МОДЕЛИРОВАНИЕ ФОРМЫ ШАПЕРОНИНОВ ПО КРИВЫМ МАЛОУГЛОВОГО РЕНТГЕНОВСКОГО РАССЕЯНИЯ С ИСПОЛЬЗОВАНИЕМ ФОРМФАКТОРА ТОРА

С. В. Амарантов^а^{*}, И. Н. Налётова^b, Л. П. Курочкина^c

^а Институт кристаллографии им. А. В. Шубникова Российской академии наук 119333, Москва, Россия

^b Институт физико-химической биологии им. А. Н. Белозерского, Московский государственный университет им. М. В. Ломоносова 119991, Москва, Россия

^с Институт биоорганической химии им. М. М. Шемякина и Ю. А. Овчинникова Российской академии наук 117997, Москва, Россия

Поступила в редакцию 22 октября 2010 г.

Для белковых комплексов, поверхности которых можно описать набором простейших двусвязных форм, получено решение обратной задачи рассеяния в однородном приближении (рассеивающий потенциал внутри объема молекулы — постоянная величина). Для эксперимента были взяты растворы двух белков: хорошо известного бактериального шаперонина GroEL и малоизученного фагового шаперонина — продукта гена 146 (gp146). По экспериментальным кривым рассеяния было продемонстрировано устойчивое восстановление формы белковых комплексов. При восстановлении формы этих частиц использовались метод оболочек, метод вращения аминокислотных последовательностей с использованием формфактора аминокислоты, метод поиска параметров модели белкового комплекса с предварительно полученным формфактором модели.

1. ВВЕДЕНИЕ

Малоугловое рассеяние (МУР) — упругое рассеяние электромагнитного излучения или пучка частиц (нейтронов, электронов) на неоднородностях вещества, размеры которых существенно превышают длину волны излучения (или дебройлевскую длину волны частиц). При этом направления рассеянных лучей лишь незначительно (на малые углы) отклоняются от направления падающего луча. В малоугловых структурных исследованиях используют, как правило, рентгеновское излучение или тепловые нейтроны с длиной волны 1–10 Å (10⁻¹–1 нм). С их помощью изучают неоднородности коллоидных размеров (10–10⁴ Å). В отличие от других дифракционных методов с помощью МУР исследуют структуру разупорядоченных объектов, в частности, строение биологических молекул в растворе. В этом случае об-

разец представляет собой раствор идентичных частиц (монодисперсную систему), а экспериментальная интенсивность рассеяния пропорциональна произведению концентрации частиц в растворе и интенсивности рассеяния от одной частицы. Метод малоуглового рентгеновского рассеяния позволяет определять размер и форму белковых молекул и комплексов в растворе в их естественном состоянии, а также следить за любыми изменениями этих параметров под воздействием внешних факторов [1]. Интерпретация экспериментальных данных в этом случае заключается в создании модели низкого разрешения структуры рассеивающей частицы. В данном случае это модель трехмерной формы частицы белкового комплекса в однородном приближении. Построение такой модели осложняется потерей информации при сферическом усреднении в отличие от рентгеноструктурного анализа монокристаллов. Поэтому оптимальное использование информации, заключенной в экспериментальных данных, а так-

^{*}E-mail: amarantov_s@mail.ru

же привлечение дополнительных характеристик, например, наличия оси симметрии у частицы, ее вытянутости или сплюснутости, оказывается важным при обработке экспериментальных данных.

Несмотря на заметный прогресс в области компьютерного моделирования для обработки данных малоуглового эксперимента, поиск оптимальной параметризации модели в данном конкретном случае остается актуальным. В рамках однородного по плотности приближения форму рассеивающей молекулы задает ее потенциал рассеяния. В случае упругого рассеяния потенциал рассеяния с точностью до материального множителя пропорционален формфактору рассеивающей молекулы. При этом число используемых формфакторов для описания форм частиц ограничивается приведенными в классических работах Гинье [2]. Широко известны аналитические выражения для формфакторов геометрических фигур, таких как шар, эллипсоид, диск, стержень. Общая таблица этих формфакторов представлена в работе [3]. Однако только этих данных оказывается недостаточно для моделирования более сложных комплексов, имеющих топологически двусвязную поверхность. В связи с изучением подобных структур ранее одним из авторов этой работы были получены аналитические выражения для малоугловых формфакторов для простейшей двусвязной поверхности — тора [4]. Данное исследование можно рассматривать как экспериментальное приложение указанной выше работы. В частности, целью нашей работы является восстановление формы белковых комплексов по кривой интенсивности рентгеновского МУР. Одним из методов решения обратной задачи рассеяния будет поиск параметров модели для полученного в указанной выше работе формфактора двойного эллиптического тора. В качестве объектов исследования взяты различные белки-шаперонины, которые, как известно, являются олигомерными белками, имеющими сложную архитектуру: они состоят из двух колец, уложенных стопкой, и имеют центральный канал.

Шаперонины: характеристика и выделение рекомбинантных белков для измерения методом рентгеновского МУР. Шаперонины (белки теплового шока) представляют собой класс молекулярных шаперонинов, которые помогают правильно сворачиваться как новосинтезированным полипептидным цепям, так и денатурированным под воздействием стресса белкам [5]. Они являются сложными гомо- или гетероолигомерными комплексами, в формировании которых участвуют 14–18

субъединиц белков массой около 60 кДа. Наиболее изученным на сегодняшний день шаперонином является GroEL Escherichia coli (E.coli), который обычно функционирует в паре с другим белком ко-шаперонином GroES [6-8], но также может работать и один. Шаперонин GroEL состоит из 14 идентичных субъединиц (57 кДа), формирующих два уложенных стопкой гептамерных кольца [8,9]. Субъединицы GroES (10 кДа) также образуют гептамерное кольцо, которое играет роль «крышки» и прикрывает сверху центральный канал шаперонина при образовании комплекса двух белков [10, 11]. Одна из основных функций комплекса GroEL/GroES заключается в связывании развернутых полипептидных цепей, для того чтобы предотвратить их неспецифическую ассоциацию и обеспечить дальнейшее правильное сворачивание белка внутри канала [11, 12]. В нашей работе форма шаперонина GroEL E. coli (далее просто GroEL) в растворе будет сравниваться с формой белка в кристаллическом состоянии как в комплексе с GroES, так и без него.

Другим объектом нашего исследования является малоизученный вирусный (фаговый) шаперонин, кодируемый геном 146 бактериофага EL Pseudomonas aeruginosa [13]. На основании сравнительного анализа первичной структуры продукта гена 146 (gene product, gp146) с аминокислотными последовательностями бактериальных GroEL данный белок был отнесен к классу шаперонинов. На сегодняшний день это единственный из известных белков-шаперонинов, кодируемый вирусным геномом. Согласно данным электронной микроскопии он является гомоолигомерным комплексом, состоящим из 14 субъединиц gp146 (61.7 кДа), которые формируют два стопкой уложенных гептамерных кольца, и имеет внутренний канал [13]. Поскольку этот белок имеет характерную для шаперонинов структуру и размеры, близкие к GroEL, для описания его формы в растворе мы можем воспользоваться той же моделью, что ранее была предложена для GroEL [4]: для обоих белков (бактериального GroEL и вирусного шаперонина) будет вычислен формфактор идеализированной модели в виде двойного эллиптического тора как функция трех геометрических параметров и модуля вектора рассеяния с дальнейшим поиском этих параметров.

