

УДК: 577.3

## **Использование связных масок в задаче восстановления изображения изолированной частицы по данным рентгеновского рассеяния. II. Зависимость точности решения от шага дискретизации экспериментальных данных**

**Лунина Н.Л.<sup>1</sup>, Петрова Т.Е.<sup>1</sup>, Уржумцев А.Г.<sup>\*2,3</sup>, Лунин В.Ю.<sup>\*\*1</sup>**

<sup>1</sup>*Институт математических проблем биологии, Российская академия наук,  
Пушино, Московская область, Россия*

<sup>2</sup>*Institut de Génétique et Biologie Moléculaire et Cellulaire, Илькири, Франция*

<sup>3</sup>*Université de Lorraine, Вандевр-ле-Нанси, Франция*

**Аннотация.** Совершенствование техники рентгеновского дифракционного эксперимента приводит в настоящее время к возможности регистрации рассеянных лучей не только для кристаллических образцов, но и для изолированных больших биологических частиц (вирусов, макромолекулярных комплексов и отдельных клеток). Эксперимент с изолированной частицей позволяет получить значения интенсивностей рассеянных лучей для непрерывного спектра векторов рассеяния. Такой эксперимент дает гораздо больше экспериментальной информации, нежели в случае рассеяния кристаллическим образцом, когда информация ограничена набором брэгговских рефлексов. При практическом исследовании исходная непрерывная картина дифракции дискретизируется – набор используемой далее экспериментальной информации ограничивается значениями, выбранными в узлах регулярной сетки в пространстве векторов рассеяния (в обратном кристаллографическом пространстве). Однако в данном случае сетка более не обусловлена экспериментальными ограничениями, такими как размер элементарной ячейки кристалла, а может быть выбрана по желанию исследователя. Шаг дискретизации определяет количество информации, вовлеченной в процесс решения фазовой проблемы, и трудоемкость необходимых вычислений. В данной работе исследуется влияние шага дискретизации на точность решения фазовой проблемы, получаемого при использовании предложенного ранее авторами метода, основанного на использовании связных бинарных масок исследуемого объекта. Показано, что ожидаемое повышение точности решения при уменьшении шага дискретизации продолжается и после пересечения предела Найквиста, определяемого как величина, обратная к удвоенному размеру изучаемой биологической частицы.

**Ключевые слова:** *рентгеновская кристаллография, фазовая проблема, XFEL, рассеяние изолированной частицей.*

---

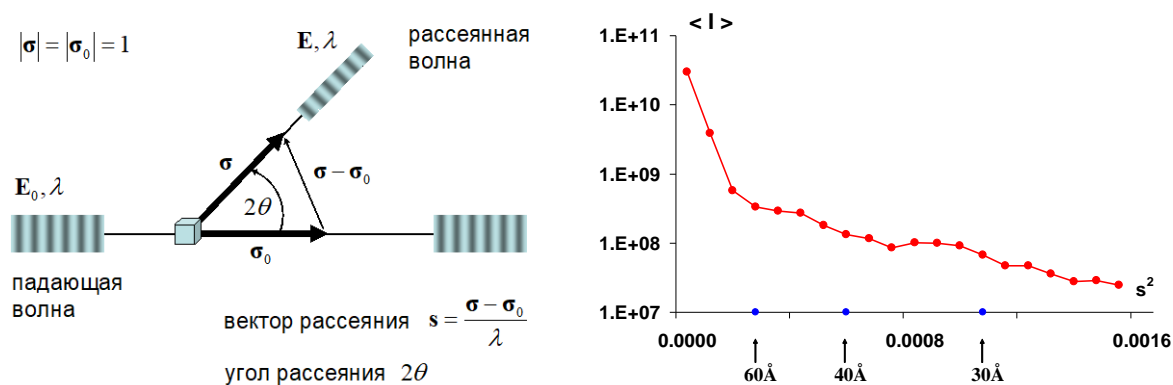
\*sacha@igbmc.fr

\*\*lunin@impb.psn.ru

## 1. ДИФРАКЦИЯ РЕНТГЕНОВСКИХ ЛУЧЕЙ НА КРИСТАЛЛАХ И ИЗОЛИРОВАННЫХ МОЛЕКУЛАХ

### 1.1. Постановка эксперимента

Целью первого этапа расшифровки структуры биологического объекта по данным рентгеновского рассеяния является нахождение функции, описывающей распределение электронов в изучаемом образце. Эта функция, найденная хотя бы приближенно, затем интерпретируется в терминах координат отдельных атомов, что приводит к построению приближенной атомной модели изучаемого объекта. На заключительном этапе расшифровки структуры параметры построенной модели уточняются посредством минимизации расхождения между данными эксперимента и теоретической картиной рассеяния, соответствующей построенной модели.



**Рис. 1.** Слева: схема рентгеновского дифракционного эксперимента. Справа: зависимость интенсивностей рефлексов от разрешения. Для тонких сферических оболочек в обратном пространстве показаны средние значения интенсивностей как функция квадрата радиуса оболочки  $s^2 = (2 \sin \theta / \lambda)^2$ . Данные приведены для тестового объекта PS-I в ячейке c1 (см. раздел 3).

Схема рентгеновского дифракционного эксперимента представлена на рисунке 1. В этом эксперименте исследуемый объект помещается на пути  $\sigma_0$  первичной рентгеновской волны с длиной волны  $\lambda$  и измеряется интенсивность вторичных волн, расходящихся от объекта в различных направлениях  $\sigma$ . Вторичные волны являются результатом сложения волн, исходящих от осциллирующих электронов объекта, возбужденных первичной волной. Комплексная амплитуда вторичной волны  $E$  отличается от амплитуды первичной волны  $E_0$  двумя множителями:

$$E = \varepsilon F(s) E_0. \quad (1)$$

Множитель  $\varepsilon$  не зависит от структуры объекта. Его величина определяется, в основном, долей потока энергии исходной волны, приходящей с рассеянной волной на детектор, при рассеянии одним электроном. Эта доля очень мала (она может быть оценена величиной  $\varepsilon \sim 10^{-12}$ ), что и создает основную сложность при регистрации рассеянного излучения. Комплексный множитель  $F(s)$  (структурный фактор) связан с распределением  $\rho(\mathbf{r})$  электронной плотности в рассеивающем объекте и условиями эксперимента (направлениями первичной и регистрируемой вторичной волны, длиной волны используемого рентгеновского излучения). При отсутствии эффекта аномального рассеяния эта связь выражается равенствами:

$$\mathbf{F}(\mathbf{s}) = F(\mathbf{s}) \exp[i\varphi(\mathbf{s})] = \int_{\mathbf{R}^3} \rho(\mathbf{r}) \exp[2\pi i(\mathbf{s}, \mathbf{r})] dV_r, \quad (2)$$

$$\mathbf{s} = \frac{\boldsymbol{\sigma} - \boldsymbol{\sigma}_0}{\lambda}. \quad (3)$$

Величины  $F(\mathbf{s})$  и  $\varphi(\mathbf{s})$  называются модулем и фазой структурного фактора, а вектор  $\mathbf{s}$  – вектором рассеяния.

Регистрируемой в эксперименте информацией является интенсивность вторичной волны, которая пропорциональна квадрату модуля структурного фактора

$$I(\mathbf{s}) = CF^2(\mathbf{s}), \quad (4)$$

где  $C$  – единый для всех структурных факторов коэффициент, определяемый условиями эксперимента. Поэтому можно сказать, что эксперимент позволяет определить значения модулей структурных факторов. В то же время, значение фаз структурного фактора не может быть измерено в рамках стандартного рентгеновского эксперимента, и восстановление значений фаз, необходимых для восстановления распределения электронной плотности, составляет центральную проблему рентгеноструктурного анализа, так называемую фазовую проблему.

## 1.2. Рассеяние кристаллическими образцами

Использование в эксперименте кристаллических образцов оказывает двойное влияние на картину рассеяния. С одной стороны, для дискретного набора выделенных направлений все волны, приходящие от разных копий объекта, оказываются в фазе, что приводит к многократному усилению вторичной волны (возникновению брэгговских рефлексов) и позволяет экспериментально измерить ее интенсивность. Для остальных значений вектора рассеяния волны от разных копий объекта приходят с хаотичными сдвигами фаз, что приводит к их взаимной аннигиляции, крайне слабой интенсивности для этих векторов и практической невозможности ее измерения.

Для кристаллического образца распределение электронной плотности является периодической функцией и может быть представлено в виде ряда Фурье:

$$\rho(\mathbf{r}) = \frac{1}{|\mathbf{V}_{cell}|} \sum_{\mathbf{s} \in \mathcal{R}'} \mathbf{F}(\mathbf{s}) \exp[-2\pi i(\mathbf{s}, \mathbf{r})]. \quad (5)$$

Суммирование осуществляется по всем узлам трехмерной целочисленной решетки (решетки обратного пространства)  $\mathcal{R}' = \{h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*\}$ , где  $\{h, k, l\}$  – всевозможные целочисленные наборы. Данная решетка построена на базисе  $\{\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*\}$ , являющимся сопряженным к базису  $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ , образованному линейно независимыми периодами функции  $\rho(\mathbf{r})$ . Параллелепипед  $\mathbf{V}_{cell}$ , построенный на векторах  $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ , называется элементарной ячейкой кристалла. Базис элементарной ячейки, как правило, выбирается из периодов функции  $\rho(\mathbf{r})$  таким образом, чтобы обеспечить минимально возможный объем  $|\mathbf{V}_{cell}|$  элементарной ячейки.

