Journal of Siberian Federal University. Chemistry 2 (2013 6) 180-191
~ ~ ~

УДК 539.26:519.65:519.68

Эволюционный метод моделирования кристаллической структуры вещества по данным порошковой дифракции

Я.И. Якимов[а], С.Д. Кирик[а], Е.С. Семенкин[а], Л.А. Соловьев[б], И.С. Якимов[а]*
[а]Сибирский федеральный университет, Россия 660041, Красноярск, пр. Свободный, 79
[б]Институт химии и химической технологии СО РАН Россия 660049, Красноярск, К. Маркса, 42

Received 23.09.2012, received in revised form 24.12.2012, accepted 18.01.2013

Предложен эволюционный метод моделирования атомной кристаллической структуры химических соединений из данных порошковой дифракции, интегрирующий метод полнопрофильного структурного анализа и генетический алгоритм глобальной оптимизации. Проведено экспериментальное исследование эволюционного механизма, обеспечивающего сходимость нового гибридного эволюционного метода структурного анализа при ab initio определении кристаллической структуры ряда химических соединений.

Ключевые слова: кристаллическая структура вещества, методы глобальной оптимизации, эволюционные и генетические алгоритмы, порошковая дифракция, рентгеноструктурный анализ поликристаллов, полнопрофильный метод Ритвельда.

Введение

Информация об атомно-кристаллическом строении необходима для объяснения и прогнозирования физических и химических свойств веществ и материалов. Данная информация накапливается в базах кристалло-структурных данных (БД) [1, 2] и включает координаты и параметры тепловых колебаний атомов в симметрически независимой части элементарной ячейки кристаллического вещества. Основным методом для изучения структуры служит дифракционный анализ монокристаллов, однако у значительной части веществ, получаемых в форме поликристаллов, структура изучается методами порошковой дифракции. Структурное исследование включает этапы определения параметров элементарной ячейки и пространствен-

© Siberian Federal University. All rights reserved
* Corresponding author E-mail address: i-s-jakimov@yandex.ru

ной группы симметрии, определения модели атомной кристаллической структуры и ее оптимизацию. Оптимизация выполняется путем решения прямой задачи моделирования экспериментального дифракционного спектра (дифрактограммы) вещества его расчетным спектром, вычисляемым из уточняемой структуры и параметров профиля дифрактограммы. Как правило, для этого используется полуэмпирический метод полнопрофильного уточнения дифрактограмм (метод Ритвельда) [3], основанный на нелинейном МНК.

Проблемой является нахождение структурной модели. Определение параметров модели атомной структуры по порошковой дифрактограмме представляет обратную задачу моделирования, когда известно множество возможных моделей и надо выбрать конкретную модель на основе дифрактограммы и дополнительных данных о веществе. Перспективным направлением ее решения служит применение численных стохастических методов глобальной оптимизации, осуществляющих «сборку» модельных структур из атомов и молекул в кристаллической ячейке вещества по критерию соответствия расчетной дифрактограммы с экспериментальной. В последние годы наблюдается неуклонный рост числа опубликованных структур, решенных методами глобальной оптимизации [4]. Наиболее развиты методы имитации отжига, реализованные, например, в широко известных структурных программах Fox [5] и DASH [6].

Эволюционный генетический алгоритм (ГА) [7] – альтернативный метод глобальной оптимизации, он моделирует механизм биологического естественного отбора, выполняемого над множеством (популяцией) строк из пространства допустимых значений параметров оптимизируемой функции, включая их перекрестное скрещивание, вероятностные мутации и селекцию новых поколений популяции по целевой функции.

Несмотря на то что оба метода начали адаптироваться для структурного анализа практически одновременно [8, 9] и с помощью ГА найден ряд структур новых веществ [10], этот метод еще недостаточно развит для широкого применения. В [11] отмечено, что в публичном доступе нет ни одной структурной программы ГА, хотя некоторые исследования выявили его мощные возможности, в частности неявный параллелизм. Причиной служит недостаточное понимание эволюционного механизма ГА при поиске структуры, и ГА в значительной степени используется по принципу «черного ящика».

Метод эволюционного структурного моделирования

В данной статье описывается метод эволюционного структурного моделирования (ЭСМ), представляющий развитие предложенного нами ранее генетического метода поиска структурных моделей [12]. Особенностью ЭСМ относительно прочих реализаций ГА в структурном анализе является его двухуровневая организация, обеспечивающая гибридизацию ГА с методом полнопрофильного уточнения структурных моделей. Сводные результаты апробации ЭСМ, полученные при определении «с нуля» 21 кристаллической структуры веществ разной сложности, приведены в табл. 1. Наиболее сложной из них стала структура соединения $Er_{10}W_2O_{21}$ с 16 атомами в общих позициях независимой части ячейки (48 атомных координат).