Как и все шаперонины, вирусный шаперонин функционирует АДФ-зависимым способом (АДФ-аденозиндифосфат). Поскольку в отличие от бактериального GroEL вирусный шаперонин функционирует без ко-фактора, он может находиться в

11 ЖЭТФ, вып. 2 (8)



Рис.1. *a*) Треугольник упругого рассеяния на тройке волновых векторов \mathbf{k}_0 , \mathbf{k}_1 , \mathbf{s} для обратного пространства и для переданного импульса $\Delta \mathbf{p}$ в прямом пространстве $\mathbf{p} = \hbar \mathbf{k}$. δ) Сканирование по углу при постоянной энергии фотонов E = const ($\lambda = \text{const}$), представлены положения трех векторов рассеяния $\mathbf{s}_1(\theta_1)$, $\mathbf{s}_2(\theta_2)$, $\mathbf{s}_2(\theta_3)$, соответствующих трем углам рассеяния

разных конформациях: открытой и закрытой. Мы впервые проводим опыты по выяснению конформации вирусного шаперонина в растворе в разных условиях методом рентгеновского МУР, а также сравниваем молекулярные структурные параметры (радиус инерции и максимальный размер комплекса) в разных конформационных состояниях белка, а также с GroEL.

Вирусный шаперонин был выделен из клеток *E.coli* штамма BL21(DE3) (Novagen, CША), продуцировавших рекомбинантный ген gp146, и очищен хроматографией на Q-сефарозе. Фракции, содержавшие чистый белок, концентрировали с помощью мембранных фильтров Amicon Ultra-15, задерживающих белки с массой более 100 кДа (molecular-weight cutoff (MWCO) 100,000, Millipore, CIIIA). Гомогенность полученного препарата определяли с помощью аналитического ультрацентрифугирования. Концентрацию белка определяли спектрофотометрически по поглощению при длине волны 280 нм, используя теоретически рассчитанный коэффициент экстинкции $35.870 \text{ моль}^{-1} \cdot \text{см}^{-1}$. Молекулярную массу фагового шаперонина считали равной 863.8 кДа (субъединица — 61.7 кДа).

Выделение шаперонина GroEL и ко-шаперонина GroES из клеток *E.coli* штамма W3110, трансформированных плазмидой pOF39, содержавшей гены, кодирующие GroEL и GroES, проводили по методике Корралеса и Фершта [14] с описанными нами ранее модификациями [15]. Концентрацию белка определяли спектрофотометрически по поглощению при 280 нм, принимая коэффициент поглощения $A_{280}^{0.1\%}$ для GroEL равным 0.18 [16]. Молекулярную массу GroEL₁₄ считали равной 840 кДа (субъединица — 60 кДа).

2. МЕТОД МАЛОУГЛОВОГО РАССЕЯНИЯ

2.1. Упругое рассеяние

Пусть на разбавленную систему монодисперсных однородных частиц падает монохроматическое рентгеновское излучение или поток нейтронов. Для изучения размеров частиц, заключенных в диапазоне от 1 до 100 нм, обычно используются длины волн $0.5 \text{ \AA} \le \lambda \le 6 \text{ \AA}$.

Обозначим волновые векторы падающего и рассеянного излучения, образующие треугольник рассеяния, как \mathbf{k}_0 и \mathbf{k}_1 (рис. 1). В эксперименте с рентгеновским источником будем считать рассеяние падающих на образец фотонов или нейтронов абсолютно упругим. В этом случае $|\mathbf{k}_0| = |\mathbf{k}_1|$, а разность этих векторов определяет вектор рассеяния $\mathbf{s} = \mathbf{k}_0 - \mathbf{k}_1$, модуль вектора рассеяния равен (см. рис. 1)

$$s = 2|\mathbf{k}_0| = \frac{4\pi}{\lambda} \sin\theta,$$

де θ — половина угла рассеяния, λ — длина волны используемого излучения, в нашем случае λ = const.

Треугольник рассеяния в обратном пространстве определяется волновыми векторами падающего и рассеянного излучения фотона или нейтрона. В прямом пространстве ему соответствует подобный треугольник с тем же углом рассеяния 2θ , составленный из волновых векторов (импульсов) фотона или нейтрона до и после рассеяния, $\mathbf{k}_0(\mathbf{p}_0)$ и $\mathbf{k}_1(\mathbf{p}_1)$ с разностью, определяемой вектором рассеяния **s** в обратном пространстве и переданным импульсом $\Delta \mathbf{p}$ в прямом пространстве. На рис. 1a представлен треугольник рассеяния AOB для одного фиксированного угла рассеяния 2θ , а на рис. 1b показаны три треугольника рассеяния AOB_i , i = 1, 2, 3, лежащие в одной плоскости и имеющие общую вершину в точке *О*. Поскольку рассеяние упругое, все треугольники равнобедренные (см. рис. 16).

Целью нашей задачи является восстановление трехмерной формы белка по скалярной интенсивности МУР от системы типа «разбавленного раствора» (межчастичная интерференция отсутствует). Возможность сведения трехмерной задачи к одномерной рассмотрена в работе [17].

В основе используемого нами метода лежит первое борновское приближение [18], которое можно использовать, если выполнены условия $L \gg D_{max}$ и $D_{max}^2/L\lambda \ll 1$. В нашем случае оно применимо, поскольку в классической схеме малоуглового эксперимента расстояние L между образцом и детектором много больше длины волны λ и максимального размера рассеивающих объектов D_{max} . Полагая $D_{max} \approx 100$ нм, $L \approx 0.3$ м, $\lambda \approx 1.5 \cdot 10^{-10}$ м, получим $D_{max}^2/L\lambda \leq 10^{-4}$. Пользуясь первым борновским приближением, для интенсивности I(s) упругого рассеяния можно записать

$$I(s) = I_e \overline{N} A(s, \Omega) A^*(s, \Omega), \qquad (2.1)$$

где I_e — «материальный множитель», который в области углов 0° $\leq \theta < 5$ ° есть постоянная величина, \overline{N} — среднее число идентичных частиц в растворе, A — амплитуда рассеянной волны, знак «*» означает комплексное сопряжение, Ω — проекция телесного угла в плоскости детектирования рассеянного излучения. Поскольку в эксперименте регистрация рассеянного излучения проводится в относительных единицах, интенсивность рассеяния определена с точностью до постоянной.

2.2. Амплитуда рассеяния, формфактор частицы

Согласно первому борновскому приближению, амплитуда $A(s, \Omega)$ рассеянной сферической волны и рассеивающий потенциал $U(\mathbf{r})$ связаны преобразованием Фурье:

$$A(s,\Omega) = \frac{n_e}{V'} \int_{V'} U(\mathbf{r}) \exp(i\mathbf{s} \cdot \mathbf{r}) \, dV', \qquad (2.2)$$

 $\overline{\rho}_{e}(\mathbf{r}) = (n_{e}/V')U(\mathbf{r})$ — средняя электронная плотность рассеивающей частицы, V' — ее объем, n_{e} — число электронов одной частицы, участвующих в рассеянии.

Обозначим фурье-амплитуду от одной частицы $(\overline{N}=1)$ как

$$\Phi(s,\Omega) = \int_{V'} \overline{\rho}_e(\mathbf{r}) \exp(i\mathbf{s} \cdot \mathbf{r}) \, dV'. \tag{2.3}$$

Возводя амплитуду $\Phi(s, \Omega)$ в квадрат и усредняя по всевозможным ориентациям частицы в растворе, для интенсивности рассеяния от одной модели частицы, или для формфактора частицы $\langle \Phi^2(s) \rangle$ получим выражение

$$I_{mod}(s) = \langle \Phi^2(s) \rangle = \int_{\Omega} \Phi^2(s,\theta) \, d\Omega, \qquad (2.4)$$

где элемент телесного угла в случае сферической системы координат имеет вид $d\Omega = d\theta \, d\varphi, \, \theta$ — полярный угол, $0 \leq \theta \leq \pi$ (он также определен как половина угла рассеяния), φ — азимутальный угол, $0 \leq \varphi \leq 2\pi$. Область рассматриваемых углов рассеяния в малоугловой схеме ограничивается диапазоном $0 \leq \theta \leq 5^{\circ}$ (см. рис. 1).

В табл. 1 приведены основные формфакторы, используемые далее [3, 4].