Условия возникновения брэгговских рефлексов (условия Лауэ – Брэгга – Вульфа) могут быть сформулированы, как условие принадлежности вектора рассеяния  $\mathbf{s}$  решетке обратного пространства:  $\mathbf{s} \in \mathcal{R}'$ . Структурные факторы таких рефлексов связаны с функцией распределения электронной плотности соотношением

$$\mathbf{F}(\mathbf{s}) = F(\mathbf{s}) \exp[i\varphi(\mathbf{s})] = N \int_{\mathbf{V}_{cell}} \rho(\mathbf{r}) \exp[2\pi i(\mathbf{s}, \mathbf{r})] dV_r, \quad (6)$$

где  $N$  – общий для всех брэгговских рефлексов множитель, пропорциональный числу копий исследуемого объекта в кристалле. Задача восстановления распределения  $\rho(\mathbf{r})$  может быть сформулирована как задача нахождения значений модулей и фаз структурных факторов, необходимых для расчета распределения по формуле (5).

Таким образом, при использовании кристаллических образцов, рентгеновский эксперимент позволяет получить лишь половину информации, необходимой для восстановления распределения электронной плотности по формуле (5), а именно, модулей структурных факторов для векторов рассеяния  $\mathbf{s} \in \mathcal{R}'$ . Восстановление недостающей половины информации (соответствующих фаз) требует дополнительных усилий, обычно связанных либо с проведением дополнительных экспериментов в модифицированных условиях, либо с использованием дополнительной информации об объекте.

Необходимость использования в эксперименте образцов исследуемого вещества, приготовленных в виде монокристалла, вызвана тем, что рассеянные объектом вторичные рентгеновские волны являются слишком слабыми. До последнего времени экспериментальное оборудование не позволяло регистрировать рассеяние изолированными биологическими молекулами. Однако, ввод в эксплуатацию новых мощных источников излучения – рентгеновских лазеров на свободных электронах (XFEL) – и постоянный прогресс в совершенствовании регистрирующих устройств позволяют надеяться на преодоление этой сложности. Первые экспериментально полученные данные по рассеянию отдельными биологическими частицами – вирусами и цельными клетками [1–6] – хотя и имеют достаточно низкое разрешение и пока ограничены отдельными двумерными рентгенограммами, дают основания надеяться на получение в будущем трехмерных наборов дифракционных данных в таких экспериментах. Следует отметить, что речь идет именно о трехмерных наборах данных, регистрируемых для отдельных частиц, а не об одномерных кривых рассеяния, получаемых при использовании метода малоуглового рентгеновского рассеяния [7], когда регистрируемые данные являются результатом усреднения по большому числу разным образом ориентированных частиц.

Еще одной проблемой, осложняющей расчет распределения электронной плотности, является то, что формула (5) подразумевает суммирование бесконечного ряда, хотя эксперимент позволяет измерить лишь конечный набор модулей структурных факторов. Размер полученного экспериментально набора зависит, в первую очередь, от качества используемого в эксперименте кристалла. Этот размер обычно характеризуется величиной разрешения  $d_{\min}$ , вычисляемого как  $d_{\min} = 1/s_{\max}$ , где  $s_{\max}$  – максимальная длина вектора рассеяния  $\mathbf{s}$  для членов ряда (5), включенных в расчет. Геометрически, величина  $d_{\min}$  является минимальной длиной периода для гармоник Фурье  $\exp[2\pi i(\mathbf{s}, \mathbf{r})]$ , включенных в расчет суммы (5), и характеризует минимальный размер деталей, визуально различимых на контурных картах, соответствующих этой сумме ряда. Частичные суммы для ряда Фурье (5) в кристаллографии принято называть синтезами Фурье электронной плотности. Теоретический предел разрешения определяется величиной  $\lambda/2$ , где  $\lambda$  – длина волны используемого в эксперименте рентгеновского излучения, что дает предельную величину  $d_{\min} \approx 0.5 \text{ \AA}$  при длине волны около 1 ангстрема, обычно используемой при проведении эксперимента на синхротронном ускорителе. Однако, для большинства структур, депонированных в банк белковых структур PDB [8], эксперимент позволил собрать набор данных лишь с разрешением около 2  $\text{\AA}$ . Это обстоятельство вызвано сложностями в получении кристаллов высокого качества для биологических макромолекул.

Помимо характеристики всего набора структурных факторов, использованных при расчете синтеза Фурье, понятие "разрешение", определяемое как  $d = 1/s$ , применяется и к отдельным структурным факторам. Отметим, что длина вектора рассеяния  $\mathbf{s}$  связана с углом рассеяния соотношением  $s = 2 \sin \theta / \lambda$ . Поэтому структурным факторам высокого разрешения (малые значения  $d$ , большие значения  $s$ ) отвечает рассеяние под большими углами, а, имея в виду данные низкого разрешения (большие значения  $d$ , малые значения  $s$ ), говорят о "малоугловом" рассеянии. При работе с большими биологическими объектами распространенной является практика поэтапного повышения разрешения используемого набора данных. В данной работе мы ограничиваемся обсуждением первого этапа исследования - получения изображения макромолекулярного объекта при разрешении порядка  $25 \text{ \AA}$ , позволяющем определить композицию макромолекулярного комплекса и общие очертания доменов или молекул, его образующих.

### 1.3. Рассеяние изолированными частицами

Рентгеновский дифракционный эксперимент с изолированной частицей теоретически позволяет измерить значения модуля интегрального преобразования Фурье функции распределения электронной плотности  $\rho_{sp}(\mathbf{r})$  в отдельной частице

$$F_{sp}(\mathbf{s}) = \int_{\mathbf{R}^3} \rho_{sp}(\mathbf{r}) \exp[2\pi i(\mathbf{s}, \mathbf{r})] dV_{\mathbf{r}} \quad (7)$$

для любых значений вектора рассеяния  $\mathbf{s}$  в пределах разрешения  $1/s \leq \lambda/2$ , а не только для дискретного набора  $\mathcal{R}'$  брегговских векторов, как это имеет место при использовании кристаллических образцов. Таким образом, эксперимент с изолированной частицей может дать существенно больший объем информации, нежели эксперимент с кристаллическим образцом, и тем самым существенно облегчить решение фазовой проблемы.

Конечные размеры изучаемой частицы позволяют свести задачу восстановления соответствующей ей функции распределения электронной плотности к стандартной кристаллографической задаче при помощи следующего приема [9]. Рассмотрим воображаемый кристалл с достаточно большой элементарной ячейкой  $\mathbf{V}_{abc}$ , содержащей внутри себя одну изучаемую частицу. Обозначим через  $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$  базис этой ячейки и через  $\{\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*\}$  сопряженный базис (базис обратного пространства). Введем в рассмотрение периодическую функцию  $\rho_{crist}(\mathbf{r})$ , совпадающую с  $\rho_{sp}(\mathbf{r})$  внутри кристалла и периодически продолженную на все пространство периодами  $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ . Распределение электронной плотности  $\rho_{crist}(\mathbf{r})$  полностью определяется набором структурных факторов с векторами  $\mathbf{s}$ , принадлежащими целочисленной решетке  $\mathcal{R}'_{abc}$ , построенной на векторах сопряженного базиса  $\{\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*\}$ . Величины соответствующих структурных факторов:

$$\mathbf{F}_{crist}(\mathbf{s}) = \int_{\mathbf{V}_{abc}} \rho_{crist}(\mathbf{r}) \exp[2\pi i(\mathbf{s}, \mathbf{r})] dV_{\mathbf{r}} = \int_{\mathbf{R}^3} \rho_{sp}(\mathbf{r}) \exp[2\pi i(\mathbf{s}, \mathbf{r})] dV_{\mathbf{r}} = \mathbf{F}_{sp}(\mathbf{s}) \quad (8)$$

совпадают со значениями интегрального преобразования Фурье распределения электронной плотности в отдельной частице. Поэтому, ограничив набор имеющейся экспериментальной информации значениями модуля преобразования Фурье на дискретной сетке  $\mathcal{R}'_{abc}$ :  $\{F_{sp}(\mathbf{s}); \mathbf{s} \in \mathcal{R}'_{abc}\}$ , мы приходим к задаче восстановления

распределения плотности  $\rho_{cryst}(\mathbf{r})$  по модулям его структурных факторов. Формально говоря, мы здесь сталкиваемся с той же проблемой, что и при работе с кристаллами: эксперимент дает только половину информации, необходимой для расчета распределения плотности  $\rho_{cryst}(\mathbf{r})$  по формуле (5). Однако существенное отличие возникающей задачи для воображаемого кристалла от обычной кристаллографической практики состоит в том, что, выбирая достаточно большую ячейку, мы получаем задачу восстановления функции, для которой известны ее значения, равные нулю, для большей части элементарной ячейки. Правда, при этом остается неизвестным, в каких именно точках ячейки эта функция равна нулю. В кристаллографических терминах этот факт может быть сформулирован как наличие в элементарной ячейке большой области, занятой растворителем. Такого рода информация приводит к избыточности количества экспериментальных измерений по отношению к количеству значений, которые должны быть определены [10], и широко используется в кристаллографической практике. Естественно, выбор элементарной ячейки воображаемого кристалла неоднозначен и, меняя ее параметры, мы изменяем количество экспериментальной информации, вовлекаемой в работу в рамках заданной зоны разрешения.

