

СТРОЕНИЕ ХИМИЧЕСКИХ СОЕДИНЕНИЙ, СПЕКТРОСКОПИЯ

УДК 539.26

О ВОЗМОЖНОСТИ ОПРЕДЕЛЕНИЯ СТРУКТУРЫ БЕЛКОВ В ЭКСПЕРИМЕНТАХ С ИСПОЛЬЗОВАНИЕМ РЕНТГЕНОВСКИХ ЛАЗЕРОВ НА СВОБОДНЫХ ЭЛЕКТРОНАХ В РЕЖИМЕ НЕСТАЦИОНАРНОГО РАССЕЯНИЯ

© 2017 г. Д. О. Синицын^{1*}, Г. А. Армеев²

¹Институт химической физики им. Н.Н. Семенова Российской академии наук, Москва

²Московский государственный университет им. М.В. Ломоносова

*E-mail: d_sinitsyn@mail.ru

Поступила в редакцию 23.12.2015

Рассмотрена задача определения структуры белков методом рентгеноструктурного анализа с использованием рентгеновских лазеров на свободных электронах в случае интенсивностей импульсов, на несколько порядков превышающих достигнутые на настоящий момент. При этом вследствие радиационного повреждения за время импульса образец теряет значительную часть электронов, и рассеяние происходит в нестационарном режиме. В работе путем моделирования изучены возникающие отсюда особенности дифракционных данных, которые должны учитываться при их анализе, однако не являются препятствием для определения трехмерных структур. Оценки достижимых разрешений показывают перспективность режима нестационарного рассеяния для определения структуры наиболее сложных объектов — крупных, плохо кристаллизуемых белков и комплексов — с высоким разрешением.

Ключевые слова: кристаллография, рентгеновский лазер на свободных электронах, радиационное повреждение, разрешение, нестационарное рассеяние.

DOI: 10.7868/S0207401X17010113

1. ВВЕДЕНИЕ

Рентгеновская кристаллография — один из основных методов определения пространственной структуры молекул, в том числе биологических полимеров. Этот метод позволил определить структуры ДНК и десятков тысяч белков, однако работа с некоторыми классами объектов, такими как многие мембранные белки и сложные молекулярные комплексы, до сих пор вызывает существенные трудности, связанные с плохой кристаллизуемостью и проблемой радиационного повреждения [1]. В последние годы возможности рентгеноструктурного анализа были существенно расширены за счет появления рентгеновских лазеров на свободных электронах (РЛСЭ) [2–4].

Основными отличиями этих установок от традиционных рентгеновских источников — синхротронов являются сверхвысокая пиковая мощность (которая в настоящее время на 9 порядков выше, чем у лучших синхротронов) и ультракороткая длительность импульсов (единицы–десятки фемтосекунд) [5–7]. При этом радиационное повреждение существенно изменяет структуру

образца¹⁾ через десятки–сотни фемтосекунд от начала импульса, поэтому импульс длительно в единицы фемтосекунд заканчивается до наступления эффектов повреждения и регистрируемые данные свободны от существенных искажений (принцип, получивший название “дифракции до разрушения” [5]). Вследствие этого появилась возможность на порядки увеличить допустимые дозы облучения и получать интерпретируемые дифракционные картины даже от наноразмерных белковых кристаллов, что привело к зарождению серийной фемтосекундной нанокристаллографии [8]. Помимо случаев плохой кристаллизуемости, использование нанокристаллов также желательно для экспериментов по изучению конформационных переходов, индуцированных светом, так как малый размер кри-

¹⁾Под существенным изменением имеется в виду изменение, относительная величина которого порядка или больше типичной погрешности используемых моделей, а именно относительного отличия картины дифракции, рассчитанной по атомной модели макромолекулы, от экспериментальной дифракционной картины. Данная погрешность (R-фактор) в биологической кристаллографии при разрешениях 1–2 Å составляет 10–20%.

стала позволяет добиться более равномерной фотоактивации [9].

Однако для сложных объектов, таких как крупные мембранные белковые комплексы, разрешение структурных данных, достигнутое на данный момент по этой методике, составляет около $2.5\text{--}5\text{ \AA}$ [10–13] и не достигает атомного разрешения, уступая многим традиционным экспериментам с использованием синхротронов [14]. Кроме того, в большинстве работ используются кристаллы размером от одного до нескольких микрон, и вопрос о пригодности нанокристаллов сложных макромолекул для получения данных высокого разрешения остается открытым.

Приведенные в настоящей работе теоретические оценки интенсивности излучения, рассеиваемого белковыми нанокристаллами, показывают, что получение атомного разрешения на нанокристаллах крупных белков требует увеличения интенсивности рассеиваемого излучения. Один из основных способов достижения этой цели — повышение мощности импульсов рентгеновских лазеров. Это потребует модификации конструкции РЛСЭ, например по схеме, приведенной в работе [15].

В то же время при высоких интенсивностях становится существенным²⁾ электронное повреждение образца во время импульса: даже в течение его фемтосекундной длительности за счет фотоионизации и вызванных ею процессов атомы образца успевают потерять часть своих электронов, и этой потерей нельзя пренебречь по сравнению с типичными погрешностями используемых теоретических моделей [5–7, 16–23]. Таким образом, дифракционный эксперимент происходит в режиме *нестационарного рассеяния*, т.е. рассеяния на изменяющемся образце [20, 23–25]. В предыдущих работах путем моделирования было показано умеренное влияние этих эффектов на регистрируемые дифракционные данные в случае падающей плотности потока, составляющей 10^6 фотонов/ $\text{\AA}^2$ за импульс (что на 2–3 порядка выше достигнутой на данный момент), и существенное влияние при плотности 10^8 фотонов/ $\text{\AA}^2$, в то время как при текущих плотностях потока этими эффектами можно пренебречь [7, 20–23].

Наличие электронного повреждения (в случаях, когда оно существенно) должно учитываться в экспериментах по определению трехмерных структур макромолекул по дифракционным дан-

ным на всех стадиях работы. На стадии планирования эксперимента необходимы оценки возможных значений рассеиваемой интенсивности излучения с учетом эффектов повреждения. На основе этих оценок можно приближенно предсказать максимальное разрешение дифракционных данных, достижимое в данных условиях. На стадии обработки данных рассеяния необходимо решить так называемую фазовую проблему, т.е. определить фазы комплексных значений структурных факторов (в то время как дифракционная картина позволяет измерить только их абсолютные величины). В случае нестационарного рассеяния на этом этапе требуется оценка влияния повреждения на фазы структурных факторов. При построении атомных моделей макромолекул необходимо учитывать характерные отличия электронной плотности, рассчитываемой по данным нестационарного рассеяния, от электронной плотности неповрежденного образца. Наконец, на стадии уточнения атомных моделей необходим учет нестационарной дифракции с максимально возможной точностью для обеспечения соответствия расчетных и экспериментальных данных.