2.3. Расчет функций распределения по расстояниям, максимального размера частицы и ее радиуса инерции

В рамках однородного по электронной плотности приближения можно по кривой МУР найти радиус инерции частицы как геометрический инвариант; он аналогичен механическому радиусу инерции R_g , вычисленному через функцию парных расстояний p(r)[2, 3]:

$$R_{g}^{2} = \frac{\int\limits_{V'}^{} \overline{\rho}_{e} r^{2} dV'}{\int\limits_{V'}^{} \overline{\rho}_{e} dV'} = \frac{\int\limits_{0}^{D_{max}} p(r) r^{2} dr}{2 \int\limits_{0}^{D_{max}} p(r) dr}.$$
 (2.5)

В отличие от традиционного подхода, когда делается оценка величины R_g^2 по тангенсу угла наклона начального участка кривой рассеяния в координатах Гинье (ln I и s²), вычисление по интегральному соотношению (2.5) устраняет произвол при выборе начального участка кривой рассеяния. Это даст более точную оценку радиуса инерции, поскольку при вычислении R_g через функцию p(r) используется вся кривая рассеяния.

В работах [2,3] получена связь интенсивности рассеяния и функции парных расстояний p(r):

$$I(s) = \frac{1}{V'} \int_{V'} \langle \gamma_0(\mathbf{r}) \rangle \exp(i\mathbf{s} \cdot \mathbf{r}) r^2 d\mathbf{r} \, d\Omega =$$
$$= \frac{1}{V'} \int_{V'} \langle p(\mathbf{r}) \rangle \exp(i\mathbf{s} \cdot \mathbf{r}) \, d\mathbf{r} \, d\Omega, \quad (2.6)$$

11*

Тело вращения	Φ ормфактор $\langle \Phi^2(s) angle$	Асимптотика Порода $\langle \Phi^2(s) angle _{sR_g \gg 1}$
Шар радиуса R_0 $x^2 + y^2 + z^2 \le R_0^2$	$\left[3\frac{\sin(sD_{max}) - sr\cos(sD_{max})}{(sD_{max})^3}\right]^2$	$\frac{9}{(sR_0)^4} \propto \frac{1}{s^4}$
Полый цилиндр высотой h $r_1^2 \le x^2 + y^2 \le r_2^2,$ $0 \le z \le h = 2L$	$\frac{4}{(r_{max}^2 - r_{min}^2)^2} \int_{0}^{\pi/2} \times \left[\left(r_2^2 D(sr_2, \theta) - r_1^2 D(sr_1, \theta) \right) \frac{\sin(Ls\cos\theta)}{Ls\cos\theta} \right]^2 \times \sin\theta d\theta, D(x, \theta) = J_1(x\sin\theta)/x\sin\theta$	$\propto rac{1}{s^3}$
Top $V' \in \frac{(r-b)^2}{a^2} + \frac{z^2}{c^2} \le 1,$ $b-a \le r \le b+a$	$\left(\frac{2}{\pi abcs}\right)^{2} \times \\ \times \int_{0}^{\pi/2} \frac{1}{\cos^{2}\theta} \left\{ \int_{b-a}^{b+a} \sin\left[sc\sqrt{1-\frac{(r-b)^{2}}{a^{2}}\cos\theta}\right] \times \\ \times J_{0}(sr\sin\theta)rdr \right\}^{2} \sin\thetad\theta$	$\propto \frac{4}{bs(as)^2(cs)^3}$

Таблица 1. Основные формфакторы, используемые в данной работе

Примечание: $J_{\alpha}(z) - функция Бесселя первого рода порядка <math>\alpha$.

где $r \in V'$, а пространственно-усредненная автокорреляционная функция $\langle \gamma_0(r) \rangle$ связана с p(r) соотношением $p(r) \equiv \langle \gamma_0(\mathbf{r}) \rangle r^2$. Для нее выполняется условие нормировки

$$4\pi \int_{0}^{D_{max}} r^2 \gamma_0(r) \, dr \equiv 4\pi \int_{0}^{D_{max}} p(r) \, dr = 1.$$
 (2.7)

В частности, для разбавленной системы идентичных невзаимодействующих частиц, например раствора белковых комплексов, выражение для интенсивности рассеяния, приходящейся на один комплекс, можно записать в виде [2]

$$I(s) = \frac{1}{V'} \int_{0}^{2\pi} \int_{0}^{D_{max}} \int_{0}^{\pi} p(r) \exp(isr\cos\theta) \times \\ \times \sin\theta \, d\theta \, dr \, d\varphi = \frac{4\pi}{V'} \int_{0}^{D_{max}} p(r) \frac{\sin(sr)}{sr} \, dr. \quad (2.8)$$

Здесь величина D_{max} — максимальный размер молекулы — также является неизвестной величиной. Соотношение между интенсивностью рассеяния I(s) и функцией распределения по расстояниям p(r) определяет связь между обратным и прямым пространствами, выражаемую через преобразование Фурье для следующих трансформант:

$$sI(s) \xrightarrow{F} p(r), \quad p(r) \xrightarrow{F} \frac{2}{\pi} sI(s), \quad (2.9)$$

или в развернутой форме:

$$\frac{V'}{4\pi}sI(s) = \int_{0}^{\infty} p(r)\frac{\sin(sr)}{r} dr,$$

$$p(r) = \frac{V'}{4\pi^2} \int_{0}^{\infty} I(s)\frac{\sin(sr)}{sr} s^2 ds.$$
(2.10)

Таким образом, соотношение между интенсивностью рассеяния I(s) и функцией p(r) для частицы, имеющей максимальный размер D_{max} , связано через интегральное уравнение Фредгольма первого рода

$$I(s) = \frac{4\pi}{V'} \int_{0}^{D_{max}} p(r) \frac{\sin(sr)}{sr} dr \equiv$$
$$\equiv C \int_{0}^{D_{max}} p(r) K(s, r) dr \quad (2.11)$$

(C -постоянная, K(s, r) -ядро интегрального преобразования), или, в операторном виде,

$$f = \hat{A}y,$$

где y определяет неизвестную функцию y = p(r), \hat{A} — интегральный оператор с ядром K(s, r).

Переход из обратного пространства в прямое и вычисление функции распределения по расстояниям внутри частицы, p(r), в данной работе осуществляли с помощью программы косвенного фурье-преобразования, в которой использован метод α -регуляризации Тихонова [19]. В этом методе ставятся два условия: минимизация невязки и минимизация нормы решения:

$$||\hat{A}y - f||^2 = \min \mathcal{T}[y, f] \cap ||y|| = \min[y, f], \quad (2.12)$$

что эквивалентно задаче минимизации сглаживающего функционала $\mathcal{T}[y, f]$:

$$||\hat{A}y - f||^2 + \alpha ||y||^2 = \min_{y} \mathcal{T}_{\alpha}[y, f].$$
(2.13)

С этой целью полученная из эксперимента на рентгеновской малоугловой установке экспериментальная кривая рассеяния после операции «редукции к идеальному прибору» согласно выражению (2.11) подвергается фурье-преобразованию с ядром $\sin(sr)/sr$ в пространстве Соболева $W_2^1[0, D_{max}]$ с нормой

$$|y|| = \begin{cases} \int_{0}^{D_{max}} y^{2}(x) \, dx + \alpha \int_{0}^{D_{max}} y'^{2}(x) \, dx \end{cases}^{1/2}.$$
 (2.14)

Фурье-образом функции интенсивности малоуглового рассеяния I(s) в «прямом пространстве» является внутричастичная функция распределения парных расстояний p(r),

$$sI(s) \xrightarrow{F} p(r),$$

причем $p(0) = p(D_{max}) = 0$, что позволяет найти отрезок максимальной длины, концы которого расположены внутри молекулы. Регуляризованный метод наименьших квадратов (РМНК) по Тихонову при оптимальном выборе параметра α дает гладкое и устойчивое решение.