Избыточность экспериментальной информации, вносимая присутствием большого пустого объема в элементарной ячейке, на протяжении многих лет используется при решении фазовой проблемы в биологической кристаллографии и оптике. При этом основным методом использования этой информации являются итерационные процедуры типа solvent flattening, density modification или hybrid input-output algorithm [11–13]. Недавно нами был предложен альтернативный подход к использованию такой избыточности, осуществляемый в рамках процедур случайного поиска с использованием добавочных ограничений, повышающих избыточность экспериментальных данных, таких как связность и бинарность области частицы [14, 15].

Увеличение размеров элементарной ячейки воображаемого кристалла позволяет теоретически неограниченно наращивать количество используемой для восстановления фаз экспериментальной информации. Однако на практике мы сталкиваемся с некоторыми ограничениями. Во-первых, минимальный размер шага решетки  $\mathcal{R}'$  ограничен техническими характеристиками детектора (размером пикселя матрицы детектора) и условиями съемки (расстоянием образец-детектор). Во-вторых, увеличение числа структурных факторов, вовлеченных в работу, приводит к существенному повышению трудоемкости расчетов и предъявляет повышенные требования к производительности имеющихся в распоряжении вычислительных средств. Наконец, потенциальная избыточность данных может приводить к успеху при работе в ограниченной по разрешению области и не распространяться на работу при более высоком разрешении. Целью данной работы являлось изучение зависимости качества найденных решений фазовой проблемы от размеров ячейки воображаемого кристалла, использованной в расчетах.

## 2. ИСПОЛЬЗОВАНИЕ СВЯЗНЫХ МАСОК ПРИ РЕШЕНИИ ФАЗОВОЙ ПРОБЛЕМЫ БИОЛОГИЧЕСКОЙ КРИСТАЛЛОГРАФИИ

Метод, который используется нами для решения фазовой проблемы, описан в работах [14, 15]. Вкратце он может быть изложен следующим образом. В элементарной ячейке воображаемого кристалла вводится сетка и, исходя из оценки объема исследуемой частицы, определяется число точек сетки, попадающих внутрь объекта. Первый этап работы заключается в случайной генерации большого числа конечных связных наборов узлов сетки (масок) заданного размера. Связность набора определяется на основе введенных правил соседства; в наших тестах соседними

считались 6 ближайших узлов сетки. Каждому сгенерированному набору узлов ставится в соответствие бинарная характеристическая функция множества, и для этой функции рассчитываются модули и фазы структурных факторов. В случае, когда рассчитанные модули структурных факторов достаточно близки к экспериментально измеренным значениям, сопутствующий набор фаз структурных факторов считается допустимым и запоминается для дальнейшего анализа. Генерация продолжается, пока не получено заданное количество допустимых наборов фаз (в наших тестах использовалось 100 отобранных наборов). На следующем этапе работы отобранные наборы фаз выравниваются путем применения сдвигов начала координат и выбора энантиомера, исходя из требования наибольшей схожести соответствующих синтезов Фурье. Выровненные наборы фаз в паре с экспериментально определенными модулями структурных факторов усредняются. Полученный усреднением набор фаз рассматривается как результат первого этапа решения фазовой проблемы [14].

В качестве критерия близости рассчитанных по маске значений модулей структурных факторов  $\{F_{calc}(\mathbf{s})\}$  и экспериментально определенных значений  $\{F_{exact}(\mathbf{s})\}$  в работе использовался нецентрированный коэффициент корреляции:

$$CM[d_{max}, d_{min}] = \frac{\sum_{\mathbf{s} \in \mathbf{S}} F_{calc}(\mathbf{s}) F_{exact}(\mathbf{s})}{\sqrt{\sum_{\mathbf{s} \in \mathbf{S}} F_{calc}^2(\mathbf{s}) \sum_{\mathbf{s} \in \mathbf{S}} F_{exact}^2(\mathbf{s})}}. \quad (9)$$

Здесь суммирование идет по структурным факторам  $\mathbf{s} \in \mathbf{S} = \{\mathbf{s} : 1/d_{max} < s \leq 1/d_{min}\}$ .

На первом этапе каждая маска строится шаг за шагом, начиная с одной случайно выбранной точки. На каждом шаге к маске добавляется одна точка, выбираемая с равной вероятностью из граничных точек для уже построенной маски.

Второй этап (уточнение набора фаз) отличается от первого двумя особенностями [15]. Во-первых, как и на первом этапе, маска строится шаг за шагом, начиная с одной точки, однако выбор добавляемой граничной точки осуществляется с учетом некоторого априорного распределения вероятностей  $P_{prior}(\mathbf{r})$ . В наших тестах априорное распределение вероятностей строилось на основе синтеза Фурье электронной плотности  $\rho(\mathbf{r})$ , построенного с использованием экспериментальных значений модулей структурных факторов и значений фаз, определенных на предыдущем этапе. Априорное распределение строилось в виде:

$$P(\mathbf{r}) = C \exp\left[\frac{\ln t}{\rho_{max} - \rho_{min}} \rho(\mathbf{r})\right], \quad (10)$$

где  $C$  – нормализующий множитель,  $\rho_{max}$  и  $\rho_{min}$  – максимальное и минимальное значение функции  $\rho(\mathbf{r})$ , а  $t$  – управляющий параметр, контраст. Нетрудно видеть, что параметр  $t$  есть не что иное, как отношение максимального и минимального значений функции  $P_{prior}(\mathbf{r})$ .

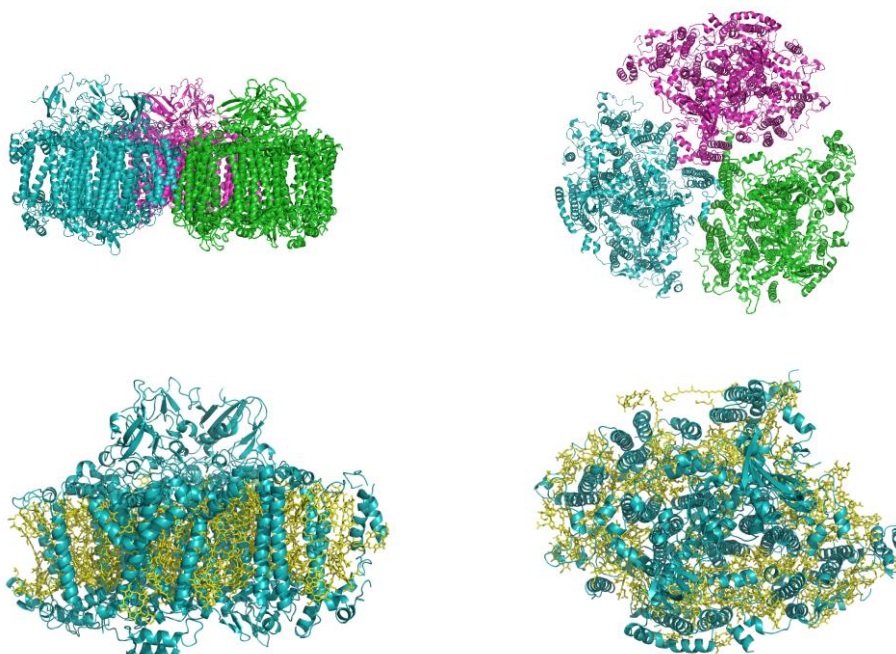
Во-вторых, на шаге уточнения, применяемом в нашем методе, одновременно используются несколько критериев, контролирующих отбор допустимых масок. Особенностью макромолекулярных комплексов является резкое уменьшение величины модуля структурных факторов с повышением сопутствующего им разрешения (рис. 1). Поэтому значение коэффициента корреляции модулей структурных факторов в большей мере определяется соответствием значений модулей для малоугловых рефлексов и слабо зависит от рефлексов высокоугловой зоны. Чтобы преодолеть это обстоятельство, при отборе масок контролировались не только коэффициент



корреляции по все рабочей зоне  $25 \text{ \AA}$ , но и коэффициенты корреляции, рассчитанные по сферическим оболочкам ограниченного разрешения ( $60\text{--}25$ ,  $40\text{--}25$ ,  $30\text{--}25 \text{ \AA}$ ), что приводило к последовательному исключению более сильных рефлексов низкого разрешения.