Таким образом, при поиске структуры из 6-7 атомов, расположенных в общих, симметрически независимых позициях ячейки, в среднем тестируется всего около 3000 структурных моделей. Очевидно, это очень мало для оптимизации 19 параметров любым методом. Заметим,

Таблица 1. Сводка результатов моделирования 21 структуры разных веществ по ЭСМ

Среднее кол-во координат атомов	Средний размер популяций ЭСМ1 / ЭСМ2	Среднее кол-во циклов ЭСМ1-ЭСМ2	Среднее кол-во поколений эволюции ЭСМ1	Среднее кол-во испытанных структур
19	100 / 7	4,25	26	3190

например, что при поиске тех же структур с помощью упомянутых программ Fox- и DASH-метода имитации отжига тестируются 10^6-10^8 структурных конфигураций. Для объяснения этого факта в работе исследуется эволюционный механизм, обеспечивающий сходимость ЭСМ при решении обратной задачи моделирования атомной кристаллической структуры вещества из экспериментальных данных порошковой дифракции.

Обратная задача моделирования кристаллической структуры из данных порошковой дифракции может быть формализована в виде следующей задачи оптимизации. Экспериментальная дифрактограмма многофазного образца представляют собой дискретную последовательность $\{\theta_j, Y_j\}$, где θ_j – угол дифракции, Y_j – измеренная интенсивность. Имеется некоторый класс параметрических функций $Y^{мод}(\boldsymbol{P},\theta)$, моделирующих дифрактограмму:

$$Y^{мод.}(\boldsymbol{P},\theta_j)=\sum_{\alpha}^{m} S_\alpha \sum_{hkl\,\alpha}^{n_\alpha} \Omega_\alpha(\boldsymbol{P}_{L,\alpha},\theta_j-\theta_{hkl,\alpha})\cdot I^{расч.}_{hkl,\alpha}(\boldsymbol{P}_{S,\alpha},\theta_j)+B(\boldsymbol{P}_B,\theta_j), \qquad (1)$$

где Ω_α – параметрическая функция профиля дифракционных линий с индексами hkl с центрами в точках $\theta_{hkl,\alpha}$ фазы α, зависящая от набора профильных параметров $\boldsymbol{P}_{L,\alpha}$ (центры, полуширина, форма, асимметрия линий и т.п.); $I^{расч.}_{hkl,\alpha}$ – расчетная интегральная интенсивность линий, зависящая от набора параметров структуры $\boldsymbol{P}_{S,\alpha}$ фазы α. В качестве параметров структуры могут использоваться координаты отдельных атомов и жестко связанных групп атомов (молекул и структурных фрагментов) и параметры тепловых колебаний атомов. B – функция фона, зависящая от профильных параметров фона $\boldsymbol{P}_B$, $\boldsymbol{P}=\boldsymbol{P}_{L,\alpha}\times\boldsymbol{P}_{S,\alpha}\times\boldsymbol{P}_B \in D$. Пространство возможных значений параметров D определяется особенностями конкретной задачи и обычно является прямоугольным из R^n. Интегральная интенсивность I рефлекса hkl кристаллической фазы α вычисляется согласно [13]:

$$I^{расч.}_{hkl,\alpha}=\left[\frac{I_0\lambda^3}{32\pi r}\frac{e^4}{m_e^2c^4}\right]\cdot\left[\frac{M_{hkl}}{2V_\alpha^2}\left|F_{(hkl)\alpha}\right|^2\left(\frac{1+\cos^2 2\theta\cos^2 2\theta_m}{\sin^2\theta\cos\theta}\right)\right]\cdot\left[\frac{W_\alpha}{\rho_\alpha\mu_m^*}\right], \qquad (2)$$

где I_0 – интенсивность падающего монохроматического рентгеновского излучения; e – заряд электрона; m_e – масса электрона; r – расстояние между образцом и детектором; c – скорость света; M – фактор повторяемости, V – объем элементарной кристаллической ячейки; θ и θ_m – углы дифракции для рефлекса с индексами hkl и монохроматора; W_α и ρ_α – массовое содержание и плотность фазы α соответственно, μ^*_m – массовый коэффициент поглощения образца, $|F|^2$ – структурный фактор рефлекса hkl фазы α:

$$|F_{(hkl),\alpha}|^2=(|\sum_j \eta_j\cdot f_j\sum_m e^{-2\pi i(hx_{jm}+ky_{jm}+lz_{jm})}\cdot t_{jm}|^2)_\alpha, \qquad (3)$$

где j, m – суммирование по атомам, соответственно, независимой и зависимой части ячейки; x_{jm}, y_{jm}, z_{jm} – координаты атомов в координатной системе векторов ***a, b, c*** элементарной ячейки; t_{jm} – анизотропный тепловой фактор атома j позиции m (или изотропный t_j); η_j – коэффициент заполнения атомом j кристаллографической позиции; f_j – фактор рассеяния излучения атомом j. Координаты атомов зависимой части ячейки вычисляются из координат атомов независимой части с помощью операторов пространственной группы симметрии кристаллической фазы вещества.