Относительно искомого решения для p(r) делали предположение, что при $r \in [0, D_{max}]$ оно должно иметь внутри этого интервала производную p'(r), интегрируемую с квадратом. При этом

$$p(0) = 0, \quad p'(0) = 0, \quad p(D_{max}) = 0, \quad p'(D_{max}) = 0.$$

Эти равенства определяют для нашей задачи «множество корректности», существенно сужающее

класс возможных решений. В данной программной реализации [20] функционал Тихонова, согласно выбранной норме (2.14), имеет вид

$$\mathcal{T}_{\alpha}\left[p(r), I(s)\right] = \frac{1}{N-1} \times \\ \times \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \left[I(s_i) - \int_{0}^{D_{max}} p(r) K(s_i, r) dr \right]^2 + \\ + \alpha \int_{0}^{D_{max}} \left[\frac{dp(r)}{dr} \right]^2 dr, \quad \alpha > 0, \quad (2.15)$$

где σ_i — стандартное отклонение интенсивности рассеяния в данной экспериментальной точке s_i. Таким образом, метод регуляризации Тихонова в рассматриваемом случае сводится к минимизации сильновыпуклого функционала вида (2.13) на выпуклом замкнутом множестве $D \in W_2^1[0, D_{max}]$, так что экстремаль $\min \mathcal{T}_{\alpha} \in D$ функционала существует и единственна при любом выборе параметра $\alpha > 0$. Если величина α мала, то решение неустойчиво, и при $\alpha \to 0$ метод регуляризации Тихонова переходит в метод наименьших квадратов $||\hat{A}y - f||^2 = \min \mathcal{T}[y, f]$ с крайне неустойчивым решением. С увеличением а решение становится глаже и устойчивей, т. е. уменьшается квадрат нормы решения $||y||^2$, но увеличивается невязка. При выборе параметра регуляризации а необходимо получить оптимальное сочетание достаточной гладкости при допустимой невязке.

Ограничиваясь максимально экспериментально возможным интервалом для модуля вектора рассеяния $s \in [s_{min}, s_{max}]$ и аппроксимируя в «нулевой угол рассеяния» $s_{min} \to 0$ мы осуществляли преобразование Фурье на ограниченном интервале как обратного $s \in [0, s_{max}]$, так и прямого $r \in [0, D_{max}]$ пространств, что допускается для функций с финитным спектром по теореме Котельникова-Шеннона [21].

2.4. Формула Дебая, используемая для вычисления кривых МУР

В рамках однородного приближения интенсивность МУР может быть вычислена по формуле Дебая для атомной структуры частицы или шариковой модели молекулы, заполняющей ее объем. В этом случае статическое или мгновенное значение результирующей амплитуды A(s) рассеянной волны для некоторого произвольного, дискретного набора рассеивателей с амплитудами $f_j^0(s)$ (так называемые атомные формфакторы) имеет вид

$$A(s) = \sum_{j=1}^{n} f_j^0(s) \exp(i\mathbf{s} \cdot \mathbf{r}_j), \qquad (2.16)$$

где *п* — число атомов в частице. Верхний индекс у функции $f_i^0(s)$ указывает на то, что динамические факторы́ мы не учитываем; поэтому $f_j^0(s)$ является действительной величиной, $\operatorname{Im} f_i^0(s) = 0$. Интенсивность рассеяния получим аналогично выражению (2.1) умножением амплитуды на комплексно-сопряженную величину. Поскольку молекулы в растворе не ориентированы по отношению к падающему лучу, полученную интенсивность следует усреднить по телесному углу. Таким образом, формула Дебая для интенсивности рассеяния от монодисперсной системы типа «разбавленного раствора» из \overline{N} (в среднем) частиц, каждая из которых состоит из *n* атомов, имеет вид

$$\langle I(s) \rangle_{\Omega} = \overline{N} I_e \sum_{i=1}^n \sum_{j=1}^n f_i^0(s) f_j^0(s) \langle \exp(i\mathbf{s} \cdot \mathbf{r}_{ij}) \rangle_{\Omega} =$$
$$= \overline{N} I_e \sum_{i=1}^n \sum_{j=1}^n f_i^0(s) f_j^0(s) \frac{\sin(sr_{ij})}{sr_{ij}}, \quad (2.17)$$

где материальный множитель I_e представляет собой томсоновское сечение (поскольку измерения проводились в относительных единицах, то, считая I(s)с точностью до постоянной, величину Іе можно не принимать во внимание); $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|, r_i$ и r_j — радиусы-векторы *i*-го и *j*-го атомов, а суммирование ведется по всем атомам одной частицы.

3. РЕШЕНИЕ ОБРАТНОЙ ЗАДАЧИ РАССЕЯНИЯ ПУТЕМ МОДЕЛИРОВАНИЯ В «ПРЯМОМ ПРОСТРАНСТВЕ»

Для задачи восстановления формы белка мы воспользуемся различными путями ее решения. Все они представлены на рис. 2. Первый метод поиска формы молекулы — метод функции оболочки. Найдем характерную форму частицы с рассеивающей плотностью $\overline{\rho}_{e}(\mathbf{r})$ в однородном приближении как суперпозицию сферических гармоник [22]:

$$\overline{\rho}_e(\mathbf{r}) = \begin{cases} 1, & 0 \le \mathbf{r} \le F(\Omega), \\ 0, & \mathbf{r} > F(\Omega), \end{cases}$$
(3.1)

где функция оболочки $F(\Omega)$ может быть разложена в ряд Фурье-Лапласа по ортогональной системе сферических функций $Y_{lm}(\Omega) \equiv Y_{lm}(\theta, \varphi)$ и, таким образом, определяется комплексными коэффициентами f_{lm} [3]:



Рис. 2. Демонстрация четырех различных алгоритмов восстановления формы белка по экспериментальной кривой рассеяния $I_{\exp(s)}$: a- в виде функции оболочки; б — в виде шариковой модели; в в виде модели аминокислотных остатков; г — через поиск параметров выбранной модели с рассчитанным для нее формфактором

$$F(\Omega) \approx F_{l_{max}}(\Omega) = \sum_{l=0}^{l_{max}} \sum_{m=-l}^{l} f_{lm} Y_{lm}(\Omega). \qquad (3.2)$$

Второй метод — метод шариков-атомов (Bead model) — использовался нами в предыдущей работе [4]. В нем реальную атомную систему мы заменили системой маленьких шаров атомного радиуса a с одинаковой амплитудой рассеяния $f_i^0(s)$ = $= f_i^0(s) \equiv f(s).$

. Третий используемый в нашей работе метод метод моделирования аминокислотными остатками (Dummy residues model). Для него с точностью до постоянной можно воспользоваться следующим выражением для амплитуды рассеяния одного аминокислотного остатка [23, 24]:

$$f_i^0(s) = G(s) \exp(s^2 V_i^{2/3}), \qquad (3.3)$$

где экспоненциальный множитель дает поправку на вытесненный молекулой объем растворителя $V_i = (4\pi/3) r_{wi}^3$ для *i*-го атома или группы атомов с радиусом r_{wi} , а G(s) — полный по одной атомной группе формфактор,

$$G(s) = (r_0/r_m)^3 \exp\left[-s^2(r_0^2 - r_m^2)\right], \qquad (3.4)$$

 r_m — средний действующий радиус для группы атомов (для атомов белковых комплексов r_m обычно приблизительно составляет 1.62 Å), r_0 — эффективный радиус, т.е. переменный параметр, используемый для учета «исключенного объема»; он мало отличается от r_m ($0.96r_m \leq r_0 \leq 1.04r_m$). В данной работе роль единичного элемента структуры с амплитудой рассеяния $f_i^0(s)$ будет играть аминокислотный остаток. Интенсивность рассеяния от такой модельной системы, согласно формуле Дебая (2.17) примет вид

$$I_{mod}(s) = \sum_{i=1}^{n} \sum_{j=1}^{n} f_i^{(0)}(s) f_j^{(0)}(s) \frac{\sin(sr_{ij})}{sr_{ij}}.$$
 (3.5)

Поскольку результирующая интенсивность по всем атомам или аминокислотным остаткам не содержит особенностей, присущих отдельным атомам, нет необходимости использовать точные выражения для формфакторов аминокислотных остатков. Для программной реализации формул (3.3)–(3.5) использована сферическая сетка. Для поиска пространственного расположения объемных элементов взят вариант программы решения нелинейной задачи наименьших квадратов с помощью глобальной минимизации, основанной на методе Монте-Карло с «имитацией отжига» в виде алгоритма Метрополиса. Метод Метрополиса представляет собой один из вариантов марковского порогового алгоритма глобального случайного поиска [25].

