая часть атомов О вблизи истинных позиций

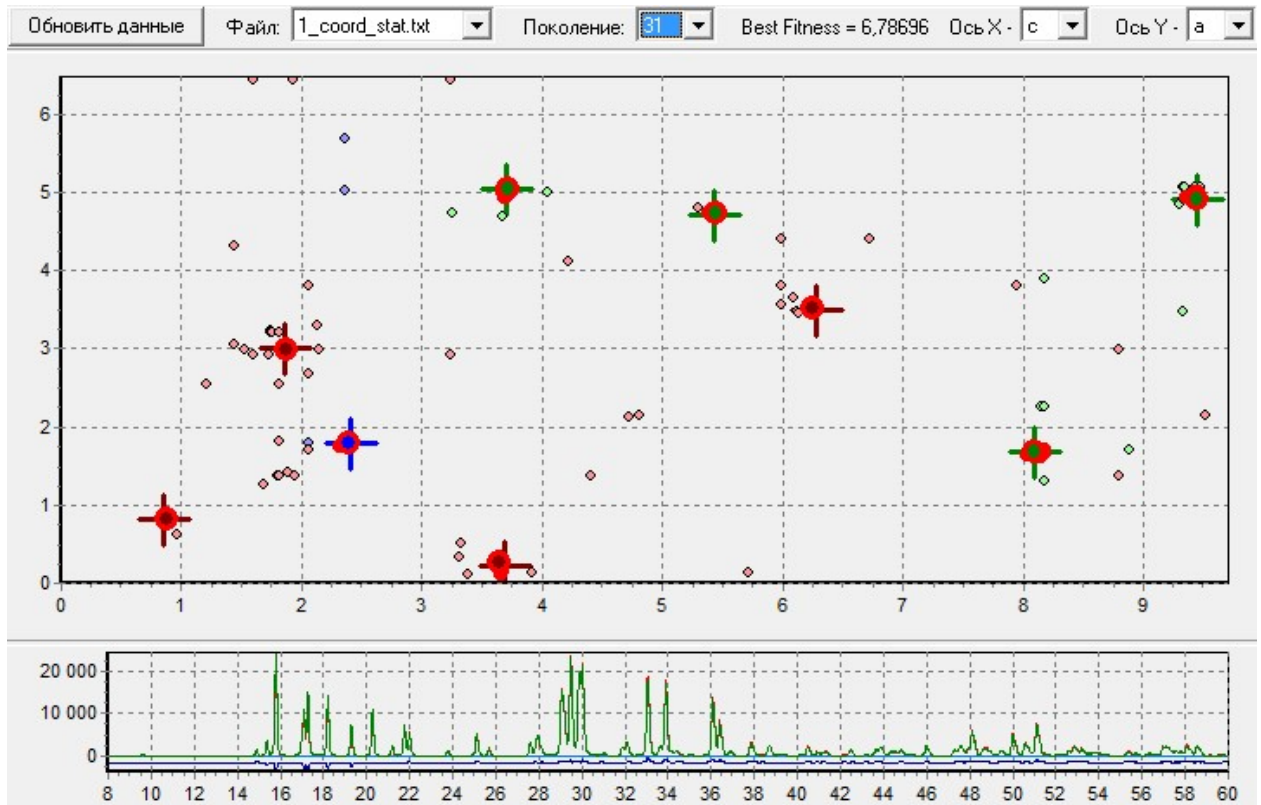


Рисунок 19 – 31-е поколение структурных моделей; R_{wp} лучшей структурной модели = 6,78%, все атомы в истинных позициях (x,y,z); большая часть атомов др. моделей вблизи истинных позиций