Требуется определить функцию (1) из класса $Y(\boldsymbol{P}, \theta)$ и вектор оптимальных параметров $\boldsymbol{P}^*$ из D, доставляющих минимум избранному функционалу качества (разность модельной и экспериментальной дифрактограммы в некоторой метрике). Вместо обычного функционала МНК полнопрофильного анализа по Ритвельду выбран функционал метода МПР – полнопрофильного анализа «минимизации производной разности» [14]:

$$MF = \sum_j w_1 \left(\frac{\partial}{\partial \theta} \left(Y_j - Y^{мод}(\boldsymbol{P}, \theta_j) \right) \right)^2 + \sum_j w_2 \left(\frac{\partial^2}{\partial \theta^2} \left(Y_j - Y^{мод}(\boldsymbol{P}, \theta_j) \right) \right)^2 \xrightarrow{\boldsymbol{P}} 0, \qquad (4)$$

где w_i – веса, отражающие статистику измерений интенсивности.

За меру качества решения взят взвешенный профильный R-фактор МПР:

$$R_{МПР} = 50 \cdot \sqrt{\frac{\sum_{k=1}^{2} \sum_{i=m+1}^{N-m} w_i^k \left[\sum_{j=-m}^{m} s_j^k \left(Y_{i+j} - Y^{мод}(P, \theta_{i+j}) \right) \right]^2}{\sum_{k=1}^{2} \sum_{i=m+1}^{N-m} w_i^k \left[\sum_{j=-m}^{m} s_j^k Y_{i+j} \right]^2} + \frac{\sum_{i=m+1}^{N-m} w_i \left[Y_i - \sum_{j=-m}^{m} s_j^0 \left(Yo_{i+j} \right) \right]^2}{\sum_{i=m+1}^{N-m} w_i Y_i^2}}, \qquad (5)$$

где $w_i^k = \left[\sum_{j=-m}^{m} (s_j^k)^2 (\sigma_{i+j})^2 \right]^{-1}$, s_j^k – коэффициенты Савицкого-Голея для оценки производной k-го порядка для интервала свертки профиля $[-m, m]$, N – число точек профиля, σ_i – дисперсия наблюдаемой интенсивности профиля Y_i.

Для решения задачи предложен гибридный двухуровневый эволюционный алгоритм, интегрирующий генетический метод глобальной оптимизации и проблемно-ориентированный метод МПР. Индивиды эволюции ЭСМ первого уровня представляют собой параметрические строки типа ***P*** – двоичные представления округленных значений вектора искомых профильных и/или структурных параметров, используемых МПР, в частности координат x, y, z атомов вещества относительно осей его элементарной ячейки. Округленные значения строк задаются узлами сетки, равномерно покрывающей заданный интервал поиска параметров. Минимизируемая целевая функция для ***P*** есть $R_{МПР}$ (5). Задача первого уровня ЭСМ – найти удовлетворительные, в смысле величины R-фактора, приближения параметров ***P*** для передачи лучших решений на следующий, 2-й уровень ЭСМ. Алгоритм ЭСМ 2-го уровня используется для выработки стратегии уточнения по МНК МПР параметров ***P***, представляющую собой последовательность локальных спусков на гиперповерхности R-фактора. Индивиды эволюции ЭСМ 2-го уровня представляют управляющие битовые строки типа ***B***, где каждая строка бит определяет группу уточняемых по МНК на данном поколении профильных и/или структурных параметров из $\boldsymbol{P}_i$. ***B***-индивиды подвергаются генетическим операциям скрещивания, мутации и др., целевая функция учитывает результаты МНК:

$$F = \left[R(\boldsymbol{P}_i')/R(\boldsymbol{P}_i) + p\right]^{-1}, \quad (6)$$

где p – штраф за отсутствие сходимости, $\boldsymbol{P}_i'$ – уточненные по МНК строки $\boldsymbol{P}_i$.

Оба уровня ЭСМ работают циклически до сходимости одной из структур к целевому значению R (или до конца вычислительного ресурса). Описанный алгоритм реализован в программе моделирования кристаллической структуры, использованной для экспериментального исследования механизма глобальной сходимости метода ЭСМ.

Экспериментальное исследование сходимости метода ЭСМ

Важным фактором сходимости является размер вычислительной сетки, используемой на 1-м уровне ЭСМ. Результаты сходимости двухуровневого ЭСМ при моделировании ряда структур с округлением координат атомов по разным сеткам разбиения ячейки показали, что при размерах осей ячейки до 10–15 ангстрем (большинство неорганических соединений) оптимальной является 5-битная сетка. В качестве примера в табл. 2 представлены результаты моделирования известной структуры Ba_2CrO_4 (параметры решетки: a=7.67Å, b=5.89Å, c=10.39Å, $\alpha=\beta=\gamma=90^o$, пространственная группа симметрии $Pn2_1a$), в которой требовалось определить 21 координату 7 атомов в независимой части ячейки. Для исключения влияния микроструктурных особенностей (текстуры, размера кристаллитов и т.п.) и установления точного целевого значения R-фактора использована расчетная зашумленная дифрактограмма (дифрактограммы прочих веществ экспериментальные). Поскольку ЭСМ-стохастический метод, для получения достаточно достоверного результата осуществлено по десять независимых запусков программы ЭСМ для каждого размера сетки.