В процессе поиска программа минимизации инвертирует плотность случайным образом выбранного аминокислотного остатка, т. е. изменяет ее с плотности частицы $\overline{\rho}_e = 1$ на плотность растворителя, равную нулю, или наоборот. После каждой модификации программа рассчитывает кривую распределения интенсивности малоуглового рассеяния от текущей структуры по формуле Дебая и вычисляет взвешенное суммарное квадратичное отклонение модельной кривой рассеяния $I_{mod}(s)$, рассчитанной согласно формуле (3.4), от экспериментальной кривой $I_{exp}(s)$; расчет проводится по K узлам:

$$\chi^{2} = \sum_{j=1}^{K} \left[\frac{I_{exp}(s_{j}) - I_{mod}(s_{j})}{\sigma(s_{j})} \right]^{2}, \qquad (3.6)$$

где $\sigma(s_j)$ — оценка экспериментальной ошибки в *j*-й точке. Число узлов *K* на кривой рассеяния определяется на основании уже упоминавшейся в конце разд. 2.3 теоремы Котельникова – Шеннона о передаче сигнала без потерь [21]. Согласно этой теореме, количество узлов, определяемых выражением

$$K = 2\left[\frac{(s_{max} - s_{min})D_{max}}{2\pi}\right] + 1, \qquad (3.7)$$

оказывается достаточным для однозначного восстановления финитной функции по ее фурье-спектру. Здесь s_{min} и s_{max} — максимальное и минимальное значения модуля вектора рассеяния для $I_{exp}(s)$, [...] — целая часть числа.

К сумме (3.6) добавляют «штрафные» члены $w_D P_D$ и $w_L P_L$, которые соответствуют требованиям неразрывности структуры и отсутствию отдельно расположенных шаров, w_D и w_L весовые коэффициенты для функционала $\mathcal{F} = \chi^2 + w_D P_D + w_L P_L$. Штраф P_D вычисляют как отношение общего числа шаров структуры к числу шаров, составляющих наибольший домен, понимаемый как полученный в процессе моделирования цельный «кусок» моделируемого тела, т. е. образованный непосредственно контактирующими атомами:

$$\mathcal{F}[I(s)] = \sum_{j=1}^{K} \left[\frac{I_{exp}(s_j) - I_{mod}(s_j)}{\sigma(s_j)} \right]^2 + w_D P_D + w_L P_L. \quad (3.8)$$

Если значение целевой функции \mathcal{F} меньше предыдущего значения ($\Delta = \mathcal{F}_k - \mathcal{F}_{k-1} < 0, k$ номер шага), то структура запоминается. Процедура моделирования «отжига» состоит в том, что на каждом шаге в качестве «наилучшей» структуры с некоторой вероятностью программа может принимать ту, которая имеет худшее значение. Вероятность задается параметром «температуры» T: если $\Delta > 0$, структура принимается с вероятностью $\exp(-\Delta/T)$. Со временем, после каждых 100K шагов (или 10K шагов с $\Delta < 0$), параметр T уменьшается ($T_{K+1} = 0.9T_K$), что позволяет программе все реже запоминать неудачные структуры [26].

Необходимость процедуры отжига связана с тем, что целевая функция \mathcal{F} может иметь несколько локальных минимумов. Возможность принятия худшего шага в качестве решения помогает программе не останавливаться в промежуточных решениях. Финальную пространственную конфигурацию шаров рассматривают как найденную модель структуры молекулы в растворе. В силу нелинейного характера зависимости параметров структурной модели в общем случае невозможно аналитически оценить степень близости решения к глобальному минимуму, поэтому обычно проводится несколько независимых поисков решения для каждой кривой рассеяния.

Четвертый метод восстановления формы молекулы по экспериментальной кривой МУР требует



Рис. 3. Два основных подхода к решению обратной задачи рассеяния для глобулярных белков в однородном приближении

выбора подходящей модели, описывающей форму частицы, с последующим вычислением по формуле (2.3) формфактора модели и численного поиска параметров этой модели по кривой рассеяния. Совокупность перечисленных методов позволяет надежно восстановить форму белкового комплекса (см. рис. 2 и схему на рис. 3).

4. ЭКСПЕРИМЕНТ

Монодисперсность растворов белков — бактериального шаперонина GroEL и вирусного шаперонина gp146 — была предварительно проверена методом аналитического ультрацентрифугирования. Измерения коэффициента седиментации этим методом проводили при помощи аналитической ультрацентрифуги Beckman Spinco E, оснащенной фотоэлектрической сканирующей приставкой. Скорость вращения ротора составляла 56000 об/мин. Оптическую плотность определяли на длине волны источника излучения равной 280 нм. Пробы готовили непосредственно перед экспериментом. Исходный белок GroEL разводили 50 миллимольным К-фосфатным буфе-



Рис. 4. Фракционная зависимость концентрации c белка шаперонина GroEL (≈ 800 кДа) и наличия примесей в растворе (их массы в кДа указаны цифрами около кривой) от коэффициента седиментации S

ром (pH = 7.5), а белок gp146 — 50 миллимольным трис-HCl буфером (pH = 7.5) до значения оптической плотности раствора равной 0.6 оптической единицы на длине волны источника излучения 280 нм.

Зависимость концентрации частиц в растворе от коэффициента седиментации для GroEL представлена на рис. 4. Видно, что кроме исследуемого белка в растворе присутствует небольшое количество примесей. Поскольку интенсивность рассеяния пропорцио-



Рис.5. Схема МУР-эксперимента при сканировании по углу при постоянной энергии падающего на образец монохроматического рентгеновского излучения $E(\lambda) = \text{const: } 1 - \text{источник рентгеновско-го излучения; } 2 - \text{Ni-фильтр; } 3 - формирователь рентгеновского пучка по Кратки [1]; } 4 - падающий на образец рентгеновский пучок (12 × 0.2 мм); 5 - образец; 6 - рассеянное излучение; 7 - заслонка прямого пучка; 8 - линейный позиционно чувствительный детектор$

нальна квадрату массы рассеивающих молекул, для малоуглового эксперимента на данном растворе белка наличие небольшой по молекулярному весу примеси не будет давать заметного вклада в экспериментальную кривую рассеяния. Для вирусного шаперонина в тех же условиях наблюдали единственный пик, соответствующий декатетрамерной форме белка, что свидетельствовало о монодисперсности препарата.

Кривые МУР от растворов белков — шаперонина GroEL и вирусного шаперонина gp146 — были получены на лабораторной установке HECUS (Австрия). Источник излучения — рентгеновская трубка с медным анодом и длиной волны характеристического излучения $\lambda(\mathrm{Cu}K_{\alpha}) = 1.54\,\mathrm{\AA}$. Рабочее напряжение U = 50 кВ, ток I = 40 мА. Для отсечки *β*-линии характеристического излучения использовался фильтр (фольга из никеля толщиной 50 мкм). Формирование падающего на образец пучка осуществлялось коллиматором Кратки [1]; высота пучка на образце (длина штриха) 10 мм; ширина пучка в вертикальной плоскости 0.2 мм; расстояние образец-детектор 267 мм; рассеянное излучение регистрировалось линейным позиционно-чувствительным газовым детектором. Схема эксперимента представлена на рис. 5.