В настоящей работе последовательно рассмотрены эти этапы, описаны типичные данные, полученные путем моделирования для различных условий эксперимента, и сделаны выводы о возможности и ключевых условиях успешного построения трехмерных атомных структур макромолекул по данным нестационарного рентгеновского рассеяния.

2. МОДЕЛИРОВАНИЕ НЕСТАЦИОНАРНОГО РАССЕЯНИЯ

Рентгеновская дифракция на образцах, типичных для биологической кристаллографии, может быть описана в рамках кинематической теории рассеяния. При малых длительностях импульсов (единицы фемтосекунд) положения ядер могут приближенно считаться неизменными. При этом электронные состояния атомов изменяются в течение импульса прежде всего за счет фотоионизации, оже-распадов и флуоресценции (*электронное повреждение*) [16]. Этот процесс моделируется с помощью кинетических уравнений, и в результате для заданных параметров импульса вычисляются зависящие от времени атомные факторы рассеяния $f_j(s; t)$ [7, 16, 21, 22, 25]. Поскольку в данной ситуации распределение электронной плотности образца изменяется со временем в течение действия рентгеновского импульса, генерируемого РЛСЭ, то мгновенное распределение интенсивности рассеиваемого образом излучения также зависит от времени. При этом детектор, используемый для регистрации дифракционной картины, измеряет суммарное количество фотонов, попавших на пиксель за все время им-

²⁾Под существенным электронным повреждением подразумевается потеря большинством атомов образца такой доли от общего числа электронов в атоме, которая сравнима или превышает типичные значения относительной погрешности теоретических моделей в биологической кристаллографии (10–20% [21]). Ввиду этого таким повреждением нельзя пренебречь в расчетах при интерпретации данных рассеяния.

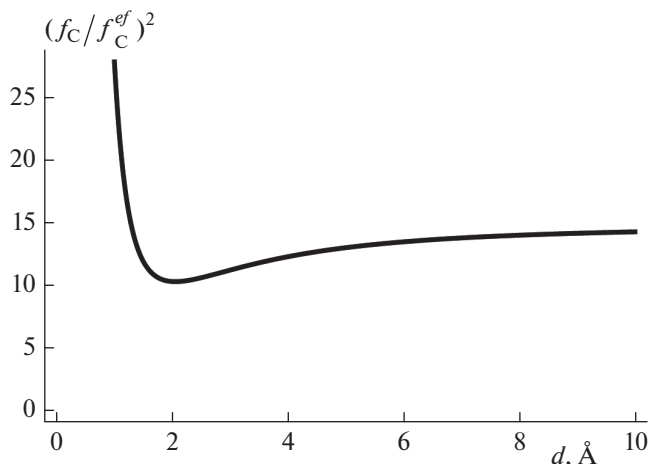


Рис. 1. Величина эффекта электронного повреждения: квадрат отношения формфактора атома углерода в отсутствие повреждения к формфактору атома углерода при плотности потока 10^8 фотонов/Å² как функция разрешения.

пульса. Таким образом, измеренные величины представляют собой интегралы по времени от мгновенных значений рассеиваемой интенсивности. Как показано в работе [23], эти величины могут быть приближенно рассчитаны с использованием следующих формул, аналогичных обычным формулам для стационарного случая:

$$I^{cum}(\mathbf{s}) \approx C |F^{ef}(\mathbf{s})|^2, \quad (1)$$

$$F^{ef}(\mathbf{s}) = \sum_{j=1}^{N_{atoms}} f_j^{ef}(s) \times \exp\left(-B_j s^2/4\right) \exp\{-2\pi i(\mathbf{s}, \mathbf{r}_j)\}, \quad (2)$$

где $\mathbf{s}$ — вектор рассеяния; $F^{ef}(\mathbf{s})$ — эффективный структурный фактор; $f_j^{ef}(s)$ — эффективные атомные формфакторы, рассчитываемые из зависящих от времени формфакторов $f_j(s; t)$ по изложенному в [23] алгоритму; B_j — фактор неупорядоченности (В-фактор), $\mathbf{r}_j$ — положение атома, N_{atoms} — число атомов в элементарной ячейке кристалла, C — постоянный множитель.

По структурным факторам $F^{ef}(\mathbf{s})$ можно рассчитать соответствующие значения фаз $\varphi^{ef}(\mathbf{s}) = \arg F^{ef}(\mathbf{s})$ и распределение электронной плотности $\rho^{ef}(\mathbf{r})$. Это распределение не отвечает какому-либо состоянию молекулы, а является вспомогательным математическим объектом, которому соответствует дифракционная картина, близкая к измеряемой картине интегральных интенсивностей $I^{cum}(\mathbf{s})$. Поэтому функция $\rho^{ef}(\mathbf{r})$

называется далее эффективным распределением электронной плотности. Ниже показано, что это распределение после подходящего масштабирования оказывается близким к истинному распределению электронной плотности неповрежденной ячейки кристалла, поэтому оно может быть интерпретировано и использовано для построения атомной модели аналогично схеме обработки данных без учета повреждения, используемой в кристаллографии на синхротронах.

Среднее количество фотонов, рассеиваемых кристаллом в области одного дифракционного максимума данного разрешения, может быть оценено по формуле [7]

$$N_{spot}(s) = (\lambda r_e)^2 \Phi \frac{a^4}{V_{UC}^2} P_T(s) \times \sum_{j=1}^{N_{atoms}} f_j^{ef}(s)^2 \exp\left(-B \frac{s^2}{2}\right), \quad (3)$$

где s — величина вектора рассеяния для данного максимума в Å⁻¹ ($s = 2 \sin(\theta)/\lambda$, где 2θ — угол рассеяния, λ — длина волны излучения; при этом $d = 1/s$ — разрешение в Å), Φ — плотность потока фотонов в падающем на образец рентгеновском импульсе (в фотонах на 1 Å² за импульс), $P_T(s) = [1 + (1 - s^2 \lambda^2/2)^2]/2$ — поляризационный множитель, $r_e = e^2/m_e c^2$ — классический радиус электрона, m_e — масса электрона, e — заряд электрона, c — скорость света, V_{UC} — объем элементарной ячейки, a — размер кристалла, B — средний фактор неупорядоченности. Следует иметь в виду, что формула (3) дает грубую оценку по порядку величины.

В настоящей работе проводились моделирование дифракции и оценка числа рассеянных фотонов по изложенной методике. В расчетах использовалась длина волны в 1.55 Å (энергия фотонов равна 8 кэВ), длительность импульсов в 10 фс и два значения плотности потока фотонов: 10^3 фотонов/Å² за импульс — текущее значение, достигнутое на лазере LCLS в США [2], и 10^8 фотонов/Å² за импульс — перспективное значение, открывающее возможность нанокристаллографии высокого разрешения.

Для иллюстрации влияния электронного повреждения на интенсивность рассеяния на рис. 1 приведен график отношения квадратов формфакторов атома углерода в отсутствие повреждения и при плотности потока 10^8 фотонов/Å². Это отношение показывает, во сколько раз интенсивности рефлексов данного разрешения будут уменьшаться из-за повреждения. Видно, что эффект наиболее сильно сказывается на высоких разрешениях порядка 1 Å. В среднем для данной

плотности потока эффект повреждения снижает интенсивность рассеяния более чем на порядок.