Для изучения преимуществ двухуровневой организации алгоритма по сравнению с одноуровневой выполнена оценка сходимости по двухуровневому ЭСМ относительно ЭСМ 1-го уровня. В обоих случаях использованы одинаковые управляющие параметры поиска, в частности размер популяции из 100 структур. Для сопоставимости результатов в конце работы ЭСМ 1-го уровня выполнялось уточнение координат атомов лучшей структуры по МПР. В табл. 3 представлены результаты сходимости для десяти независимых запусков программы ЭСМ в каждом из этих режимов при моделировании структуры Ba_2CrO_4.

На рис. 1 показаны типичные графики сходимости. При использовании только ЭСМ 1-го уровня (слева) целевой R-фактор (1,4 %) достигнут и структура найдена на 102-м, а в двухуровневом ЭСМ (справа) – на 39-м поколении эволюции. На правом графике рис. 1 штриховыми полосами разделены итерационные циклы ЭСМ1 – ЭСМ2, резкие спады R-фактора лучшей структуры на участках ЭСМ2 иллюстрируют действие эволюционной локальной минимизации по МПР. Таким образом, именно применение двухуровневого алгоритма обеспечивает эффективность ЭСМ.

Таблица 2. Сходимость моделирования структуры Ba_2CrO_4 с разными размерами сетки

Сетка 4 бит		Сетка 5 бит		Сетка 6 бит	
Количество успешных пусков (%)	Среднее время (мин)	Количество успешных пусков (%)	Среднее время (мин)	Количество успешных пусков (%)	Среднее время (мин)
20	40	70	30	40	35

Таблица 3. Характеристики сходимости ЭСМ 1-го уровня и двухуровневого ЭСМ

Критерии эффективности	ЭСМ 1-го уровня	2-уровневый ЭСМ
Количество успешных пусков (%)	10	60
Среднее время (мин)	23	30

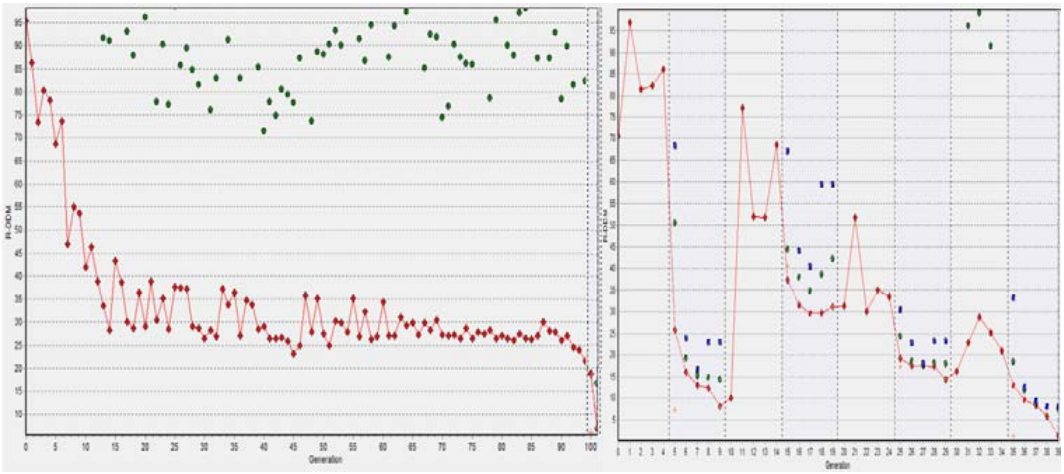

Рис. 1. Графики сходимости структуры Ba_2CrO_4: слева – одноуровневым ЭСМ, справа – двухуровневым (штриховые полосы – циклы ЭСМ1 и ЭСМ2); ось абсцисс – номер поколения, ось ординат – R-фактор МПР; красные ромбы – R-фактор лучшей структуры, зеленые кружки – средний, синие квадраты – худшей по популяции структуры

Таблица 4. Сравнение найденной по ЭСМ структуры Ba_2CrO_4 с эталонной*

Атом	X*	$X_{ЭСМ}$	|Δ|	Y*	$Y_{ЭСМ}$	|Δ|	Z*	$Z_{ЭСМ}$	|Δ|
Ba1	0,0047	0,0047	0,0000	0,2432	0,2427	0,0005	0,1893	0,1893	0,0000
Ba2	0,8519	0,8519	0,0000	0,2500	0,2500	0,0000	0,5838	0,5839	0,0001
Cr1	0,2770	0,2771	0,0001	0,2420	0,2408	0,0012	0,5800	0,5798	0,0002
O1	0,5230	0,5231	0,0001	0,1900	0,1906	0,0006	0,5770	0,5771	0,0001
O2	0,1650	0,1647	0,0003	0,2330	0,2313	0,0017	0,4260	0,4258	0,0002
O3	0,3250	0,3253	0,0003	0,9930	0,9909	0,0021	0,1570	0,1573	0,0003
O4	0,6950	0,6954	0,0004	0,0050	0,0035	0,0015	0,8500	0,8502	0,0002

Полученные координаты атомов с высокой точностью совпали с истинными: среднее абсолютное отклонение |Δ|ср.=0,00047 (табл. 4).