Образец помещался в стеклянный капилляр диаметром 1 мм с облучаемым объемом V ≈ 30 мкл. Накопление полезного сигнала осуществлялось сериями продолжительностью $\tau = 1.5$ ч каждая: сигнал от раствора белка снимался в течение 3τ , от буфера — 2τ , далее полученные результаты усреднялись. Из-за несущественной разницы в поглощении между белком и буфером поправка на поглощение не



Рис. 6. Результат введения коллимационной поправки на длину штриха по бегенату серебра: 1 — до введения поправки; 2 — после введения поправки

вводилась. Для выделения рассеяния от исследуемого белка из полученной кривой рассеяния от белка в растворе вычиталась отдельно снятая кривая рассеяния от буфера. В связи с неточечным размером рентгеновского пучка на образце (длина штриха 10 мм) использовалось введение коллимационной поправки. С помощью специальных программных средств выполнялась операция «редукции к идеальному прибору». Переход к шкале модуля вектора рассеяния и введение коллимационной поправки выполнялось по стандартному образцу — бегенату серебра [27], рис. 6.

5. ОБРАБОТКА И ОБСУЖДЕНИЕ ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ

Из графика Гинье по тангенсу угла наклона кривой МУР (см. вставку на рис. 7) была сделана оценка радиуса инерции вирусного шаперонина $R_g =$ = 73 ± 3 Å. Более точная величина, поскольку в ней используется вся кривая рассеяния, была получена после вычисления функции парных расстояний p(r)по формуле (2.5), согласно которой $R_g = 67 \pm 1$ Å.

Была проведена экспериментальная оценка массы M_{exp} молекулы белка по интенсивности рассеяния в нулевой угол, I(0) по формуле:

$$M_{exp} = \frac{c_{BSA}}{c_{exp}} M_{BSA} \frac{I(0)_{exp}}{I(0)_{BSA}}.$$
 (5.1)

Величины, входящие в правую часть выражения (5.1), имели следующие значения: концентрация бычьего сывороточного альбумина (BSA)



Рис.7. Экспериментальные точки МУР для растворов gp146 с концентрацией $c_{exp} = 18$ мг/мл в буфере АДФ (1) и бычьего сывороточного альбумина (BSA), $c_{BSA} = 1$ мг/мл (3), а также сглаженные по РМНК кривые 1 и 3 с аппроксимацией в нулевые углы рассеяния $I_{exp}(0)$ и $I_{BSA}(0)$ — соответственно кривые (2) и (4). На вставке представлена кривая (1) в области Гинье (GR), где $\ln I(s) \approx \ln I_{exp}(0) - (R_q^2/3)s^2$

 $c_{BSA} = 1$ мг/мл, концентрация растворов gp146 $c_{exp} = 18 \pm 0.4$ мг/мл, $M_{BSA} \approx 66$ кДа, экспериментально найденное отношение интенсивностей рассеяния в нулевой угол $I(0)_{exp}/I(0)_{BSA} = 234\pm10$. Отсюда из (5.1) для белка gp146 нами получена оценка его молекулярного веса $M_{exp} = 830$ -900 кДа.

Аналогичные измерения были проведены для бактериального шаперонина GroEL (рис. 8).

Поскольку исследуемые белки являются глобулярными, кривые рассеяния (рис. 9) имеют ярко выраженный минимум, которым можно воспользоваться для независимой оценки максимального размера комплекса. С этой целью воспользуемся известным выражением для формфактора шара диаметром D_{max} , приведенным выше в табл. 1. Найдем из него положение первого минимума (вертикальная прямая на рис. 9). Приравняем нулю числитель в выражении для формфактора шара и, разделив его на $\cos(sr)$, получим трансцендентное уравнение tg(sr) = sr, решая которое численно, найдем его корень с достаточной для нас точностью: sr = 4.4934. Используя определенное из графика значение модуля вектора рассеяния $s = s_{min} = 0.053 \,\text{\AA}^{-1}$, найдем минимальный радиус описанной вокруг молекулы сферы, $r_{min} \approx 85$ Å. С другой стороны, второй корень уравнения $p(D_{max}) = 0$ дает значение для максимального размера молекулы $D_{max} = 180 \,\text{\AA}$ (см.



Рис. 8. Экспериментальные точки МУР (1) от раствора GroEL с концентрацией $c_{exp} \approx 9$ мг/мл; сглаженная по РМНК кривая для точек 1 с аппроксимацией в нулевой угол рассеяния $I_{exp}(0)$ (кривая 2); рассчитанная по формуле Дебая кривая МУР от кристаллической структуры шаперонина GroEL (кривая 3); сглаженные экспериментальные точки МУР от раствора бычьего сывороточного альбумина (BSA), $c_{BSA} = 1$ мг/мл, с аппроксимацией в нулевой угол рассеяния $I_{BSA}(0)$ (кривая 4). На вставке представлена кривая 1 в области Гинье



Рис. 9. Сравнение кривых интенсивности рассеяния растворов фагового (gp146) и бактериального (GroEL) шаперонинов: 1 — сглаженная по РМНК кривая МУР в растворах белка gp146 (см. кривые 2 на рис. 7 и 8); 2 — рассчитанная от кристаллической структуры шаперонина GroEL; 3 — формфактор шара диаметром D = 170 Å для оценки максимального размера этих белков



Рис. 10. Зависимость общего решения $\Sigma(\alpha)$ задачи минимизации функционала (2.15) от параметра регуляризации α ; три наилучших решения обведены овалом; Σ_{max} при $\alpha = 0.0546$



Рис. 11. Функция парных расстояний внутри молекулы белка gp146, рассчитанная из кривой 1 на рис. 9

рис. 9). Для нижней границы максимального размера частицы получим $D'_{max} = 2r_{min} \approx 170$ Å. Таким образом, оба белка, бактериальный и фаговый шаперонины, имеют близкие друг к другу максимальные размеры $D_{max} = 170-180$ Å.

С целью поиска устойчивого решения для функции парных расстояний p(r) кривая МУР (см. кривую 1 на рис. 7) многократно подвергалась фурье-преобразованию с регуляризацией по Тихонову с различными значениями параметров регуля-



Рис. 12. Рассчитанные кривые МУР для кристаллических структур GroEL (1) и GroEL/GroES (2), а также сглаженная по РМНК экспериментальная кривая рассеяния от раствора белка GroEL (см. рис. 8, кривая 2) с аппроксимацией к I(0) (кривая 3). На вставке — экспериментальная кривая рассеяния 4 от комплекса GroEL/GroES сравнивается с кривой 2

ризации (рис. 10). В результате было найдено наилучшее решение для p(r), которое представлено на рис. 11. Из него следует, что максимальный размер молекулы составляет $D_{max} = 180 \pm 5$ Å. Отсюда следует, что если бактериальные шаперонины практически одинаковы по длине полипептидной цепи GroEL E. coli-548 аминокислотных остатков, и имеют высокую степень гомологии (около 80%) этих остатков, то, несмотря на низкую степень гомологии (ниже 21 %), между бактериальным GroEL и фаговым gp146 шаперонинами из малоуглового эксперимента (см. рис. 9) следует, что габаритные размеры, передающие внешнюю форму, молекулы близки друг к другу. Согласно рис. 9, различие в области $s > 0.1 \,\mathrm{\AA^{-1}}$ обратного пространства соответствует размеру $R < 2\pi/s \approx 60$ Å. Поэтому можно предположить, что различия в форме этих макромолекул проявляются на величине масштабов одной субъединицы в макромолекуле.