3. ЗАВИСИМОСТЬ РАЗРЕШЕНИЯ ОТ ПАРАМЕТРОВ

Рассмотрим условия получения данных дифракции высокого разрешения для различных сочетаний параметров эксперимента. Из формулы (3) следует, что число рассеянных фотонов убывает с ростом s , что и является лимитирующим фактором для максимального разрешения. Обозначим через N_{thresh} минимальное (пороговое) число рассеянных фотонов, при котором возможны детектирование рефлекса и оценка его интенсивности с достаточной точностью. Тогда получаем уравнение $N_{spot}(s) = N_{thresh}$, из которого максимальное разрешение может быть найдено как неявная функция всех входящих в уравнение величин. При этом параметры, повышающие интенсивность рассеяния, повышают и максимальное достижимое разрешение. Таким образом, разрешение зависит возрастающим образом от интенсивности импульсов, эффективных формфакторов и размера кристаллов и убывающим образом — от размера ячейки, В-фактора и минимального детектируемого числа фотонов.

Эффективные формфакторы убывают с ростом интенсивности импульсов. Тем не менее, для рассматриваемого диапазона значений Φ (10^3 – 10^8 фотонов/Å²) произведение $\Phi f_j^{ef}(s)^2$, входящее в формулу (3), возрастает с ростом Φ , хотя и медленнее, чем по линейному закону. Так, на всем указанном диапазоне Φ возрастает на 5 порядков, а средняя интенсивность рассеяния — менее чем на 4 порядка. Эффективные формфакторы убывают также с ростом длительности импульсов. Действительно, оже-распад и флуоресценция начинают существенно влиять на электронные состояния на более длительных временах, чем фотоионизация, и поэтому вклад этих процессов возрастает с ростом длительности импульса [7, 22].

Чтобы проиллюстрировать зависимость разрешения от нескольких параметров, возьмем одну правдоподобную и перспективную их комбинацию и исследуем, что происходит при изменении каждого из параметров в отдельности. Рассмотрим следующую стартовую конфигурацию: белок фотосистему-II (модель с PDB-кодом 4PBU): размер кристалла — 1 мкм, плотность потока фотонов — 10^8 на Å², В-фактор — 50 Å², минимальное детектируемое число фотонов $N_{thresh} = 1$. В качестве критерия высокого разрешения (близкого к атомному) используется значение 1.5 Å.

Из вида формулы (3) можно заметить, что наиболее сильные зависимости в ней — это экспоненциальная зависимость от В-фактора и зависи-

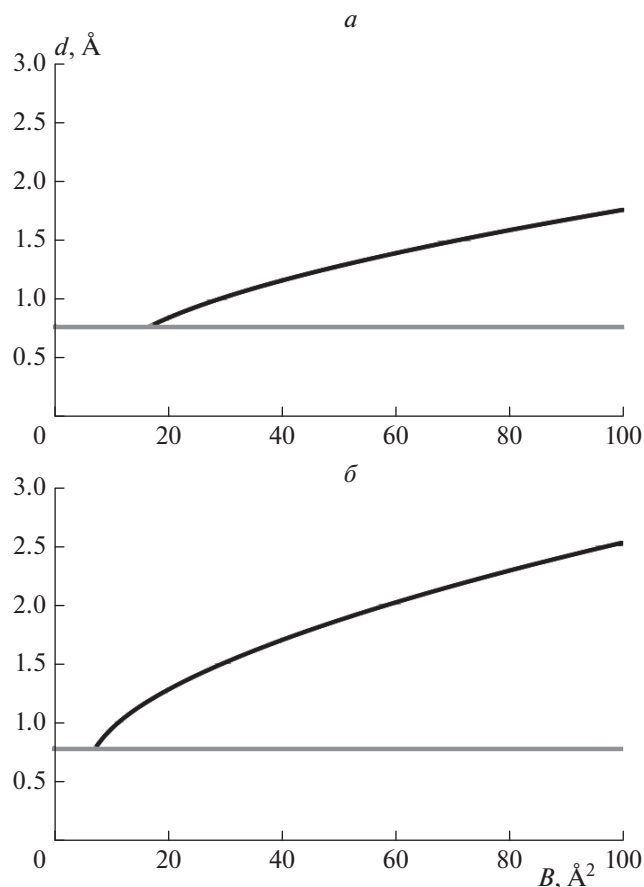


Рис. 2. *a* — Зависимость разрешения дифракционных данных d от фактора неупорядоченности B . Горизонтальная линия — максимальное разрешение для данной длины волны (предполагается, что детекторы могут быть расположены как за, так и перед образцом по ходу первичного пучка перпендикулярно направлению последнего, поэтому максимальное достижимое разрешение равно $\lambda/2$). Плотность потока фотонов в падающем импульсе — $10^8/\text{Å}^2$; *б* — то же для плотности потока 10^3 фотонов/Å² (достигнутой на существующих РЛСЭ).

мость 4-й степени от размера кристалла. Зависимость разрешения от В-фактора приведена на рис. 2. Видно, что для размера кристалла в 1 мкм и при плотности потока 10^8 фотонов/Å² разрешение в 1.5 Å можно получить на кристаллах с В-фактором, не превышающим более ~ 70 Å² (рис. 2*a*), а при достигнутой на данный момент плотности потока фотонов в 10^3 на 1 Å² — только с В-фактором не более 30 Å² (рис. 2*б*).

Ввиду этой сильной зависимости разрешения от В-фактора в дальнейшем на каждом графике мы изображаем три кривые, соответствующие хорошо упорядоченным кристаллам (для молекул такого размера), среднеупорядоченным и плохо

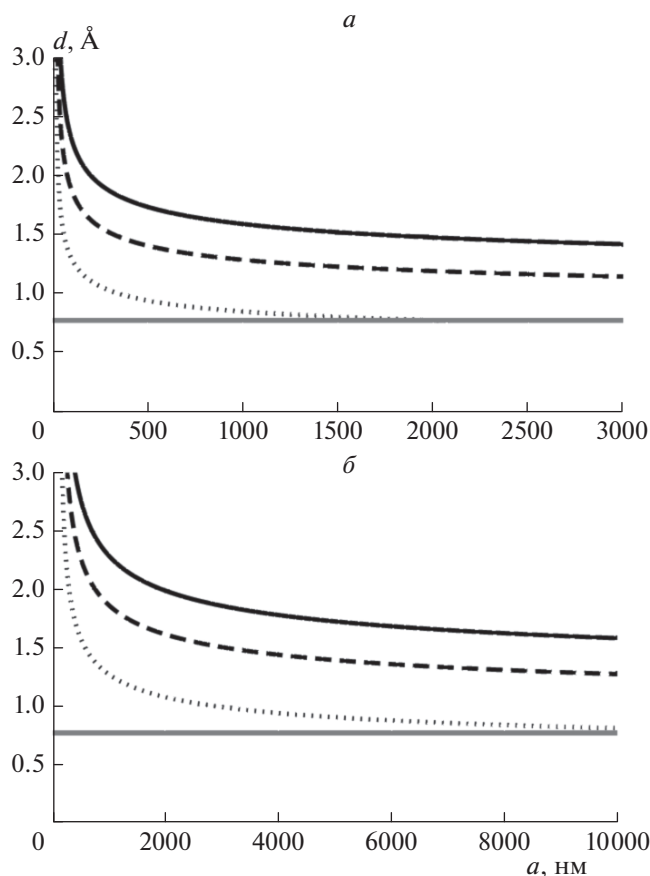


Рис. 3. То же, что и на рис. 2, но для зависимости разрешения d от размера кристалла a при следующих значениях В-фактора: сплошная линия — $80 \text{ \AA}^2$, штриховая линия — $50 \text{ \AA}^2$, точечная линия — $20 \text{ \AA}^2$.