### 3. ТЕСТОВЫЙ ОБЪЕКТ

Для проверки возможностей предложенного подхода в данной работе в качестве тестового объекта использовалась известная структура тримера фотосистемы I (*cyanobacterial photosystem I*, [16], PDB код 1JBO [8]), далее PS-I. Этот тример содержит 36 белковых цепей и 381 кофактор, состоит из около 72 тысяч неводородных атомов. Тример имеет молекулярный вес 1068000 дальтон. Общая структура тримера показана на рисунке 2. Внешние размеры тримера составляют примерно  $200 \text{ \AA}$  по осям  $X$  и  $Y$  и  $100 \text{ \AA}$  по оси  $Z$ .



**Рис. 2.** Белковая часть тримера фотосистемы I (верхняя строка; разные цвета соответствуют разным мономерам) и структура отдельного мономера (нижняя строка, белковая часть и кофакторы).

Целью нашего исследования являлось изучение влияния размеров ячейки воображаемого кристалла на возможность решения фазовой проблемы и точность определения фаз. Как уже отмечалось, при наличии данных по рассеянию изолированной частицей исследователь имеет определенную свободу в выборе ячейки воображаемого кристалла. Однако выбор слишком больших параметров ячейки может привести к чрезмерному повышению трудоемкости вычислений. Поэтому речь идет о выборе компромисса между увеличением трудоемкости расчетов и достигаемым улучшением результатов. Ниже обсуждаются результаты *ab initio* определения фаз для трех вариантов размера элементарной ячейки (табл. 1). Эти тесты были проведены для кубической ячейки, предполагающей, что нам ничего не известно о форме объекта. Размер шага выборки значений в обратном пространстве по осям  $X$  и  $Y$  был больше предела Найквиста (определяемого как величина, обратная к удвоенному размеру объекта)  $1/400 \text{ \AA}^{-1}$  для ячейки  $c1$ , совпадал с ним для ячейки  $c3$  и был в полтора раза меньше для ячейки  $c4$ . Параметры ячеек и число структурных факторов в разных зонах разрешения приведены в таблице 1.



При работе с каждым из вариантов выбора ячейки, по атомной модели тримера были рассчитаны значения модулей и фаз структурных факторов, соответствующих данной ячейке в зоне разрешения 25 Å. Рассчитанные значения модулей далее рассматривались как экспериментально определенные величины. Значения фаз, рассчитанных по модели, никак не участвовали в решении фазовой проблемы и использовались лишь для контроля. В качестве критерия точности определения фаз использовались коэффициенты корреляции синтезов, рассчитанные по разным зонам разрешения

$$CP[d_{\max}, d_{\min}](\{\varphi(\mathbf{s})\}) = \frac{\sum_{\mathbf{s} \in \mathbf{S}} F_{\text{exact}}^2(\mathbf{s}) \cos(\varphi_{\text{exact}}(\mathbf{s}) - \tilde{\varphi}(\mathbf{s}))}{\sum_{\mathbf{s} \in \mathbf{S}} F_{\text{exact}}^2(\mathbf{s})}. \quad (11)$$

Здесь суммирование идет по структурным факторам с  $\mathbf{s} \in \mathbf{S} = \{\mathbf{s} : 1/d_{\max} < s \leq 1/d_{\min}\}$ ,  $F_{\text{exact}}(\mathbf{s}), \varphi_{\text{exact}}(\mathbf{s})$  означают точные значения модулей и фаз структурных факторов, рассчитанные по известной модели тримера,  $\{\tilde{\varphi}(\mathbf{s})\}$  – набор фаз, полученный из набора  $\{\varphi(\mathbf{s})\}$  его оптимальным выравниванием по отношению к фазам  $\{\varphi_{\text{exact}}(\mathbf{s})\}$  посредством сдвига начала координат и потенциальной смены энантиомера [14, 15].

**Таблица 1.** Размеры использованных в работе ячеек и распределение рефлексов по разрешению

	Ячейка		
	c1	c3	c4
параметры ячейки [Å]	250 × 250 × 250	400 × 400 × 400	600 × 600 × 600
содержание растворителя (%) <sup>1</sup>	91.59	97.95	99.39
размер сетки для генерации масок	30 × 30 × 30	48 × 48 × 48	72 × 72 × 72
число рефлексов по зонам разрешения ∞–25 Å	2071	8538	28877
∞–30 Å	1234	4921	16700
∞–40 Å	510	2082	7061
∞–60 Å	152	618	2084
60–40 Å	358	1464	4977
40–30 Å	724	2839	9639
30–25 Å	837	3617	12177
избыточность <sup>2</sup>	0.23	0.96	3.25

<sup>1</sup>содержание растворителя определялось как  $(1 - 1.23 \cdot Mw/V_{\text{cell}}) \cdot 100\%$ , где  $Mw$  – молекулярный вес тримера, а  $V_{\text{cell}}$  – объем элементарной ячейки.

<sup>2</sup>избыточность определялась как  $N_{\text{ref}}/(3 \cdot Np)$ , где  $N_{\text{ref}}$  – число рефлексов в зоне разрешения ∞–25 Å,  $Np$  – количество узлов сетки в области удельного объема 1.6 Å<sup>3</sup>/Da.

Тестовый объект является тримером и имеет ось симметрии третьего порядка. При проведении наших тестов эта симметрия никак не учитывалась. Разумеется, учет этой симметрии может облегчить решение фазовой проблемы и привести к более точным значениям фаз за счет усреднения по некристаллографической симметрии [17]. Однако, поскольку целью тестов являлась проверка работоспособности методики, основанной на использовании связанных масок, мы не использовали дополнительные приемы уточнения фаз, дабы исключить их влияние на результат. Наоборот, спонтанное проявление симметрии третьего порядка на финальных картах являлось дополнительным подтверждением правильности решения фазовой проблемы.

При генерации масок шаг сетки в элементарной ячейке брался равным  $d_{\min}/3$ , а при выравнивании фазовых наборов равным  $d_{\min}/4$ , т.е. в нашем случае равным

соответственно 8.3 и 6.25 Å. Значение контраста  $t$  в распределении (10) бралось равным  $10^6$ . На каждом шаге построение случайных масок велось до тех пор, пока не набиралось 100 масок, удовлетворяющих поставленным ограничениям. Для усреднения фазовые наборы выравнивались по зоне разрешения  $\infty$ –25 Å.

#### 4. РЕЗУЛЬТАТЫ ТЕСТОВ

Одним из основных параметров, определяющих описываемый метод, является размер генерируемых масок или, что то же самое, предполагаемый объем частицы. Эта характеристика может выражаться в разных единицах. В данной работе мы характеризуем предполагаемый размер области либо количеством точек в маске, либо удельным объемом, определяемым как отношение объема области (в кубических ангстремах) к молекулярной массе частицы (в дальтонах). В кристаллографии, при оценке процентного содержания растворителя в элементарной ячейке, принято считать удельный объем области молекулы равным  $1.23 \text{ \AA}^3/\text{Da}$  [18, 19], однако, эта оценка не является оптимальной при работе в зоне низкого разрешения, когда границы молекулы существенно сглажены. В связи с этим, на первом этапе (при определении начальных значений фаз) нами были протестированы различные значения размеров маски.

Другим существенным параметром является жесткость отбора, т.е. порог корреляции, определяющий отбор масок. В наших тестах этот порог задавался либо непосредственным значением  $CM_{crit}$ , либо в относительных единицах ( $z$ -score),

$$z_{crit} = \frac{CM_{crit} - \langle CM \rangle}{\sigma_{CM}}, \quad (12)$$

где  $\langle CM \rangle$  и  $\sigma_{CM}$  есть соответственно среднее значение и среднеквадратичное отклонение величины критерия  $CM$ , рассчитанное по большому числу сгенерированных вариантов. В таблице 2 представлены результаты первоначального определения значений фаз при разных комбинациях задаваемых размеров маски и жесткости отбора.

**Таблица 2.** Результаты усреднения наборов фаз, рассчитанных по случайно сгенерированным связным маскам, приводящим к заданному уровню корреляции модулей структурных факторов. Приведены значения фазовой корреляции  $CP$ , рассчитанной по зонам разрешения  $\infty$ –60,  $\infty$ –40,  $\infty$ –30,  $\infty$ –25 Å (верхняя строка) и 60–40, 40–30, 30–25 Å (нижняя строка). Порог отбора  $z_{crit}$  определялся по зоне разрешения  $\infty$ –25 Å. Работа велась в ячейке c1

$CP^*100$	Объем маски:удельный [ $\text{\AA}^3/\text{Da}$ ]/число точек						
	1.0. 1845	1.2 2214	1.4 2583	1.6 2952	1.8 3321	2.0 3691	
$z_{crit}$		66/61/59/57 18/24/04	66/60/57/56 –01/28/18	69/64/59/57 15/04/00	73/67/64/62 16/26/–02	75/69/63/62 17/–10/00	77/72/69/67 28/21/14
	1.5	72/68/64/62 21/15/13	73/66/62/61 –02/09/11	82/74/68/67 –06/–05/12	81/75/72/70 22/22/11	81/75/71/68 22/13/–09	81/75/70/68 21/05/05
	2.0	75/68/65/63 11/12/14	81/73/68/66 03/–02/–04	83/75/70/68 –02/05/10	79/72/68/66 06/16/11	81/75/71/69 18/18/09	80/75/71/69 22/22/04
	2.5	76/69/66/64 08/21/14	83/77/73/71 20/18/01	85/76/71/69 –10/05/09	87/82/75/73 33/–10/–09	83/77/73/71 24/17/04	83/75/71/69 –01/19/10
	3.0	79/73/70/68 13/25/13	88/81/76/74 16/08/04	83/76/70/68 16/–05/05	88/83/78/76 30/27/08	85/79/75/73 16/25/02	82/76/71/69 19/08/01

Первая строка таблицы отвечает случаю отсутствия отбора, т.е. в этом случае производится выравнивание и усреднение наборов фаз по случайно сгенерированным

маскам. Как было показано ранее [14, 15], даже в этом случае получаемые значения фаз ближе к истине, нежели случайно генерируемые фазовые наборы. Тестирование различных размеров маски показало, что оптимальный размер маски лежит в области удельных объемов  $1.4\text{--}1.8 \text{ \AA}^3/\text{Da}$ , что совпадает с результатами, полученными для других объектов [14, 15]. Для второго этапа работы был зафиксирован удельный объем частицы, равный  $1.6 \text{ \AA}^3/\text{Da}$ .