Если сравнить кривые малоуглового рассеяния для шаперонина GroEL в кристалле и в растворе, то можно заметить существенное различие прежде всего в размере молекулы. А именно, уже в области Гинье (рис. 12) заметно, что как для отдельного белка GroEL (кривые 1 и 3), так и для комплекса GroEL/GroES (кривые 2 и 4), кривые 1 и 2, рассчи-



Рис. 13. Функции парных расстояний p(r), вычисленные по экспериментальной кривой МУР для раствора GroEL (см. рис. 12, кривая 3), $D_{max} = 170-175$ Å, $R_g = 62$ Å (кривая 1), для кристаллической структуры GroEL (см. рис. 12, кривая 1), $D_{max} = 190-200$ Å, $R_g = 68$ Å (кривая 2), для раствора GroEL/GroES, $D_{max} = 180 \pm 5$ Å, $R_g = 63$ Å (кривая 3) и для кристаллической структуры GroEL/GroES (см. рис. 12, кривая 2), $D_{max} = 215 \pm 5$ Å, $R_g = 83$ Å (кривая 4)

танные для кристаллической структуры, идут ниже, чем экспериментальные кривые 2 и 4, снятые для раствора белков. Следовательно, по области Гинье можно сказать, что максимальные размеры молекул в кристалле больше, чем в растворе.

Не ограничиваясь качественным анализом в области Гинье, мы провели фурье-преобразование с помощью РМНК, см. разд. 2.3, формулу (2.15) для расчета максимального размера по кривым рассеяния, представленным на рис. 12. Функция парных расстояний p(r), была вычислена как для индивидуальных шаперонинов (GroEL и gp146), так и для белка GroEL в составе комплекса GroEL/GroES (рис. 13).

В сводной табл. 2 представлены основные параметры белков-шаперонинов, полученные в результате МУР-эксперимента. Для состояния «раствор» кривые МУР получены из эксперимента, для состояния «кристалл» — вычислены по формуле Дебая. Полученные значения этих параметров для максимального размера бактериального шаперонина GroEL и в комплексе GroEL / GroES согласуется с работой [28].

Таблица 2. Параметры молекул белков, вычисленные из кривых МУР через преобразование Фурье

Белок	Состояние	$R_g, \mathrm{\AA}$	$D_{max}, \mathrm{\AA}$
gp146	Раствор $(c \approx 18 \text{ мг/мл})$	67 ± 2	175-185
	Кристалл	68 ± 2	190-200
GroEL	Раствор $(c \approx 11 \text{ мг/мл})$	62 ± 2	170-175
CroFI /CroFS	Кристалл	83 ± 3	210-220
GIOED/GIOES	Раствор $(c \approx 15 \text{ мг/мл})$	64 ± 2	175-185

6. МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

6.1. Восстановление формы шаперонина GroEL методом оболочек

Согласно методу функции оболочек [23] для экспериментальной кривой рассеяния для раствора белка GroEL, были вычислены коэффициенты ряда Фурье-Лапласа при сферических гармониках, см. выражение (3.2). Полученная пространственная суперпозиция сферических гармоник, представленная на рис. 14, удовлетворительно передала пространственный характер формы частицы. Мы видим (рис. 14*б*) более высокую рассеивающую плотность на поверхности частицы GroEL, чем внутри ее объема, а также наличие оси симметрии у частицы. Однако мы не можем сказать о характере неоднородности внутри объема частицы.

Согласно рис. 14 можно сделать вывод, что электронная плотность для данной частицы шаперонина сконцентрирована на боковой поверхности шаперонина. Расчет шариковой модели [26] для кристаллической структуры белка был осуществлен в предыдущей работе [4]. Здесь мы рассчитаем параметры модели формы белка GroEL, используя формфактор двойного эллиптического тора.



Рис. 14. Поверхность, полученная из экспериментальной кривой рассеяния (см. рис. 8, кривая 2) для шаперонина GroEL как суперпозиция L = 15 сферических гармоник: a — вид сбоку; δ — вид A

6.2. Приближение экспериментальных кривых рассеяния для растворов фагового (gp146) и бактериального (GroEL) шаперонинов с помощью моделей двойного эллиптического тора и полого цилиндра, заданных своими формфакторами

В работе [4] получено выражение для формфактора (интенсивности) двойного эллиптического тора:

$$I(s) = \left(\frac{2}{\pi a b c s}\right)^2 \times \\ \times \int_{0}^{\pi/2} \frac{1}{\cos^2 \theta} \left\{ \int_{b-a}^{b+a} \{ [\sin \left(s c (1+\delta) \cos \theta\right) + \\ + \sin \left(s c (1-\delta) \cos \theta\right) \right] J_0(sr \sin \theta) r \, dr \}^2 \right\} \times \\ \times \sin \theta \, d\theta, \quad (6.1)$$

где

$$\delta = \sqrt{1 - \frac{(r-b)^2}{a^2}} \,.$$

В этой же работе также была вычислена кривая МУР I(s) для кристаллической структуры шаперонина GroEL (Protein Data Bank (PDB) с идентификатором (DOI): 1kpo), который является олигомерным белком, состоящим из 14 идентичных субъединиц (57 кДа каждая), которые образуют два тороидальных кольца (по 7 субъединиц в каждом), лежащих друг на друге (рис. 15).

В качестве второй модели для описания формы молекулы в однородном приближении мы используем полый цилиндр (см. табл. 1). Квазиньютоновским методом минимизации с переменной метрикой — методом Бройдена – Флетчера – Гольдфарба – Шанно



Рис.15. Пространственная структура кристаллического шаперонина GroEL: a — вид сбоку; δ — вид A



Рис. 16. Поиск параметров модели по кривой МУР для белков gp146 в растворе: 1 - сглаженная экспериментальная кривая рассеяния для раствораgp146; <math>2 -расчет по формфактору двойного эллиптического тора, формула (6.1) с параметрами $a = 23 \pm 1.5$ Å, $b = 49 \pm 3$ Å, $c = 35 \pm 2$ Å, $R_f = 0.024$; 3 -расчет по формфактору полого цилиндра (см. табл. 1) с $h = 128 \pm 6$ Å, $r_{max} = 60 \pm 3.5$ Å, $r_{min} = 30 \pm 2$ Å, R = 0.07

(BFGS) [29] — была осуществлена подгонка экспериментальной кривой рассеяния (кривые 1 на рис. 16–18) к аналитическим выражениям для двойного тора (6.1) и для формфактора полого цилиндра (см. табл. 1) путем независимого варьирования трех параметров: образующей тора b и двух полуосей a и c (см. рис. 21 a).

В результате численной подгонки были найдены параметры для моделей в виде двойного эллиптического тора и полого цилиндра. Для каждого слу-



Рис. 17. Поиск параметров модели по кривой МУР для белка GroEL в растворе: 1 -сглаженная экспериментальная кривая рассеяния, нормировка I(0) = 1; 2 -расчет по формфактору двойного эллиптического тора, формула (6.2); 3 -расчет по формфактору полого цилиндра (см. табл. 1)



Рис. 18. Поиск параметров модели по кривой МУР для белка GroEL в кристалле: 1 — расчет по формуле Дебая (2.17); 2 — расчет по формфактору двойного эллиптического тора, формула (6.2); 3 расчет по формфактору полого цилиндра, формула (5.3)

чая из экспериментальной кривой предварительно вычитался постоянный фон. Степень близости при варьировании геометрических параметров моделей (a, b, c или $r_{min}, r_{max}, h)$ численно оценивалась величиной *R*-фактора. Под *R*-фактором понимается поточечная среднеквадратичная невязка:

$$R \approx \frac{1}{||\tilde{I}(s)_{exp}||} \sum_{j=1}^{N} \left[\tilde{I}(s)_{mod} - \tilde{I}(s)_{exp} \right]^2.$$
(6.2)

Чем меньше величина R-фактора, тем лучше описывает модель при найденных программой ее геометрических параметрах форму шаперонина. Поскольку кривая рассеяния имеет высокую скорость спада, для уравнивания вклада всех точек на этой кривой рассеяния при вычислении R-фактора по формуле (6.2) проводилось «взвешивание» кривой рассеяния по закону $\tilde{I}(s) = \sqrt[3]{I(s)}$. Наилучшие (т. е. для которых R-факторы минимальны) подгоночные кривые $I(s)_{mod}$, полученные с помощью этих моделей, к экспериментальным кривым рассеяния для шаперонинов представлены на рис. 16–18.