упорядоченным с В-факторами в 20, 50 и $80 \text{ \AA}^2$ соответственно.

Зависимость разрешения от размера кристалла показана на рис. 3а. Видно, что для кристаллов со средней степенью упорядоченности (В-фактор $50 \text{ \AA}^2$) разрешение в $1.5 \text{ \AA}$ может быть достигнуто на нанокристаллах размером $\sim 500 \text{ нм}$, для менее упорядоченных (В-фактор $80 \text{ \AA}^2$) — на кристаллах размером $\sim 1.5 \text{ мкм}$. Для сравнения на рис. 3б показаны аналогичные графики для текущей плотности потока фотонов в 10^3 на $1 \text{ \AA}^2$ за импульс. В этом случае для среднеупорядоченных кристаллов получение разрешения в $1.5 \text{ \AA}$ возможно лишь при их размере $\sim 3 \text{ мкм}$, а в случае слабой упорядоченности — только при размерах в десятки микрометров.

Зависимость разрешения от размера элементарной ячейки (т.е. фактически от размера молекулярного объекта) показана на рис. 4 (в расчетах предполагается, что ячейка имеет форму куба и что количество атомов каждого типа растет пропорционально объему ячейки). Видно, что при

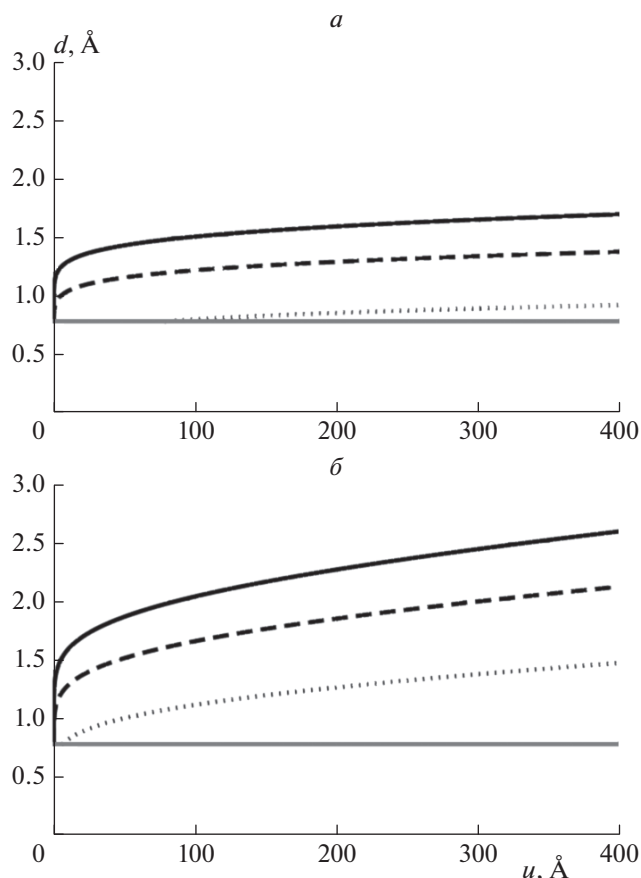


Рис. 4. То же, что и на рис. 3, для зависимости разрешения d от размера элементарной ячейки кристалла u .

фиксированном размере кристалла (1 мкм на рисунке) для крупных белков и комплексов разрешение будет ниже, чем для небольших макромолекул. Однако для рассматриваемого диапазона размеров ячейки данные остаются близки к разрешению $\sim 1.5 \text{ \AA}$. Что касается текущей интенсивности импульсов (рис. 4б), то в случае среднеупорядоченных кристаллов получение разрешения в $1.5 \text{ \AA}$ возможно лишь для простых белков с ячейкой размером меньше $50 \text{ \AA}$, а в случае плохо упорядоченных кристаллов размером 1 мкм — проблематично для любых макромолекул.

Зависимость разрешения от минимального детектируемого числа фотонов N_{thresh} показана на рис. 5. В случае хорошо и среднеупорядоченных кристаллов все изображенные значения N_{thresh} позволяют достигать максимального разрешения, для плохо упорядоченных кристаллов имеется плавное уменьшение разрешения с ростом N_{thresh} . Для текущей интенсивности импульсов (рис. 5б) аналогичные зависимости ограничивают разрешение при всех степенях упорядоченности.

Следует заметить, что в методике серийной нанокристаллографии экспериментальные интенсивности рассеяния получаются усреднением значений, измеренных в большом количестве дифракционных картин с переменными параметрами: картины регистрируются от разных кристаллов в разных ориентациях, под действием импульсов с флуктуирующими характеристиками и т.п. Так, в недавней работе [13] усреднялось порядка 600 значений интенсивности каждого рефлекса, извлеченных из ~28 млн детектированных изображений. Увеличение этого количества позволяет более точно оценивать малые интенсивности, величина которых сравнима с уровнем шума детектора, т.е. уменьшать N_{thresh} , тем самым повышая разрешение. Теоретически таким образом можно получать точные оценки интенсивностей даже таких рефлексов, на которые в среднем приходится меньше одного фотона. Однако проверка пределов эффективности этой процедуры затруднена с практической точки зрения как трудоемкостью накопления больших объемов данных, так и сложной природой флуктуаций измеряемых интенсивностей.

Таким образом, рассматриваемое перспективное значение плотности падающего потока фотонов в импульсе рентгеновского лазера, а именно 10^8 фотонов/Å² позволяет успешно детектировать данные околоатомного разрешения ~1.5 Å на нанокристаллах крупных белков, если они имеют степень упорядоченности не ниже средней, а для кристаллов микронного размера — и в случае относительно плохой упорядоченности. Обе эти ситуации делают крайне актуальными дальнейшее увеличение мощности РЛСЭ и работу в режиме нестационарного рассеяния, так как текущие интенсивности не позволяют получать в этих условиях данные атомного разрешения.