**Таблица 3.** Результаты уточнения наборов фаз. Приведены значения фазовой корреляции  $CP$ , рассчитанной по зонам разрешения  $\infty\text{--}60$ ,  $\infty\text{--}40$ ,  $\infty\text{--}30$ ,  $\infty\text{--}25$  (верхняя строка) и  $60\text{--}40$ ,  $40\text{--}30$ ,  $30\text{--}25 \text{ \AA}$  (нижняя строка). Работа велась в ячейке  $c1$

$CP*100/\text{число}$ построенных масок	Стартовый набор фаз для построения априорного распределения (10)				
	$r1$	$r2$	$r3$	$r4$	$r5$
Start	87/82/75/73 33/–10/–09 14087	83/77/71/69 16/01/–14 18469	84/78/73/71 16/08/02 22039	87/80/75/73 17/11/–03 21979	83/76/71/69 07/10/05 19074
Step 1	95/90/83/80 46/–12/–13 513	89/83/78/75 20/18/–05 633	90/85/78/76 32/00/–03 555	92/87/82/80 36/28/15 625	90/83/76/74 12/–03/02 544
Step 2	94/90/84/82 47/12/01 105	91/84/80/77 24/22/02 103	91/86/80/77 35/08/–05 100	94/89/85/83 44/32/15 111	92/85/78/76 16/00/05 104
Step 3	96/92/85/82 50/–05/–01 114	91/85/81/78 26/24/07 121	91/86/81/78 37/14/–04 110	94/90/86/84 47/36/17 105	92/85/79/77 20/02/05 162
Step 4	95/90/85/83 49/18/06 181	91/85/81/79 27/26/08 116	92/87/82/79 39/19/–02 135	96/92/87/85 48/24/09 111	93/86/80/78 22/04/05 245
Step 5	95/91/85/83 50/21/09 159	92/86/81/79 28/26/09 209	92/87/82/80 40/22/01 137	95/91/87/85 51/38/18 224	93/87/81/78 24/06/05 678
Step 6	95/91/86/84 50/23/11 332	92/86/81/79 28/26/10 301	92/87/83/80 41/24/04 226	97/93/88/86 53/28/12 119	93/87/81/79 26/07/06 1434
Step 7	95/91/86/84 51/24/12 608	92/86/81/79 29/25/11 333	92/88/83/81 41/25/06 289	97/93/88/86 54/29/12 317	94/87/82/79 27/09/06 1242
Step 8	95/91/86/84 52/25/14 1 241	92/86/82/79 29/24/11 582	93/88/83/81 43/25/08 362	96/92/88/86 55/41/20 276	94/87/82/80 28/11/06 1104
Step 9	95/91/86/84 53/25/15 2 155	92/86/82/79 29/23/11 819	93/88/83/81 44/24/09 570	97/93/89/87 56/33/15 664	94/88/82/80 28/13/06 1112
Step 10	95/91/86/84 56/25/14 1 843	92/86/82/79 29/21/11 1 269	93/88/83/81 44/24/10 774	96/93/89/87 57/44/22 1220	94/88/82/80 29/14/06 2224

Результаты применения предложенной процедуры определения фаз зависят от работы датчика случайных чисел и различаются при разной инициализации этого датчика. Для продолжения работы (уточнения значений фаз) были сгенерированы 5 независимых решений (наборов фаз)  $r1\text{--}r5$ , отвечающих удельному объему частицы  $1.6 \text{ \AA}^3/\text{Da}$  и жесткости отбора  $z_{crit} = 2.5$  для критерия  $CM[\infty\text{--}25]$ . Первая строка таблицы 3 показывает качество полученных наборов фаз. Далее для каждого из пяти решений было независимо проведено 10 циклов уточнения значений фаз, согласно процедуре, описанной в разделе 2.

При проведении процедуры уточнения значения фаз, полученных на предыдущем цикле, использовались для построения синтеза Фурье и априорного распределения (10), при этом использовалось значение контраста  $t = 10^6$ . Отбор допустимых масок велся с использованием 4 критериев отбора  $CM[\infty, 25]$ ,  $CM[60, 25]$ ,  $CM[40, 25]$ ,  $CM[30, 25]$ , отвечающих постепенному исключению их расчета сильных низкоугловых рефлексов. Пороговые значения критериев устанавливались по результатам предыдущего шага, как средние значения соответствующих критериев среди вариантов, отобранных на предыдущем шаге. Исключением являлся первый шаг уточнения, когда пороги определялись как средние значения этих критериев по результатам вспомогательной генерации масок без отбора.

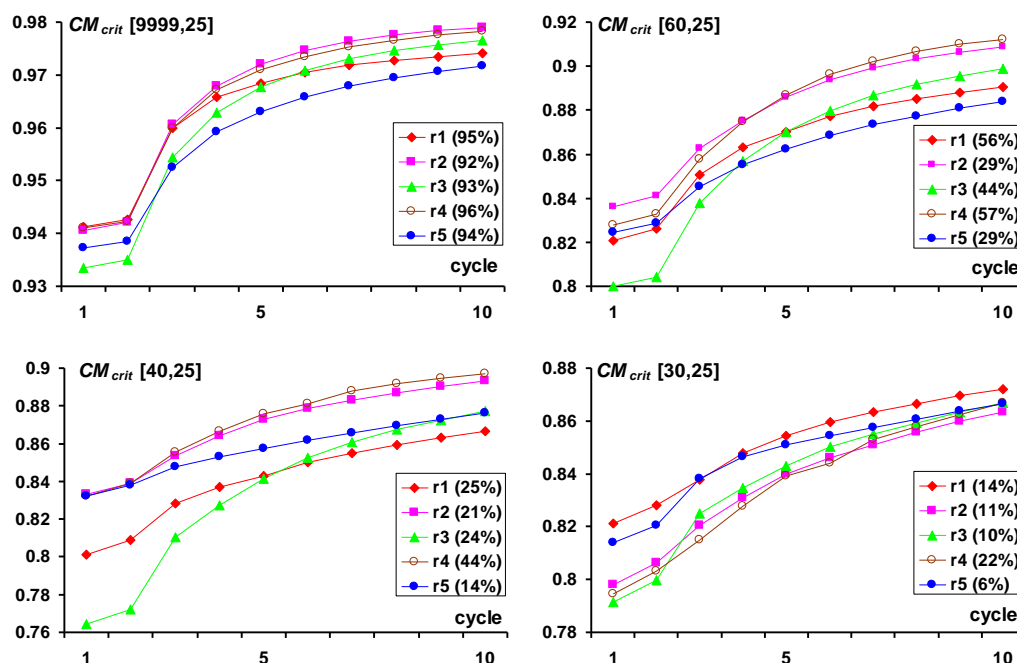


Рис. 3. Для пяти независимых решений показано изменение величин критериев отбора масок в процессе уточнения значений фаз. Цифры в рамке соответствуют достигнутым значениям корреляции фаз согласно таблице 3. Работа велась в ячейке  $c1$ .