На рис. 2 на примере атома Ba_1 показано, как происходит сходимость атомов в их истинные позиции (x, y, z) при моделировании структуры Ba_2CrO_4 по двухуровневому ЭСМ. По оси абсцисс отложены номера поколений, по оси ординат – положение атома по одной из координат X,Y или Z в независимой части ячейки; истинное значение координаты указано горизон-

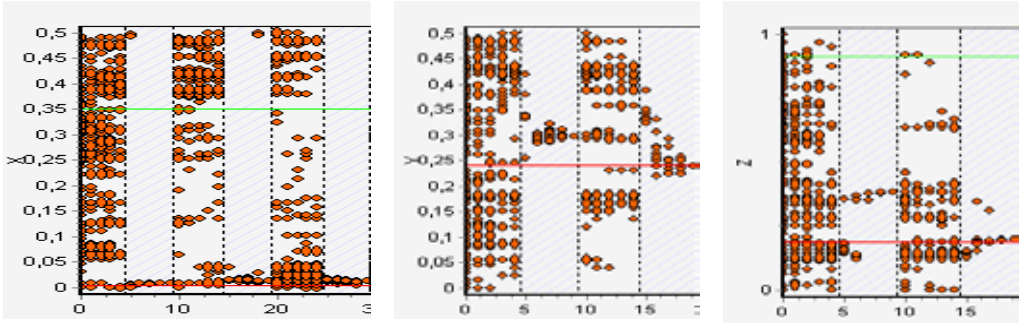

Рис. 2. Графики покоординатной сходимости атома Ba_1 к истинным значениям координат в популяции структур Ba_2CrO_4 (ось абсцисс – номер поколения, ось ординат – координата X, Y или Z атома). Вертикальным пунктиром разделена работа методов ЭСМ 1-го и 2-го уровней

тальной красной линией, вертикальные штрихованные полосы отмечают чередование 1-го и 2-го уровней ЭСМ. В процессе сходимости координаты группируются в плотные сгущения (10-кратные сгущения для наглядности выделены овалами) в окрестности истинных значений, менее плотные сгущения соответствуют локальным минимумам.

В аналогичных экспериментах с разными веществами обнаружено, что в популяциях структурных моделей сначала формируются значения координат тяжелых атомов, а затем более легких. Это в соответствии с (1)-(3) объясняется большим вкладом в дифрактограмму рассеяния от тяжелых атомов, после же их расстановки в истинные позиции на детали расчетной дифрактограммы влияет только расположение легких атомов.

Повышение сходимости при двухуровневой организации алгоритма ЭСМ обеспечивается локальной оптимизацией структурных моделей на 2-м уровне. Для изучения преимуществ эволюционного управления процессом оптимизации структур на 2-м уровне ЭСМ относительно независимого уточнения одновременно всех структурных параметров по индивидуальному методу МПР выполнялось уточнение структур с искаженной геометрией. Все атомные координаты смещались от истинных значений на величины, вычисленные генератором случайных чисел с нормальным законом распределения при заданном стандартном отклонении. В табл. 5 приведены результаты уточнения по обоим способам десяти независимо искаженных более сложных структур соединений K_4SnO_4 (27 атомных координат) и $Pd(NH_3)_2(NO_2)_2$ (12 координат неводородных атомов). Это показывает, что уточнение под управлением ЭСМ 2-го уровня обеспечивает лучшую сходимость при худших исходных приближениях по сравнению с МПР.

Важнейшим фактором сходимости при использовании эволюционных методов глобальной оптимизации является чувствительность целевой функции к накоплению элементов структуры на фоне случайных структурных вариаций. Чувствительность R-фактора МПР изучалась на основе численного эксперимента, заключавшегося в последовательной замене значений координат X, Y и Z каждого атома в случайно сгенерированных популяциях структурных моделей на их истинные значения, взятые из БД ICSD [1]. Атомы упорядочены по убыванию атомных весов. После очередной замены Z данный атом попадает в истинную позицию в структуре. На рис. 3 показано типичное распределения R-фактора, полученное после замены координа-

Таблица 5. Результаты уточнения по методам ЭСМ2 и МПР десяти структур соединений K_4SnO_4 и $Pd(NH_3)_2(NO_2)_2$ с различной величиной отклонения координат атомов

Уточняемая структура	МПР	ЭСМ2	МПР	ЭСМ2
K_4SnO_4	СКО координат 5 %		СКО координат 10 %	
Количество успешных пусков (%)	80	100	20	80
$Pd(NH_3)_2(NO_2)_2$	СКО координат 10 %		СКО координат 15 %	
Количество успешных пусков (%)	80	100	20	80