4. ВЛИЯНИЕ ПОВРЕЖДЕНИЯ ОБРАЗЦА НА МОДУЛИ И ФАЗЫ СТРУКТУРНЫХ ФАКТОРОВ

Для адекватной интерпретации данных нестационарного рассеяния с целью определения структуры молекул необходимо более детальное понимание отличия интенсивностей рассеяния и рассчитываемых на их основе распределений электронной плотности в условиях нестационарного рассеяния от аналогичных величин в стандартном случае, где повреждение мало и не учитывается.

Рассмотрим подробнее влияние электронного повреждения на дифракционные данные. В качестве модельного белка мы используем бычий родопсин (модель с PDB-кодом 1HZX). На рис. 6а показаны модули структурных факторов для рефлексов, лежащих в одной плоскости обратного

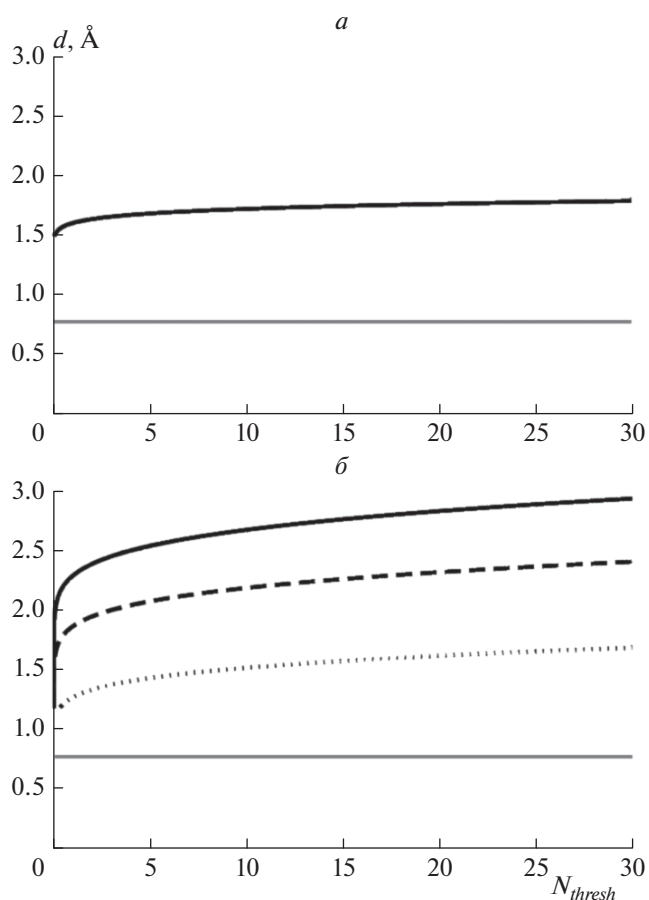


Рис. 5. То же, что и на рис. 3, но для зависимости разрешения d от минимального детектируемого числа фотонов в дифракционном максимуме N_{thresh} (в случае *a* штриховая и точечная линии не показаны, так как совпадают с горизонтальной линией максимального разрешения).

пространства, в случае нестационарного рассеяния при плотности потока 10^8 фотонов/Å², а на рис. 6б — аналогичные данные без учета повреждения. Изображенный участок плоскости содержит рефлексы высокого разрешения (~1.5 Å). Из сравнения рис. 6б и *a* видно сильное снижение интенсивности вследствие повреждения. При этом интенсивности рефлексов уменьшаются в разное число раз, поэтому эффект повреждения не может быть смоделирован за счет умножения всех атомных формфакторов на масштабирующий множитель [23]. Для корректного учета повреждения при решении прямой задачи рассеяния необходим модифицированный алгоритм, основанный на уравнениях (1), (2).

Важный вопрос состоит в том, как влияет электронное повреждение образца на фазы структурных факторов. На рис. 7 показаны разности фаз структурных факторов из рис. 6, рассчитанных с учетом повреждения по формуле (2) и

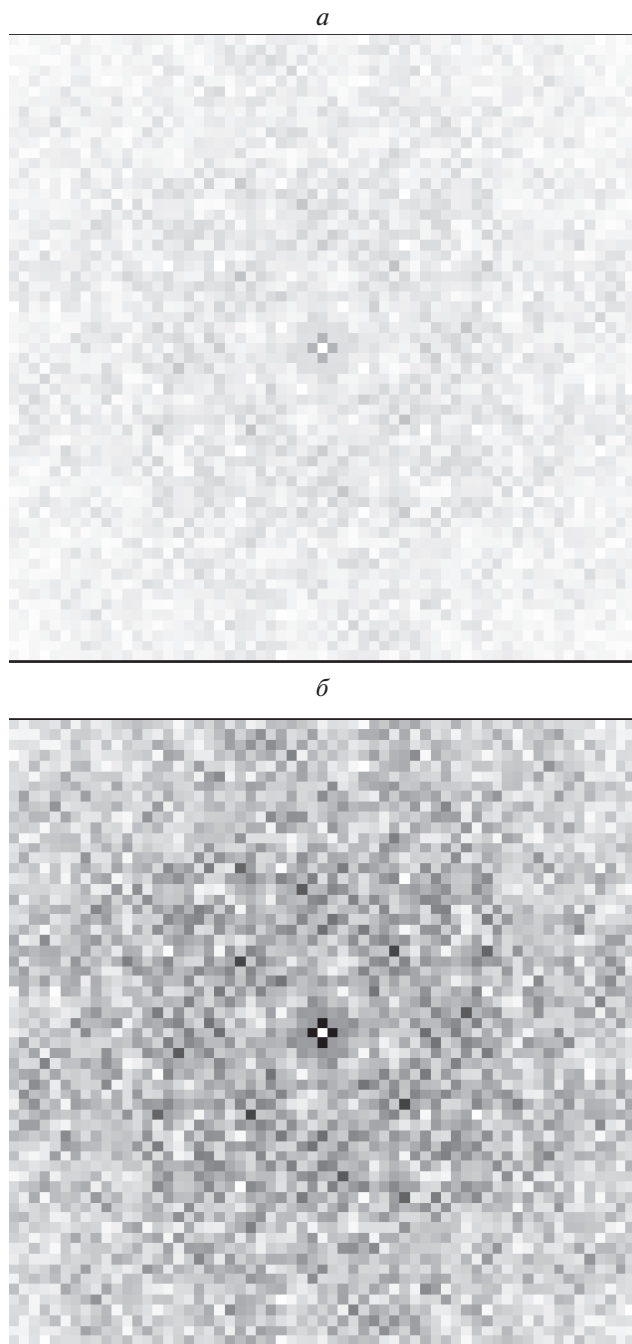


Рис. 6. *a* — Влияние повреждения на интенсивности дифракционных максимумов на примере белка бычьего родопсина (PDB-код — 1HZX). Показаны модули структурных факторов для рефлексов, лежащих в плоскости обратного пространства с индексами $(h, k, 90)$, $h, k = -30 \dots 30$ (более темный цвет соответствует большим значениям). Разрешение на границе ~ 1.5 Å. Плотность потока фотонов — $10^8/\text{Å}^2$; *б* — то же без учета повреждения.