Таблица 4. Результаты усреднения наборов фаз, рассчитанных по случайно сгенерированным связным маскам, приводящим к заданному уровню корреляции модулей структурных факторов. Приведены значения фазовой корреляции  $CP$ , рассчитанной по зонам разрешения  $\infty-60$ ,  $\infty-40$ ,  $\infty-30$ ,  $\infty-25$  (верхняя строка) и  $60-40$ ,  $40-30$ ,  $30-25$  Å (нижняя строка). Порог отбора  $z_{crit}$  определялся по зоне разрешения  $\infty-25$  Å. Работа велась в ячейке  $c3$

$CP*100$		Объем маски: удельный [ $\text{Å}^3/\text{Da}$ ]/число точек					
		1.0 1846	1.2 2215	1.4 2584	1.6 2954	1.8 3322	2.0 3691
$z_{crit}$	No	72/67/65/63 21/27/06	76/69/64/62 01/-15/04	77/70/67/65 02/25/04	79/73/69/67 03/13/04	72/68/65/64 25/22/12	75/69/65/64 07/10/04
	1.5	80/73/70/68 00/19/-05	82/76/70/68 17/-07/-11	83/77/74/71 19/22/-05	86/81/76/74 26/07/-07	87/82/78/75 32/19/-15	87/82/76/74 31/-05/-09
	2.0	81/75/71/69 -01/23/11	86/80/75/73 14/10/-03	86/81/77/75 24/23/13	86/81/77/75 27/19/-04	88/84/79/76 32/19/-13	87/82/78/76 33/22/-13
	2.5	80/75/71/69 17/28/-01	86/81/77/74 22/22/-06	87/82/77/75 24/15/-01	88/84/78/76 33/05/-14	88/83/78/76 29/18/-12	90/86/81/78 41/08/-12
	3.0	86/80/76/74 10/21/03	89/82/78/76 16/15/01	89/84/79/76 30/13/-05	90/85/80/77 34/11/-12	90/86/81/79 39/21/-05	90/86/81/78 36/21/-13

Результаты уточнения значений фаз суммированы в таблице 3. На рисунке 3 показаны изменения пороговых значений критериев отбора допустимых масок в процессе уточнения значений фаз.

Работа в ячейках  $c3$  и  $c4$  была организована аналогично работе в ячейке  $c1$ . Результаты тестирования метода суммированы в таблицах 4–6 и графиках на рисунках 4 и 5.

**Таблица 5.** Результаты уточнения наборов фаз. Приведены значения фазовой корреляции  $CP$ , рассчитанной по зонам разрешения  $\infty-60$ ,  $\infty-40$ ,  $\infty-30$ ,  $\infty-25$  (верхняя строка) и  $60-40$ ,  $40-30$ ,  $30-25 \text{ \AA}$  (нижняя строка). Работа велась в ячейке  $c3$

$CP*100/\text{число}$ построенных масок	Стартовый набор фаз для построения априорного распределения (10)				
	$r1$	$r2$	$r3$	$r4$	$r5$
Start	87/81/77/75/ 23/23/07 36490	88/82/76/73/ 17/-12/-08 37336	87/81/74/73/ 18/-14/12 39327	87/81/76/74/ 22/02/05 35142	87/83/79/77/ 40/28/10 39069
Step 1	94/90/86/84/ 42/35/19 559	94/88/82/79/ 24/-00/-12 508	93/87/81/79/ 27/-09/19 529	92/88/83/81/ 35/12/16 537	94/91/87/85/ 55/42/20 573
Step 5	96/94/91/89/ 63/55/29 562	95/91/87/84/ 46/21/-03 154	95/91/85/82/ 45/05/-00 100	95/91/87/85/ 50/34/26 1250	96/93/91/89/ 65/56/28 163
Step 10	97/94/92/91/ 68/62/34 799	96/92/88/86/ 53/33/02 1000	95/91/86/83/ 52/13/00 1032	95/91/87/85/ 50/35/21 1292	96/94/91/90/ 69/59/36 871

**Таблица 6.** Результаты уточнения наборов фаз. Приведены значения фазовой корреляции  $CP$ , рассчитанной по зонам разрешения  $\infty-60$ ,  $\infty-40$ ,  $\infty-30$ ,  $\infty-25$  (верхняя строка) и  $60-40$ ,  $40-30$ ,  $30-25 \text{ \AA}$  (нижняя строка). Работа велась в ячейке  $c4$

$CP*100/\text{число}$ построенных масок	Стартовый набор фаз для построения априорного распределения (10)				
	$r1$	$r2$	$r3$	$r4$	$r5$
Start	90/85/81/79/ 25/28/09 34190	89/83/78/76/ 18/-02/09 38570	91/85/79/77/ 27/-16/09 35081	89/83/78/76/ 18/15/07 35394	91/86/82/80/ 24/25/20 32130
Step 1	96/92/88/86/ 46/37/06 505	95/90/84/82/ 40/02/07 561	94/90/85/83/ 47/13/20 438	94/89/85/83/ 37/16/21 524	97/93/89/87/ 47/34/29 517
Step 5	98/95/92/90/ 66/51/22 334	96/92/88/85/ 54/21/02 133	95/92/89/86/ 62/36/20 101	96/93/89/87/ 57/37/26 1425	98/96/94/93/ 73/57/51 141
Step 10	98/96/94/92/ 73/57/38 169	96/93/88/86/ 58/27/07 557	95/92/89/87/ 62/36/17 405	94/90/88/86/ 49/55/32 42780	99/97/95/94/ 79/65/56 420

При работе в ячейке  $c3$  вначале было проведено тестирование параметров метода: размера генерируемых масок и жесткости отбора (табл. 4). Эти результаты дают картину, близкую к той, что наблюдалась в тестах с ячейкой  $c1$ : оптимальные размеры маски лежат в области  $1.6-1.8 \text{ \AA}^3/\text{Da}$ , точность полученных значений фаз растет с повышением жесткости отбора. Однако, в целом, точность получаемых решений фазовой проблемы была выше, чем при работе в ячейке  $c1$ . Как и раньше, для генерации нескольких пробных решений  $r1-r5$  и их уточнения мы остановились на объеме маски  $1.6 \text{ \AA}^3/\text{Da}$  и жесткости отбора  $z_{crit} = 2.5$ . Такие же значения параметров были использованы и при работе в ячейке  $c4$ , где этап тестирования параметров был

опущен. Несмотря на то, что наилучшие результаты в таблицах 3 и 4 отвечают значению параметра жесткости 3.0, в дальнейшей работе использовалось меньшее значение 2.5. Это было сделано по двум причинам. Во-первых, использование  $z_{crit} = 3.0$  приводит к резкому росту объема вычислений, особенно для ячейки  $c4$ . Во-вторых, при работе с реальными объектами жесткость отбора должна быть согласована с точностью экспериментальных данных и не должна быть чрезмерно высокой.

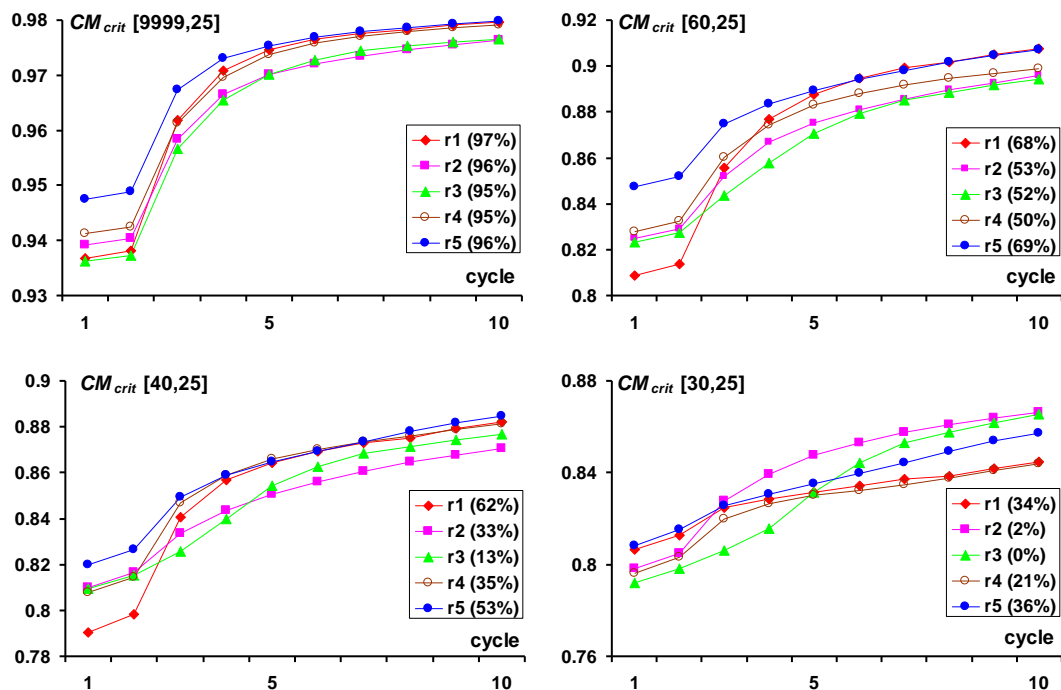


Рис. 4. Для пяти независимых решений показано изменение величин критериев отбора масок в процессе уточнения значений фаз. Цифры в рамке соответствуют достигнутым значениям корреляции фаз согласно таблице 5. Работа велась в ячейке  $c3$ .