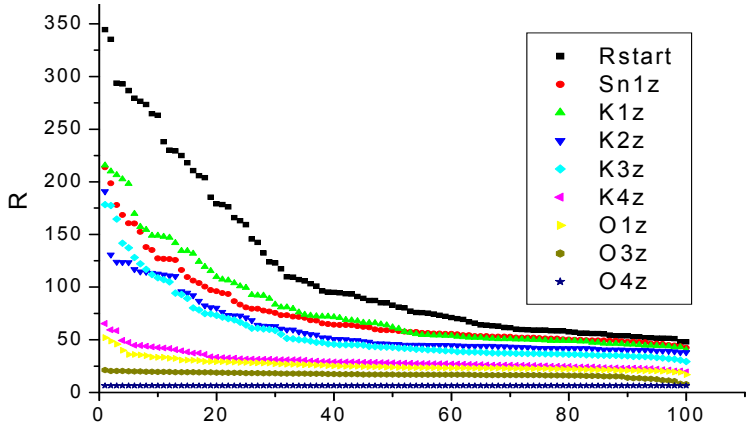

Рис. 3. График распределения R-фактора в популяции структур K_4SnO_4 при замене случайных атомных позиций (x,y,z) на истинные (Rstart – стартовая генерация, Sn1z – после замены позиции атома Sn, K1z – после замены позиции 1-го атома K и т.д.)

ты Z в популяции из 100 структурных моделей K_4SnO_4, упорядоченных в порядке убывания R-фактора.

Кривая R-фактора стартовой генерации показывает, что уже в популяции случайно генерированных структур имеется некоторое их число с относительно небольшим R-фактором. По мере расстановки атомов в истинные позиции кривые располагаются систематически все ниже и ниже. При этом R-фактор в правых частях распределений меняется незначительно и наиболее систематически. Заметим, что интенсивность дифракционных рефлексов и, следовательно, R-фактор зависят от межатомных расстояний, которые для атомов, расположенных вблизи их истинных позиций, близки к оптимальным. В то же время негативный вклад в R-фактор от межатомных расстояний для атомов в случайных позициях часто взаимно компенсируется, о чем свидетельствует форма распределения R-фактора в правой части рис. 1.

Рис. 3 показывает высокую чувствительность R-фактора к попаданию атомов в истинные позиции. Возьмем в качестве меры чувствительности долю положительных значений среднего по популяции изменения R-фактора для данной координаты (X, Y или Z):

$$N_X = n_X\{[\Delta R_i^{cp.} = \sum_{j=1}^{m} \Delta R_{ij} / m] > 0\} / n \qquad (6)$$

Таблица 6. Характеристики снижения R-фактора при последовательной замене случайных координат атомов на их истинные значения в популяциях различных структур

Среднее $\{N_X + N_Y\}$	Среднее $\{N_Z\}$	Среднее $\{\Delta R/R)_i^{ср.}\}$ для координат X и Y	Среднее $\{\Delta R/R)_i^{ср.}\}$ для координат Z
69,3 %	94,6 %	2,3 %	14,8 %

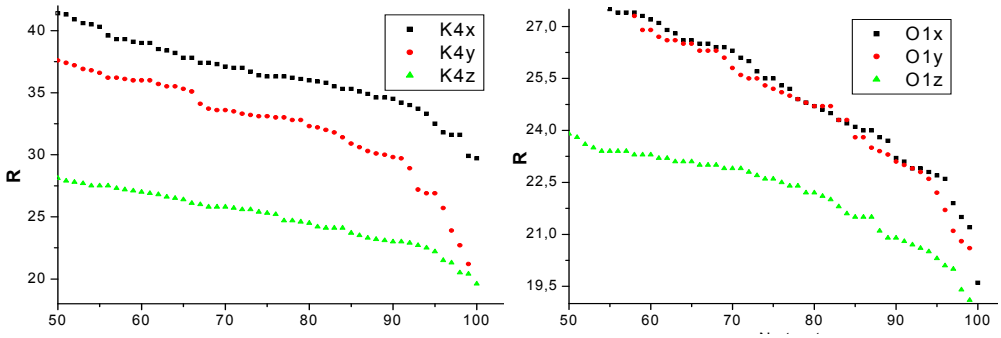

Рис. 4. Графики распределения R-фактора в популяции структур K_4SnO_4, слева для атома K4, справа – для O1; по оси абсцисс – порядковый номер структуры; в рамках – список последовательно заменяемых координат атомов

где ΔR_{ij} – изменение R-фактора при замене случайного значения координаты атома *i* на истинное в структуре *j* из *n* атомов в популяции из *m* структур. Отметим, что при отсутствии чувствительности R-фактора к координате, например к X, величина N_X~50 %, а величины $\Delta R_i^{ср.} \sim 0$, в противном случае N_X >>50 % и $\Delta R_i^{ср.}$>>0. Средние характеристики снижения R-фактора при замене случайных координат атомов на истинные значения для 10 независимо сгенерированных популяций структур соединений K4SnO4, As2O3 и $CaC_4H_4O_6{\cdot}4H_2O$ разной степени сложности сведены в табл. 6.