без учета повреждения. В среднем разности фаз невелики (среднее близко к нулю, стандартное отклонение $\sim 5^\circ$), однако следует иметь в виду, что отдельные отклонения фаз могут достигать более

высоких значений (в рассмотренном случае — до 36°). Наличие таких разностей фаз может оказать некоторое влияние на отдельные элементы карты электронной плотности и, соответственно, получаемой атомной модели. В то же время рис. 7*в*, изображающий модули структурных факторов совместно с абсолютными величинами разностей фаз, показывает, что между ними есть отрицательная корреляция, поэтому рефлексy, имеющие большие отклонения фаз, будут, как правило, вносить меньший вклад в синтез электронной плотности.

Заметим, что при решении фазовой проблемы методом молекулярного замещения в ситуации, когда интенсивности измерены в условиях электронного повреждения, можно добиться более точного соответствия данных для исследуемой и гомологичной молекул, если рассчитывать фазы для гомологичной модели по алгоритму (1), (2) с учетом повреждения. Что касается экспериментальных методов решения фазовой проблемы, то соответствующие эксперименты желательно проводить в одних и тех же условиях, определяющих повреждение. Однако небольшая величина различия фаз (рис. 7) делает возможным использование и фазовых данных, получаемых от неповрежденных структур, для построения моделей с точностью, достаточной для локализации оптимального решения (с последующим уточнением по алгоритму, учитывающему повреждение).

5. КАРТЫ ЭЛЕКТРОННОЙ ПЛОТНОСТИ И ИХ ИНТЕРПРЕТАЦИЯ С УЧЕТОМ ПОВРЕЖДЕНИЯ

Промежуточным этапом интерпретации дифракционных данных с целью построения атомной модели является синтез карт электронной плотности. Рассмотрим их характерные особенности в ситуации, когда они получены по данным нестационарного рассеяния. В качестве модельной молекулы для простоты и наглядности изображений используем небольшой пептид из шести аминокислот (PDB-код — 3PZZ). Все данные приведены для импульсов с плотностью потока 10^8 фотонов/Å².

На рис. 8*а–в* показаны карты электронной плотности с учетом повреждения, без учета последнего и совмещение этих карт соответственно. Наиболее заметное различие касается средней величины электронной плотности: поскольку факторы атомов уменьшаются из-за повреждения в 3–4 раза (см. рис. 1), то в такое же число раз падает и электронная плотность. Этот эффект может быть частично скомпенсирован масштабным коэффициентом, на который умножаются значения всех структурных факторов и который выбирается таким, чтобы отличие от неповрежденной

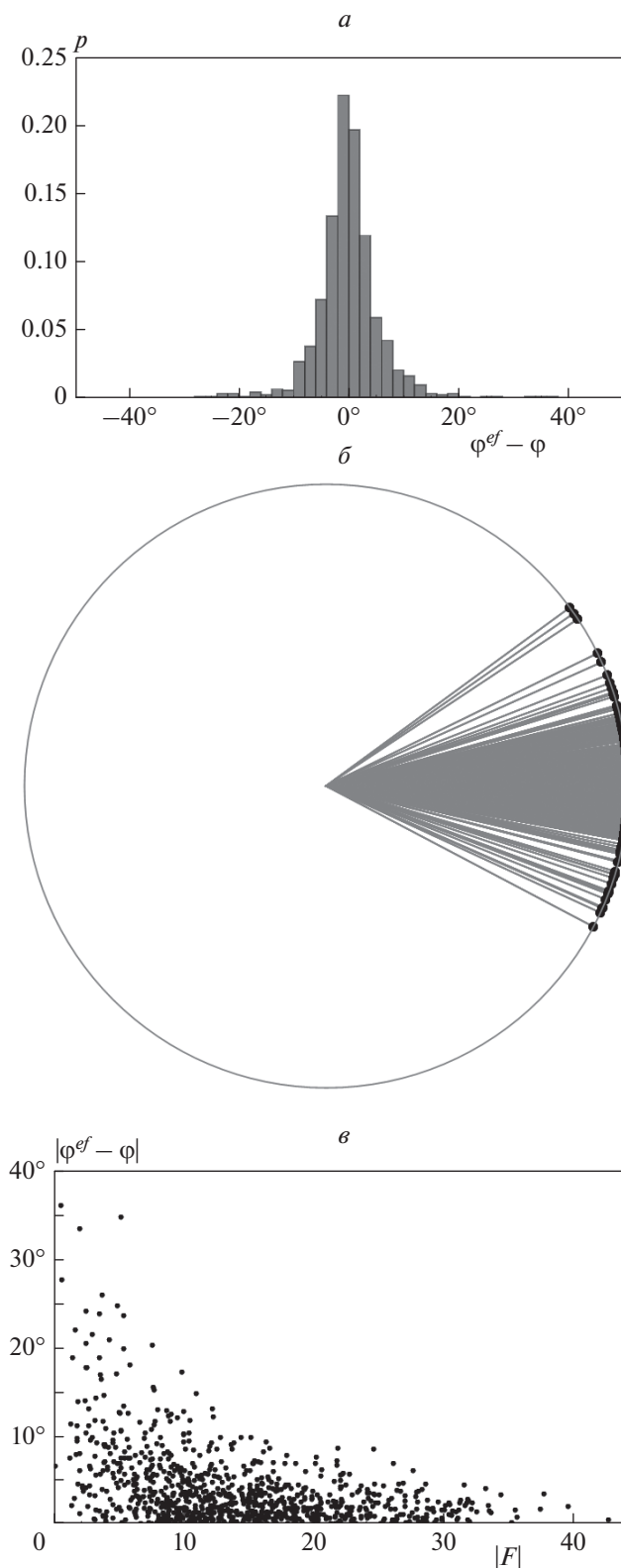


Рис. 7. *a* — Влияние электронного повреждения на фазы структурных факторов. Показана гистограмма распределения разностей фаз структурных факторов из рис. 6, рассчитанных по формуле (2) с учетом повреждения (φ^{ef}) и без учета повреждения (φ); p — частота. Среднее значение разности фаз -0.02° , стандартное отклонение -5.8° . *б* — Индивидуальные значения разностей фаз, изображенные в виде радиусов под соответствующими углами к горизонтальной оси. *в* — Зависимость модуля разности фаз от модуля структурного фактора F , рассчитанного без учета повреждения.