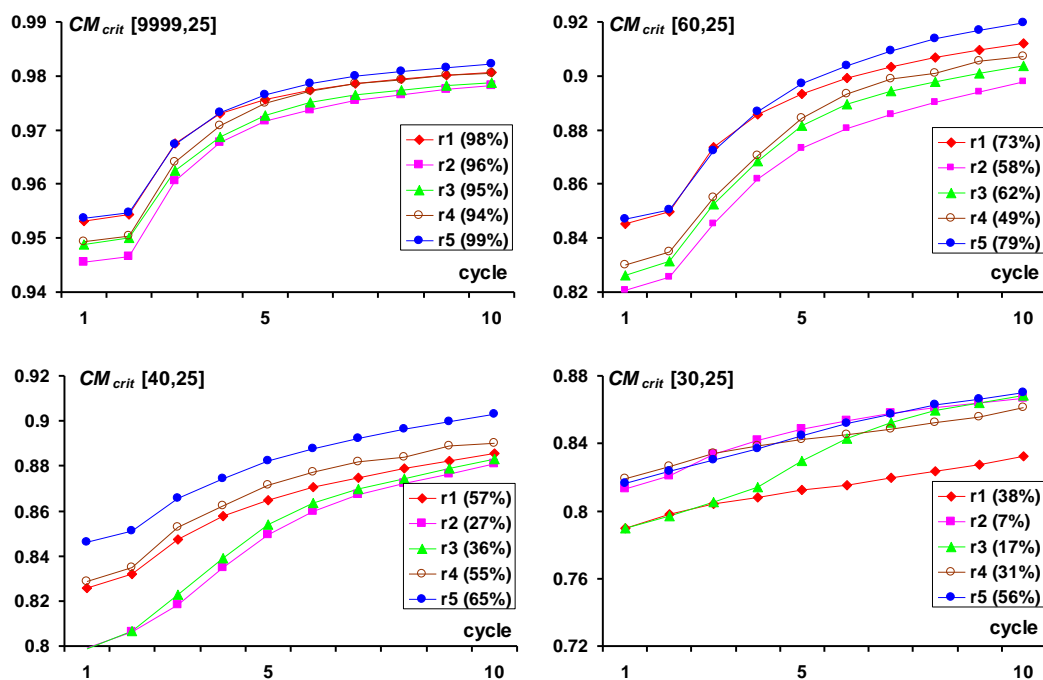


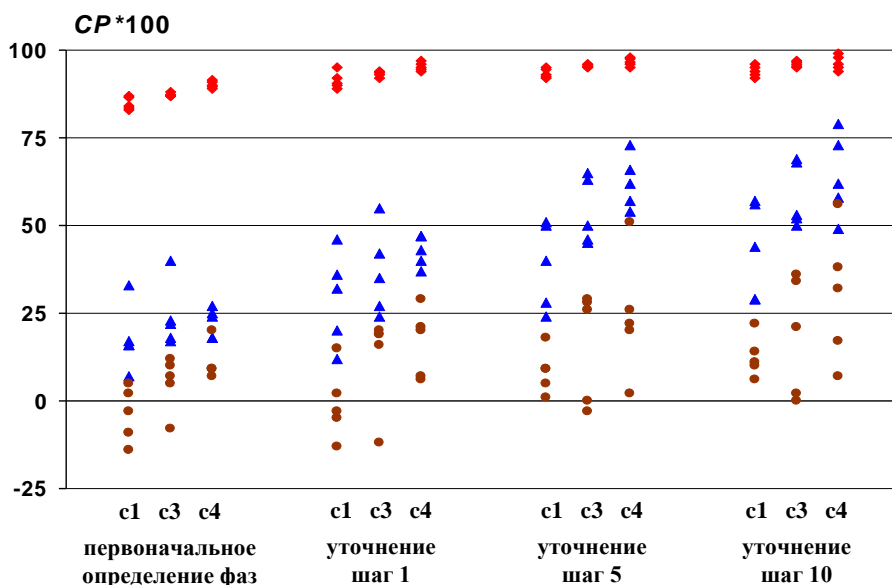
Рис. 5. Для пяти независимых решений показано изменение величин критериев отбора масок в процессе уточнения значений фаз. Цифры в рамке соответствуют достигнутым значениям корреляции фаз согласно таблице 6. Работа велась в ячейке  $c4$ .



## ДИСКУССИЯ И ЗАКЛЮЧЕНИЕ

На рисунке 6 представлена точность решений фазовой проблемы, получаемых на разных стадиях и при различном выборе вспомогательной воображаемой ячейки кристалла. Из представленной диаграммы и таблиц 3, 5 и 6 видно, что увеличение размеров воображаемой ячейки приводит к улучшению результата, причем это улучшение имеет место и за пределом Найквиста (определяемого удвоенным размером частицы). Из рисунка 6 и таблицы 3 видно, что даже при использовании ячейки размером меньшим, чем это диктуется пределом Найквиста, предложенная процедура позволяет получить информацию о значениях фаз, хотя и более ограниченную по разрешению, нежели при использовании больших ячеек. Возможность получения фазовой информации даже при ограниченных размерах ячейки связана с тем, что введение дополнительных требований бинарности и связности области вносит дополнительную информацию в процесс определения фаз, позволяющую отсеять ложные решения.

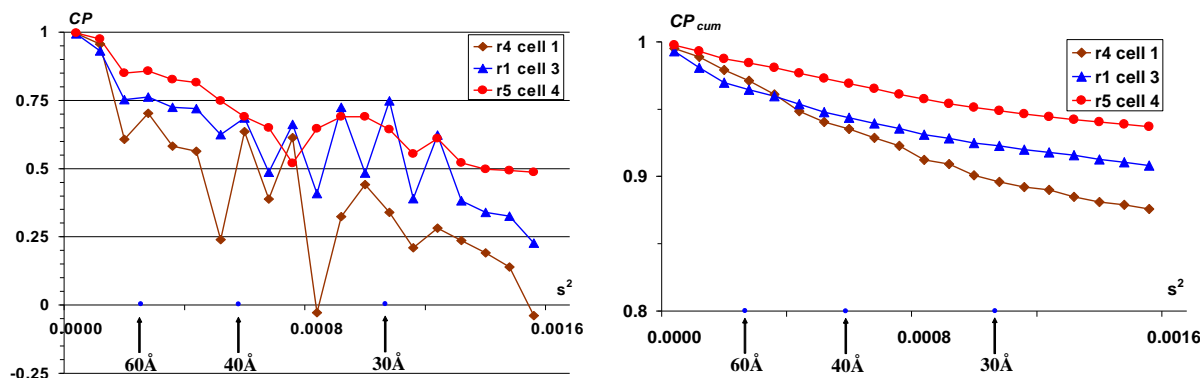
На рисунке 7 показаны подробные графики фазовой корреляции для различных оболочек обратного пространства для лучших из решений, определенных при разных размерах ячейки. Аналогично таблицам 5 и 6, эти графики подтверждают повышение точности определяемых наборов фаз при увеличении параметров воображаемой ячейки и вовлечении в работу большего количества экспериментальной информации. Ход кривой, соответствующей лучшему полученному решению ( $r_5$  для  $c_4$ ), повторяет ход кривой для средней интенсивности рефлексов (рис. 1). Это отражает общую тенденцию *ab initio* методов решения фазовой проблемы: более надежное определение значений фаз для более сильных рефлексов. Там же показаны значения фазовой корреляции для расширяющихся зон рефлексов  $\infty - d_{\min}$ . Здесь также видна тенденция повышения точности решений при увеличении размеров ячейки.



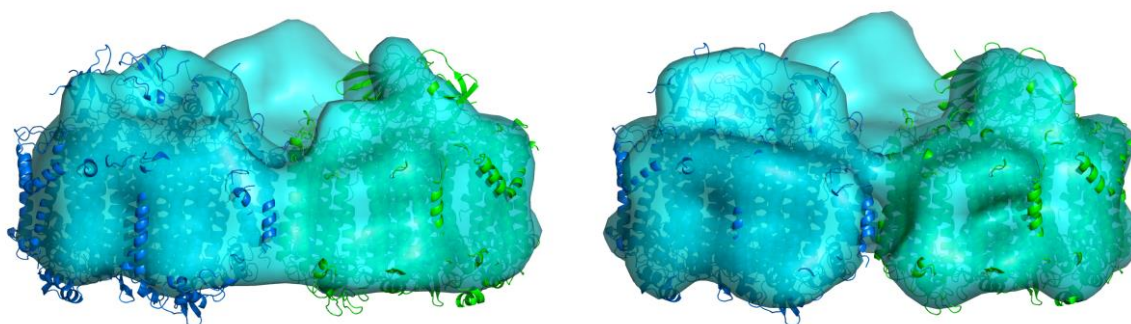
**Рис. 6.** Корреляция найденных на разных стадиях работы значений фаз с точными значениями. Показаны корреляции  $CP[\infty-60]$  (красные ромбы),  $CP[60-40]$  (синие треугольники),  $CP[30-25]$  (коричневые круги). Для каждого шага работы и типа ячейки пять одинаковых маркеров соответствуют пяти независимым решениям ( $r_1-r_5$ ).

На рисунках 8 и 9 показано изображение тримера, получаемое с использованием лучшего из найденных решений ( $r_5$  для  $c_4$ ), в сравнении с изображением, получаемым на основе точного синтеза разрешения  $25 \text{ \AA}$ .