В качестве иллюстрации на рис. 4 представлены правые части типичных графиков снижения R-фактора при замене координат атомов на их истинные значения в упорядоченной популяции из 100 структурных моделей соединения K_4SnO_4, построенных для тяжелого и легкого атомов: 4-го калия (слева) и 1-го кислорода (справа). Верхние графики соответствуют замене отдельных координат X и Y, а нижние – полных атомных позиций (X, Y, Z).

На рис. 4 видим, что R-фактор чувствителен к попаданию в истинную позицию не только всего атома, но и отдельных его координат, причем не только тяжелых (калий), но и легких (кислород), хотя и в меньшей степени.

Обсуждение механизма глобальной сходимости ЭСМ

Таблица 1 показывает, что метод ЭСМ может обеспечивать глобальную сходимость небольших популяций (~100) структурных моделей за небольшое число поколений (~26), несмотря на огромное число возможных вариантов структуры. Например, для структуры Ba_2CrO_4 (21 атомная координата) оно составляет 2^{105} при кодировании по 5-битной сетке. В соответствии с (1)-(4) снижение R-фактора должно происходить за счет накопления атомов в их истинных по-

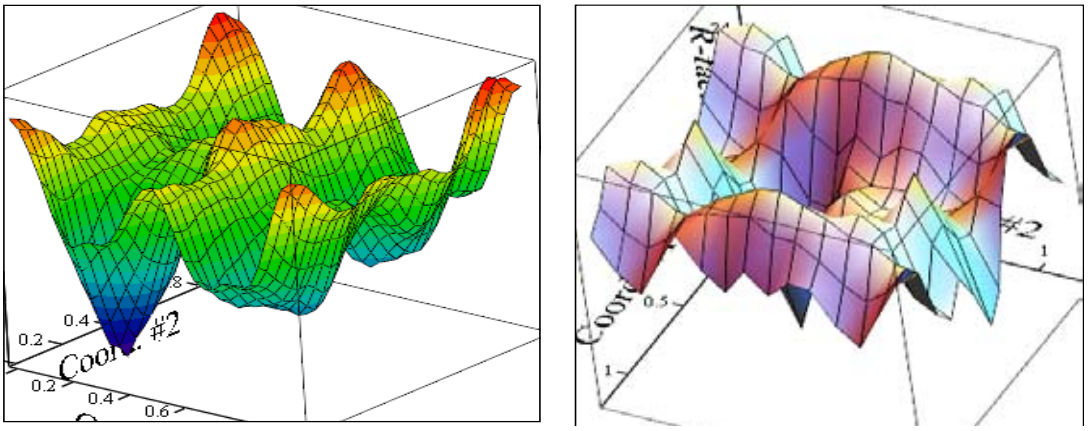

Рис. 5. Двумерные разрезы гиперповерхности R-фактора в плоскостях X0Y для структуры K_4SnO_4: слева – для атома олова, справа – для одного из кислородов

зициях. Вероятность случайного попадания данного атома в истинную позицию (x,y,z) составляет $p_1=1/2^{15}$, т.е. в среднем в одну структурную модель популяции за ~ 320 поколений. Очевидно, для накопления всех атомов в истинных позициях в одной из структур популяции процесс эволюции должен занимать в десятки раз большее число поколений, чем фактически в ЭСМ. Факт ускоренной сходимости и установление статистической чувствительности R-фактора к попаданию координат атомов в их истинные позиции в структуре позволяют предположить, что строительными блоками структуры в ЭСМ наряду с позициями атомов должны быть и отдельные атомные координаты. Очевидно, это связано со специфической формой гиперповерхности R-фактора. На рис. 5 представлены примеры двух поверхностей R-фактора структуры K_4SnO_4 как функции координат X и Y для тяжелого атома олова (слева) и легкого атома кислорода (остальные атомные координаты фиксированы в истинных позициях структуры). Можно отметить наличие широких пологих краев «воронки» вокруг глобального минимума. Замена у произвольной точки (X, Y) одной из координат на истинное значение при случайном значении другой координаты приводит к достаточно высокой вероятности попадания всей точки на край воронки с более низким значением R-фактора, которое может еще снизиться при уточнении этой структурной модели по МПР.

Поскольку размер сетки $1/2^4$ еще чувствителен (табл. 2), можно предположить, что величина R-фактора с высокой вероятностью p_2 значимо снижается после попадания одной из координат атома в истинное значение, если при этом случайные значения двух других координат будут находиться в окрестности истинной позиции атома с радиусом около $1/2^3$. Вероятность попадания координаты атома в истинное значение $1/2^5$ (в среднем в трех структурных моделях популяции размера 100), тогда вероятность снижения R-фактора составит $p_3=1/2^{11}p_2$. При p_2~0,5 значение $p_3=1/2^{12}$ в 8 раз выше p_1. Снижение R-фактора обеспечивает статистически преимущественное накопление таких структурных моделей в популяции. Кластеры истинных координат атомов, образуемые в некоторых структурных моделях за счет скрещиваний и удачных мутаций, обеспечивают дальнейшее снижение R-фактора и распространяются по популяции за счет новых скрещиваний и селекции. Локальная оптимизация лучших из этих структурных

моделей на 2-м уровне ЭМС обеспечивает точную установку координат (рис. 2), а попадания атомов в истинные позиции еще более снижают R-фактор (рис. 4). По мере накопления атомов в истинных позициях это приводит к сходимости одной из структурных моделей к истинной структуре.