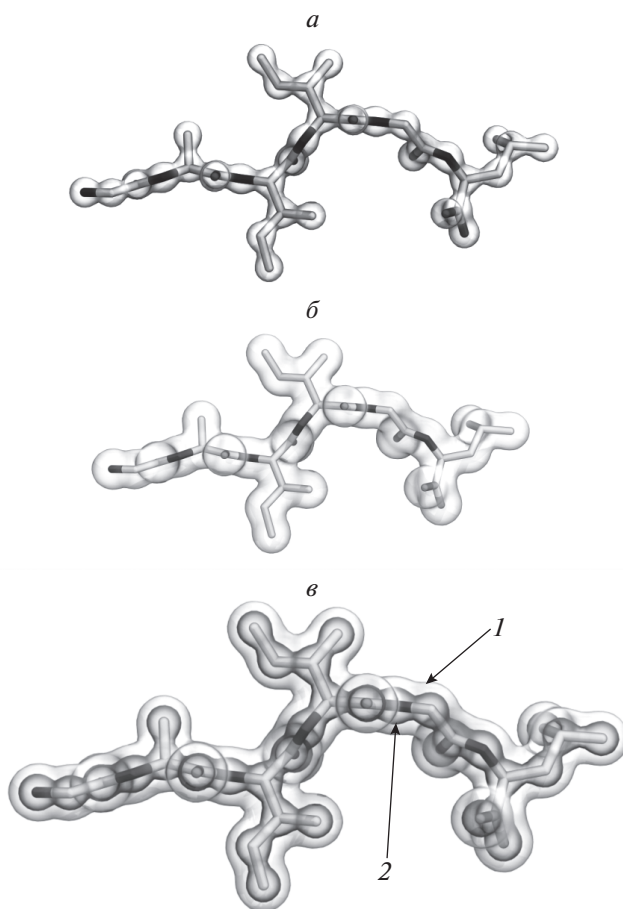


Рис. 8. *а* — Влияние электронного повреждения на карты электронной плотности. Показана карта с учетом повреждения для модельного пептида (PDB-код — 3PZZ). *б* — То же без учета повреждения. *в* — Наложение двух карт: 1 — без учета повреждения, 2 — с учетом повреждения. Карты изображены в виде поверхностей уровня 0.36 электронов/Å³.

плотности было минимальным. Результат показан на рис. 9*а*. Видно, что после масштабирования плотности становятся довольно близкими.

Однако имеется следующее отличие: в случае повреждения разница между плотностями в атомах С, N, O меньше, чем в его отсутствие. Этот же эффект можно наблюдать на рис. 9*б*, где изображены одномерные профили электронной плотности этих атомов. Другой иллюстрацией этого являются близкие средние за время импульса значения числа электронов в каждом из этих атомов, составляющие 1.22, 1.38 и 1.23 соответственно (таким образом, в среднем за импульс кислород имеет меньшее число электронов, чем азот). Заметим, что идентификация атомов С, N, O затруднена даже в традиционном рентгеноструктурном анализе ввиду близости их распределений электронной плотности (сплошная линия на рис. 9*б*). Это приводит, в частности, к проблеме выбора одной из двух ориентаций концевых групп остатков глутамина и аспарагина, а также ориентации цикла гистидина [26–29]. Считается, что эти ори-

ентации более или менее надежно определены в структурах высокого разрешения (1.5 Å и выше) [28], хотя даже при разрешении выше 0.9 Å в структурах из банка PDB были обнаружены неправдоподобные ориентации [28], а в исследовании [26] среди 100 структур количество таких ошибок не имело значимой корреляции с разрешением. Поэтому в большинстве случаев эти ориентации определяются из дополнительных соображений (водородные связи, близкие контакты атомов) с помощью разработанных для этого алгоритмов [27, 28, 30, 31]. В случае нестационарного рассеяния этот подход становится фактически единственно возможным.

Таким образом, после применения масштабирования коэффициента распределение электронной плотности в условиях нестационарного рассеяния становится качественно близким к распределению электронной плотности в неповрежденном объекте. Операция построения начальной атомной модели, в которой используются эти распределения, сама по себе имеет ограни-

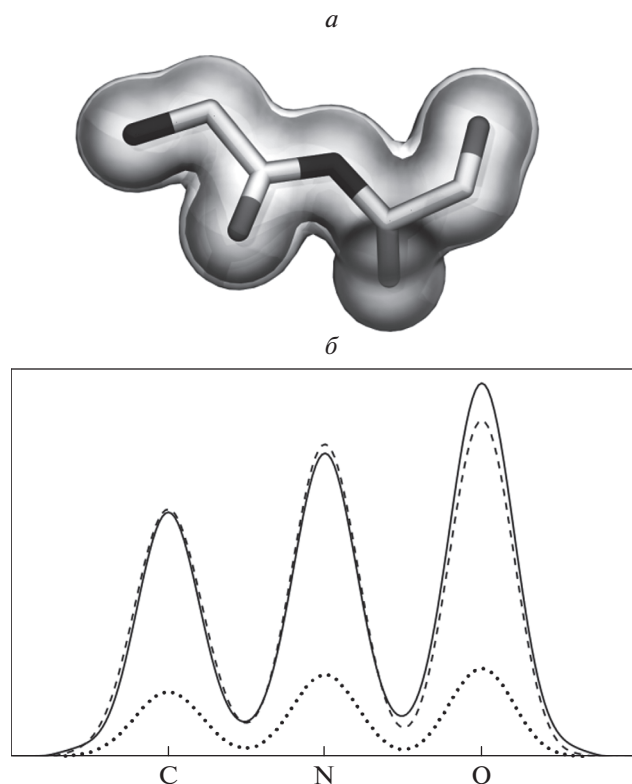


Рис. 9. *a* — Сравнение карт электронной плотности для отдельной аминокислоты: внешняя поверхность — без учета повреждения, внутренняя поверхность — с учетом повреждения и умножением на оптимальный масштабирующий коэффициент. Карты изображены в виде поверхностей уровня $0.36 \text{ электронов/\AA}^3$. *б* — Влияние электронного повреждения на распределение электронной плотности в атомах. Показаны ее одномерные профили для атомов C, N, O: сплошная кривая — профиль, рассчитанный без учета повреждения, точечная — с учетом повреждения (вычисленный по эффективным формфакторам), штриховая — профиль с учетом повреждения, умноженный на оптимальный масштабный коэффициент.

ченную точность как при ручном, так и при автоматическом ее выполнении ввиду приближенности процедур вписывания отдельных элементов молекулы и ограниченной точности фазовой информации. Поэтому поправки, связанные с электронным повреждением, не должны представлять существенных затруднений при построении атомных моделей.

6. УТОЧНЕНИЕ АТОМНЫХ МОДЕЛЕЙ ПО ДАННЫМ НЕСТАЦИОНАРНОГО РАССЕЯНИЯ

На этапе уточнения атомной модели ее трехмерная структура оптимизируется для наиболее близкого соответствия дифракционным данным. Точность финальной модели определяется именно на этой стадии и не зависит от погрешностей

вычислений на предыдущих этапах обработки данных (решение фазовой проблемы, первичное построение модели), если глобальный экстремум оценочной функции был правильно локализован. Ввиду этого именно при уточнении модели важен как можно более точный учет всех факторов, влияющих на рассеяние, и в частности процессов электронного повреждения. Точность учета эффекта повреждения зависит от точности имеющихся данных о параметрах рентгеновских импульсов. Так, при наличии достоверных данных об интенсивности, длительности и временной форме импульса погрешность расчета нестационарного рассеяния минимальна и определяется только точностью алгоритма (1), (2). Однако на практике параметры импульсов подвержены флуктуациям и обычно известны лишь приближенно, что приводит к падению точности итоговой модели. В худшем случае, как показано в работе [23], при плотности потока $10^8 \text{ фотонов/\AA}^2$ полное игнорирование повреждения приведет к ошибке порядка 8–9%. Если же повреждение учитывается, но используются приближенные оценки параметров, то можно ожидать промежуточных значений ошибок в пределах этой величины.