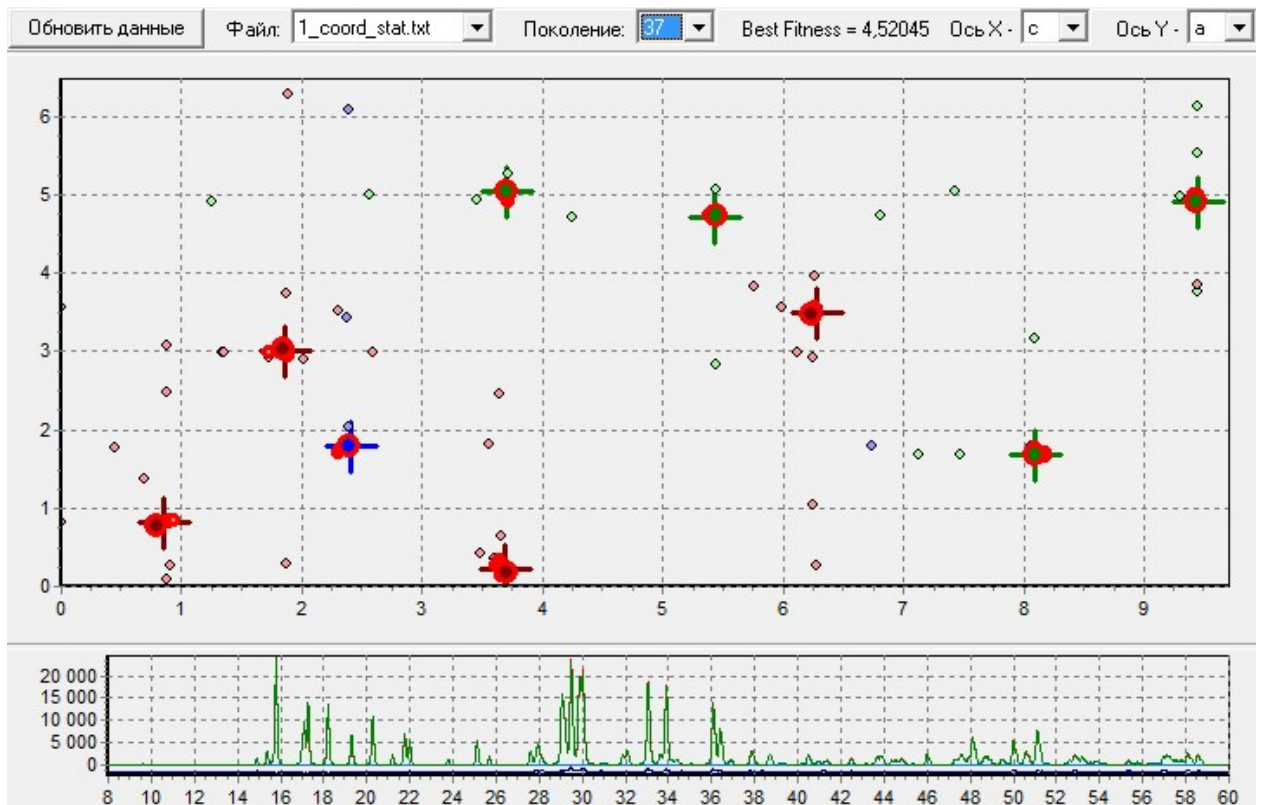


Рисунок 20 – 37-е поколение структурных моделей; R_{wp} лучшей структурной модели = 4,52%, все атомы в истинных позициях (x,y,z); все тяжелые атомы - в центрах крестов (доуточнены)

Б.3 Выводы

Сходимость ГА обусловлена тем фактом, что вероятность случайной генерации координат атомов в окрестности истинных атомных позиций в структуре с радиусом $\sim 0,03$ (размер крестов в долях осей ячейки) составляет $p \sim 1/33$. В [1] экспериментально показано, что структурные модели с такой локализацией атомов хорошо уточняются по МНК в методе Ритвельда (по крайней мере, при размере осей ячейки до 10-15 ангстрем).

В исходной популяции, даже всего из 50 структурных моделей в среднем присутствует по 1-2 таких координаты для каждого атома (в разных моделях), и сходимость обуславливается за счет их эволюционного накопления в структурных моделях популяции (происходящее поэтапно, от тяжелых атомов к легким) за счет удачных парных скрещиваний и МНК (после накопления – часто сразу по нескольким атомам).

Это связано со статистической чувствительностью R-фактора к попаданию отдельных атомных координат и целых атомов в истинные позиции в структуре (при случайном распределении прочих атомов) [1], дополняемой исключением кристаллохимически некорректных структурных моделей (путем штрафа к R-фактору).