Заключение

Стохастические методы имитации отжига и генетические алгоритмы обеспечивают ab initio поиск атомной кристаллической структуры химических соединений по порошковой дифрактограмме, хорошо автоматизируются и не требуют высокой квалификации исследователя в области структурного анализа. Это создает возможность выполнения структурных исследований непосредственно в тех же лабораториях и материаловедческих центрах, где новые химические соединения и создаваемые на их основе материалы синтезируют и исследуют, имея в виду, что многие из них оснащены относительно недорогими порошковыми дифрактометрами. Однако сложность кристаллических структур, определяемых этими методами, пока не превышает ~ 40 степеней свободы структурных параметров [4, 15]. Обнаружение и объяснение эффекта быстрой сходимости нового эволюционного метода ЭСМ открывает принципиальную возможность его развития для определения более сложных кристаллических структур. Одним из наиболее перспективных направлений является развитие на основе ЭСМ параллельной мультипопуляционной коэволюции структурных моделей на базе суперкомпьютерных вычислений.

Список литературы

1. Inorganic Crystal Structure Database. FIZ Karlsruhe // http://www.fiz-karlsruhe.de/icsd.html.

2. Cambridge Structural Database. Cambridge Crystallographic Data Centre // http://www.ccdc.cam.ac.uk/products/csd/

3. Young R.A. The Rietveld Method // Oxford University Press. 1995. 298 p.

4. David W. I. F., Shankland K. Structure determination from powder diffraction data // Acta Cryst. (2008). A64. P. 52–64.

5. Favre-Nicolin V. and Cerny R. FOX, `free objects for crystallography': a modular approach to ab initio structure determination from powder diffraction // J. Appl. Cryst. (2002). 35. P. 734–743.

6. Thomas A. N. Griffin, Kenneth Shankland, Jacco van de Streek and Jason Cole. GDASH: a grid-enabled program for structure solution from powder diffraction data // J. Appl. Cryst. (2009). 42. P. 356–359.

7. Goldberg D. E. Genetic Algorithms in Search, Optimization and Machine Learning. Addison. Wesley Publishing Co. Inc., 1989. ISBN 0-201-15767-5.

8. Solovyov L. A., Kirik S. D. Application of a Simulated Annealing Approach in Powder Crystal Structure Analysis. Materials Science Forum 133-136 (1993). P. 195-200.

9. Kariuki B. M., Serrano-Gonzalez H., Johnston R. L. and Harris K. D. M.. The application of a genetic algorithm for solving crystal structures from powder diffraction data. *Chem. Phys. Letters*, 280:189.195, 1997.

10. Kenneth D.M. Harris. Fundamentals and applications of genetic algorithms for structure solution from powder X-ray diffraction data // Computational Materials Science. V. 45. Issue 1. 2009. P. 16–20.

11. Cerny R. and Favre-Nicolin V. Direct space methods of structure determination from powder diffraction: principles, guidelines, perspectives // Z. Kristallogr. 222 (2007) P. 105-113.

12. Yakimov Y. I., Semenkin E. S., Yakimov I. S. Two-level genetic algorithm for a fullprofile fitting of X-ray powder patterns. // Z. Kristallogr. Suppl. 30 (2009) P. 21-26.

13. Powder Diffraction Theory and Practice, ed. R.E. Dinnebier and S.J.L. Billinge // Royal Society of Chemistry, 2008. 507 P.

14. Solovyov, L.A. Full-profile refinement by derivative difference minimization / L.A. Solovyov // J. Appl. Cryst. 2004. Vol. 37. P. 743-749.

15. Bail A. Le, Cranswick L. M. D. Third structure determination by powder diffractometry round robin (SDPDRR-3) // Powder Diffr. Volume 24, Issue 3, pp. 254-262 (September 2009).

The Evolutionary Method of Modeling a Crystal Structure by Powder Diffraction Data

Yaroslav I. Yakimov[a], Sergei D. Kirik[a], Evgeny S. Semenkin[a], Leonid A. Solovyov[b] and Igor S. Yakimov[a]

[a]Siberian Federal University
79 Svobodny, Krasnoyarsk, 660041 Russia
[b]Institute of Chemistry and Chemical Technology SB RAS,
42 K. Marx, Krasnoyarsk, 660049 Russia

The evolutionary method for modeling of crystal structure from powder diffraction integrated with full profile analysis method and genetic algorithm is proposed. An evolutionary mechanism of new hybrid method is studies on ab initio determination a number of known structures.

Keywords: X-ray powder diffraction, crystal structure analysis, genetics algorithm.