7. ЗАКЛЮЧЕНИЕ

Серийная фемтосекундная кристаллография с использованием РЛСЭ на основе принципа дифракции до разрушения позволила расширить возможности рентгеноструктурного анализа. Стали пригодными для получения околоатомного разрешения кристаллы размером в несколько микрон, облегчилась работа с особо чувствительными к радиационному повреждению объектами. В то же время для того, чтобы радикально превзойти результаты кристаллографии на современных синхротронах, эта методика требует дальнейшего развития для достижения возможности измерять дифракционные данные атомного разрешения, используя наноразмерные кристаллы крупных, плохо кристаллизующихся макромолекул. Один из основных факторов, приближающих эту цель, — дальнейшее повышение интенсивности импульсов, генерируемых РЛСЭ, возможность которого подтверждается теоретическими исследованиями [15]. В настоящей работе проведен анализ влияния этого и других существенных параметров на разрешение получаемых дифракционных картин. Работа со сверхвысокими интенсивностями, однако, приводит к особому режиму дифракционного эксперимента — нестационарному рассеянию, при котором регистрируется интегральная по времени картина рассеяния от образца, постоянно теряющего электроны вследствие процессов радиационного повреждения. Как показывают проведенные модельные расчеты, искажения, связанные с нестационарностью, не

являются препятствиями для расшифровки трехмерной атомной структуры молекул. В то же время необходимо принятие определенных мер по учету этих особенностей — подбор масштабирующего коэффициента; по возможности использование фазовых данных, полученных в аналогичных условиях; учет стирания различий между углеродоподобными атомами; учет нестационарности при решении прямой задачи рассеяния в ходе уточнения модели.

Таким образом, нанокристаллография с использованием сверхмощных импульсов РЛСЭ в режиме нестационарного рассеяния имеет хорошие шансы на осуществление и представляет большие перспективы для определения структуры крупных, плохо кристаллизующихся белков и макромолекулярных комплексов с высоким разрешением.

Авторы выражают благодарность В.Ю. Лунину за полезные обсуждения.

Работа поддержана грантом Российского фонда фундаментальных исследований № 14-04-31608.

СПИСОК ЛИТЕРАТУРЫ

1. *Smith J.L., Fischetti R.F., Yamamoto M.* // Curr. Opin. Struct. Biol. 2012. V. 22. № 5. P. 602.
2. *Emma P., Akre R., Arthur J. et al.* // Nat. Photonics. 2010. V. 4. № 9. P. 641.
3. *Pile D.* // Ibid. 2011. V. 5. № 8. P. 456.
4. *Fromme P.* // Nat. Chem. Biol. 2015. V. 11. № 12. P. 895.
5. *Neutze R., Wouts R., van der Spoel D. et al.* // Nature. 2000. V. 406. № 6797. P. 752.
6. *Крупянский Ю.Ф., Балабаев Н.К., Петрова Т.Е. и др.* // Хим. физика. 2014. Т. 33. № 7. С. 7.
7. *Синицын Д.О., Лунин В.Ю., Грум-Гржимайло А.Н. и др.* // Хим. физика. 2014. Т. 33. № 7. С. 21.
8. *Chapman H.N., Fromme P., Barty A. et al.* // Nature. 2011. V. 470. № 7332. P. 73.
9. *Panneels V., Wu W., Tsai C.-J. et al.* // Struct. Dyn. 2015. V. 2. № 4. P. 041718.
10. *Liu W., Wacker D., Gati C. et al.* // Science. 2013. V. 342. № 6165. P. 1521.
11. *Johansson L.C., Arnlund D., Katona G. et al.* // Nat. Commun. 2013. V. 4. P. 2911.
12. *Fenalti G., Zatsepin N.A., Betti C. et al.* // Nat. Struct. Mol. Biol. 2015. V. 22. № 3. P. 265.
13. *Kupitz C., Basu S., Grotjohann I. et al.* // Nature. 2014. V. 513. № 7517. P. 261.
14. *Neutze R.* // Philos. Trans. Roy. Soc. London, Ser. B. 2014. V. 369. № 1647. P. 20130318.
15. *Geloni G., Kocharyan V., Saldin E.* <http://arxiv.org/abs/1007.2743>
16. *Son S.-K., Young L., Santra R.* // Phys. Rev. A. 2011. V. 83. № 3. P. 033402.
17. *Caleman C., Huldt G., Maia F.R.N.C. et al.* // Amer. Chem. Soc. Nano. 2011. V. 5. № 1. P. 139.
18. *Lomb L., Barends T.R.M., Kassemeyer S. et al.* // Phys. Rev. B. 2011. V. 84. № 21. P. 214111.
19. *Barty A., Caleman C., Aquila A. et al.* // Nat. Photonics. 2012. V. 6. № 1. P. 35.
20. *Lorenz U., Kabachnik N.M., Weckert E. et al.* // Phys. Rev. E. 2012. V. 86. № 5. Pt 1. P. 051911.
21. *Лунин В.Ю., Грум-Гржимайло А.Н., Грызлова Е.В. и др.* // Мат. биол. и биоинформатика. 2013. V. 8. № 1. P. 93.
22. *Синицын Д.О., Лунин В.Ю., Грум-Гржимайло А.Н. и др.* // Наноструктуры. Мат. физика и моделирование. 2013. V. 9. № 2. P. 5.
23. *Lunin V.Y., Grum-Grzhimailo A.N., Gryzlova E.V. et al.* // Acta Crystallogr., Sect. D: Biol. Crystallogr. 2015. V. 71. Pt. 2. P. 293.
24. *Quiney H.M., Nugent K.A.* // Nat. Phys. 2010. V. 7. № 2. P. 142.
25. *Curwood E.K., Quiney H.M., Nugent K.A.* // Phys. Rev. A. 2013. V. 87. № 5. P. 053407.
26. *Word J.M., Lovell S.C., Richardson J.S. et al.* // J. Mol. Biol. 1999. V. 285. № 4. P. 1735.
27. *Weichenberger C.X., Sippl M.J.* // Bioinformatics. 2006. V. 22. № 11. P. 1397.
28. *Weichenberger C.X., Sippl M.J.* // Nucleic Acids Res. 2007. V. 35. Suppl. 2. P. 403.
29. *Lovell S.C., Word J.M., Richardson J.S. et al.* // Proc. Natl. Acad. Sci. U. S. A. 1999. V. 96. № 2. P. 400.
30. *Chen V.B., Arendall W.B., Headd J.J. et al.* // Acta Crystallogr., Sect. D: Biol. Crystallogr. 2010. V. 66. Pt. 1. P. 12.
31. *Hoof R.W.W., Vriend G., Sander C. et al.* // Nature. 1996. V. 381. № 6580. P. 272.