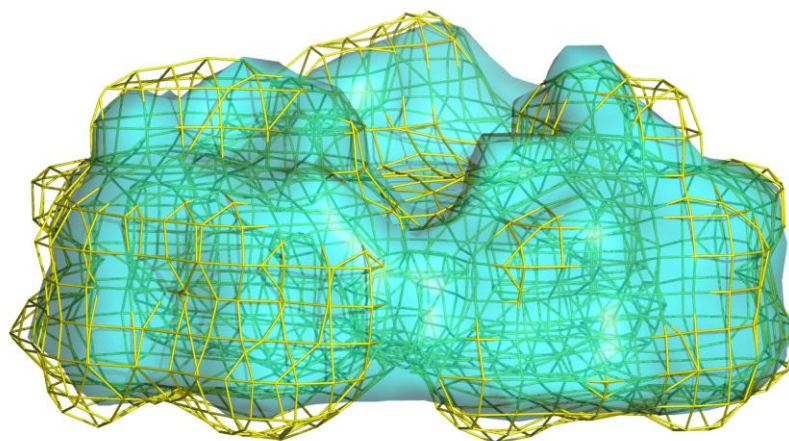




**Рис. 7.** Корреляции  $CP$  найденных значений фаз с точными. Слева: корреляции для тонких сферических оболочек в обратном пространстве. Справа: корреляции для расширяющихся сфер в обратном пространстве. Коэффициенты корреляции показаны как функция квадрата радиуса оболочки либо сферы  $s^2 = (2 \sin \theta / \lambda)^2$ .



**Рис. 8.** Изображения тримера разрешения 25 Å, построенные с использованием лучшего из найденных наборов фаз (слева) и с использованием набора точных значений фаз разрешения 25 Å (справа). Показаны области, соответствующие удельному объему 1.23 Å<sup>3</sup>/Da. Для сравнения наложена атомная модель белковой части тримера PS-I для двух из трех мономеров.



**Рис. 9.** Изображения тримера PS-I разрешения 25 Å, построенные с использованием лучшего из найденных наборов фаз (поверхность) и с использованием набора точных значений фаз (сетка). Показаны области, соответствующие удельному объему 1.23 Å<sup>3</sup>/Da.

При работе с реальным объектом точные значения фаз неизвестны, и следовательно, величины фазовых корреляций  $CP$  не могут быть использованы для выбора лучшего из серии решений, полученных независимо ( $r1 - r5$  в наших тестах). В работе [15]

обсуждается подход к выбору наилучшего решения на базе расчета величин обобщенного правдоподобия для получаемых решений. Графики на рисунке 5 показывают, что в качестве дополнительного индикатора качества полученных решений могут использоваться достигаемые в процессе уточнения пороговые значения критериев отбора. Хотя эта характеристика и не находится во взаимно однозначном соответствии с качеством решений, наилучшее решение при работе в ячейке *c4* могло бы быть опознано на основе поведения графиков, представленных на рисунке 5: решение (*r5*) доминирует над остальными по достигаемым порогам отбора во всех рассматриваемых зонах разрешения. Ситуация становится менее четкой при работе в ячейке *c3* (рис. 4). Здесь характеристики для двух лучших решений доминируют только в трех первых зонах разрешения. Однако следует заметить, что для четвертой зоны коэффициенты фазовой корреляции для всех пяти вариантов являются неприемлемо низкими. Здесь можно сказать, что реальное разрешение полученных наборов фаз ограничено величиной 30 Å. Похожая ситуация наблюдается и при работе в ячейке *c1* (рис. 3), где реальное разрешение лучшего решения ограничено величиной около 40 Å. Таким образом, можно сказать, что обсуждаемый критерий выделяет лучшее решение, ежели это решение достаточно хорошее, но не отражает иерархию слабых решений.

Результаты проведенных тестов подтверждают работоспособность предложенного метода решения фазовой проблемы и демонстрируют преимущества, получаемые при использовании больших размеров ячейки воображаемого кристалла и, соответственно, вовлечении в процесс решения структуры большего количества экспериментальных данных. Однако следует отметить, что проведенные тесты были выполнены на пределе вычислительных возможностей лабораторных компьютеров. Дальнейшее увеличение объема используемой экспериментальной информации потребовало бы модификации программного обеспечения и переноса вычислений на компьютеры с параллельной архитектурой. Также следует отметить, что возможности уменьшения шага дискретизации экспериментальных данных могут быть ограничены техническими характеристиками экспериментального оборудования, например, размерами пикселя детектора, регистрирующего рассеянное рентгеновское излучение.

Работа выполнена при поддержке гранта РФФИ № 13-04-00118.

## СПИСОК ЛИТЕРАТУРЫ

1. Miao J., Kirz J., Sayre D. The oversampling phasing method. *Acta Crystallographica Section D: Biological Crystallography*. 2000. V. 56. P. 1312–1315.
2. Thibault P., Elser V., Jacobsen C., Shapiro D., Sayre D. Reconstruction of a yeast cell from X-ray diffraction data. *Acta Crystallographica Section A: Foundations of Crystallography*. 2006. V. 62. P. 248–261.
3. Song C., Jiang H., Mancuso A., Amirbekian B., Peng L., Sun R., Shah S.S., Zhou Z.H., Ishikawa T., Miao J. Quantitative Imaging of Single, Unstained Viruses with Coherent X Rays. *Physical Review Letters*. 2008. V. 101. Article No. 158101. doi: [10.1103/PhysRevLett.101.158101](https://doi.org/10.1103/PhysRevLett.101.158101).
4. Maia F.R.N., Ekeberg T., van der Spoel D., Hajdu J. *Journal of Applied Crystallography*. 2010. V. 43. P. 1535–1539.
5. Seibert M.M., Ekeberg T, Maia F.R.N.C., Svenda M., Andreasson J., Jönsson O., Odic D., Iwan B., Rucker A., Westphal D. et al. Single mimivirus particles intercepted and imaged with an X-ray laser. *Nature*. 2011. V. 470. P. 78–82.
6. Van der Schot G., Svenda M., Maia F.R.N.C., Hantke M., DePonte D., Seibert M.M., Aquila A., Schulz J., Kirian R., Liang M. et al. Imaging single cells in a beam of live

- цианобактерия с помощью X-лучевого лазера. *Nature Communication*. 2015. V. 6. Article No 5704. doi: [10.1038/ncomms6704](https://doi.org/10.1038/ncomms6704).
7. Свергун Д.И., Фейгин Л.А. *Рентгеновское и нейтронное малоугловое рассеяние*. М.: Наука, 1986.
  8. Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E. The Protein Data Bank. *Nucleic Acids Research*. 2000. V. 28. P. 235–242. URL: <http://www.rcsb.org/> (дата обращения: 22.11.2015).
  9. Sayre D. Some implications of a theorem due to Shannon. *Acta Crystallographica*. 1952. V. 5. P. 843.
  10. Bricogne, G. Geometric sources of redundancy in intensity data and their use for phase determination. *Acta Crystallographica Section A: Foundations of Crystallography*. 1974. V. 30. P. 349–405.
  11. Bricogne G. Methods and programs for direct-space exploitation of geometric redundancies. *Acta Crystallographica Section A: Foundations of Crystallography*. 1976. V. 32. P. 832–847.
  12. Fienup J.R. Reconstruction of an object from the modulus of its Fourier transform. *Optics Letters*. 1978. V. 3. № 1. P. 27–29.
  13. Zhang K.Y.J., Cowtan K.D., Main P. Phase improvement by iterative density modification. In: *International Tables for Crystallography*. 2012. V. F. P. 385–400. doi: [10.1107/97809553602060000847](https://doi.org/10.1107/97809553602060000847).
  14. Лунин В.Ю., Лунина Н.Л., Петрова Т.Е. Использование связанных масок в задаче восстановления изображения изолированной частицы по данным рентгеновского рассеяния. *Математическая биология и биоинформатика*. 2014. Т. 9. № 2. С. 543–562. doi: [10.17537/2014.9.543](https://doi.org/10.17537/2014.9.543).
  15. Lunin V.Y., Lunina N.L., Petrova T.E., Baumstark M.W., Urzhumtsev A.G. Mask-based approach to phasing of single-particle diffraction data. *Acta Crystallographica Section D: Biological Crystallography*. 2016. V. 72. (в печати).
  16. Jordan P., Fromme P., Witt H.T., Klukas O., Saenger W., Krauß N. Three-dimensional structure of cyanobacterial photosystem I at 2.5 Å resolution. *Nature*. 2001. V. 411. P. 909–917.
  17. Rossmann M.G., Arnold E. Noncrystallographic symmetry averaging of electron density for molecular-replacement phase refinement and extension. In: *International Tables for Crystallography*. 2012. V. F. P. 352–363. doi: [10.1107/97809553602060000842](https://doi.org/10.1107/97809553602060000842).
  18. Matthews B.M. Solvent content of protein crystals. *Journal of Molecular Biology*. 1968. V. 33. P. 491–497.
  19. Weichenberger C.X., Rupp B. Ten years of probabilistic estimates of biocrystal solvent content: new insights via nonparametric kernel density estimate. *Acta Crystallographica Section D: Biological Crystallography*. 2014. V. 70. P. 1579–1588.

Материал поступил в редакцию 24.11.2015, опубликован 03.12.2015.