Для быстрой диагностики топологии поверхности исследуемых шаперонинов из кривой МУР можно использовать асимптотики Порода для шара и тора (см. табл. 1) при полученных значениях параметров $D_{max} = 185$ Å, a = 23 Å, b = 50 Å, c = 35 Å с нормировкой на интенсивность рассеяния в нулевой угол I(0) = 642200:

$$I(s)_P \approx I(0) \frac{9}{(sD_{max})^4},$$
 (6.3)

$$I(s)_P \approx I(0) \frac{4}{bs(as)^2(cs)^3}.$$
 (6.4)

Наложив эти асимптотики на экспериментальную кривую рассеяния для gp146 (рис. 19), можно по хвостовой части графиков убедиться, что скорость спада кривой МУР для данного белка лучше описывается асимптотикой тора (6.4), чем шара (6.3).

Результаты поиска параметров моделей полого цилиндра и тора и соответствующих значений *R*-факторов представлены в табл. 3.

Форму фагового шаперонина gp146 в растворе восстанавливали, используя три различных метода. Соответственно были получены три модели: первая — построенная методом оболочек (см. рис. 14), вторая — в виде полого цилиндра и третья модель в виде двойного эллиптического тора. Оказалось, что полученные модели различаются величиной радиуса центрального канала. А именно: в методе оболочек, как и ожидалось, никакая комбинация сферических гармоник не в состоянии описать наличие центрального канала, однако из рис. 14 можно заключить, что основной вклад в кривую рассеяния вносит небольшой, относительно своего радиуса, цилиндрический слой. Если толщина этого слоя из модели полого цилиндра, согласно табл. 3, составляет $r_{max} - r_{min} = 30 \pm 3$ Å и совпадает с

Белок		Параметры модели полого нилинира			Параметры модели двойного				
		параметры модели полого цилиндра				эллиптического тора			
		$h, \mathrm{\AA}$	$r_{max}, \mathrm{\AA}$	$r_{min}, \mathrm{\AA}$	R	a, Å	$b, \mathrm{\AA}$	$c, \mathrm{\AA}$	R
gp146		128 ± 6	60 ± 3.5	30 ± 3	0.070	23 ± 1.5	49 ± 3.5	35 ± 2	0.024
GroEL	Кристалл	148 ± 7	66 ± 4	20 ± 2.5	0.018	22 ± 1.5	46 ± 3	40 ± 2.5	0.015
	Раствор	138 ± 6	62 ± 3.5	15 ± 2	0.045	25 ± 1.5	40 ± 2.5	30 ± 2	0.014

Таблица 3. Геометрические параметры моделей полого цилиндра и двойного симметричного эллиптического тора, вычисленные из кривых МУР для белков-шаперонинов



Рис. 19. Асимптотики Порода для кривой МУР для белка gp146 в растворе: 1 -сглаженная экспериментальная кривая рассеяния для gp146, $c_{exp} = 18 \text{ мг/мл}; 2 -$ асимптотика Порода для тора, см. формулу (6.4); 3 -асимптотика Порода для шара, см. формулу (6.3)

радиусом центрального канала $r_{min} = 30 \pm 3$ Å, то для модели двойного тора минимальный радиус центрального канала, согласно табл. 3, немного меньше: $r = b - a = 25 \pm 3$ Å. Последняя величина представляется более правдоподобной, если сравнивать данный параметр с аналогичным для макромолекулы бактериального GroEL в растворе. В модели полого цилиндра, согласно табл. 3, радиус центрального канала GroEL в кристалле составляет $r_{min} = 15$ Å, что меньше радиуса канала для GroEL в растворе.

6.3. Конформационные формы вирусного шаперонина в растворе

Поскольку фаговый шаперонин подобно другим известным шаперонинам функционирует АДФ-зависимым способом, он, по-видимому, может находиться в двух конформациях — открытой и закрытой. В присутствии АДФ он, возможно, находится в открытой конформации. С целью обнаружения этого конформационного перехода были сняты кривые МУР для белка gp146 в «исходном» буфере и буфере с добавлением АДФ. Кривые рассеяния для шаперонина gp146 в исходном буфере и в буфере с добавлением АДФ представлены на рис. 20*a*.

Моделирование формы молекулы осуществлялось двумя методами: методом пространственной конфигурации аминокислотных остатков и методом поиска геометрических параметров в модели двойного эллиптического тора с независимыми параметрами для каждого тора. В последнем методе, взяв за основу выражение (6.1) для формфакторов двух симметричных торов, обобщим это выражение для случая несимметричных торов, имеющих одну образующую b и независимые для каждого тора полуоси a_1, c_1, a_2, c_2 (рис. 21*e*):

$$I(s) = \left(\frac{2}{\pi a b c s}\right)^2 \int_0^{\pi/2} \frac{1}{\cos^2 \theta} \times \\ \times \left\{ \int_{b-a_1}^{b+a_1} \left[F_1(sc_1, \delta_1(r), \theta) r \, dr \right. + \\ \left. + \int_{b-a_2}^{b+a_2} F_2(sc_1, \delta_2(r), \theta) r \, dr \right]^2 \right\} \sin \theta \, d\theta, \quad (6.5)$$

где

 $F_1(sc_1, \delta_1(r), \theta) = \sin[sc_1(1+\delta_1)\cos\theta] J_0(sr\sin\theta),$ $F_2(sc_1, \delta_2(r), \theta) = \sin[sc_2(1-\delta_2)\cos\theta] J_0(sr\sin\theta),$

$$\delta_{1,2} = \sqrt{1 - (r-b)^2 / a_{1,2}^2}.$$



Рис. 20. *а*) Экспериментальные кривые МУР для раствора gp146: 1 — в исходном буфере (50 mM Tris-HCl, pH = 7.5, 10 mM MgCl₂, 100 mM KCl); 2 — в исходном буфере с добавлением 3 mM АДФ (концентрация белка в растворе 11.3 мг/мл). *б*) Функции парных расстояний $p_1(r)$ и $p_2(r)$, рассчитанные методом РМНК для функционала (2.15), из кривых рассеяния соответственно 1 и 2 на рис. 16

Таким образом, в нашем случае число независимых параметров модели возросло до K = 5. Следует заметить, что в общем случае число независимых параметров модели не должно быть больше числа шеноновских узлов, определяемых выражением (3.7). В нашем случае, подставляя в (3.7) численные значения $s_{max} = 0.200$ Å, $s_{min} = 0.018$ Å, $D_{max} = 190$ Å, для максимального числа независимых параметров получим $K_{max} = 12$; условие $K < K_{max}$ выполнено.

При моделировании аминокислотными остатками в качестве подгоночной кривой в данном случае выступает не сама кривая рассеяния, как в методе шаров, а функция парных расстояний p(r), см. формулы (2.7)-(2.10). Для использования этого метода кроме предварительно вычисленной функции p(r) необходимо знать *a priori* последовательность аминокислотных остатков и молекулярный вес исследуемой молекулы. По описанной выше методике был проведен расчет функции парных расстояний для кривых МУР белка gp146 в исходном буфере (кривая 1) и в буфере с АД Φ (кривая 2), представленных на рис. 20б. Поскольку кривая $p_1(r)$ идет выше кривой $p_2(r)$, вклад отрезков различной длины в форму молекулы для исходной конформации белка имеет более однородный характер. Сравнивая формы функций $p_1(r)$ и $p_2(r)$ можно предположить, что макромолекула gp146 в исходном буфере (кривая 1) более близка к глобулярной (шарообразной) форме, чем в буфере с АД Φ (кривая 2). Более определенные выводы о форме белковых комплексов можно сделать по результатам моделирования аминокислотными остатками по функции парных расстояний (рис. 20б). Вычисление функции парных расстояний p(r) для всех экспериментальных данных проводилось согласно ГОСТ Р 8.698-2010 [30]. Результаты восстановления формы макромолекулы методом сворачивания аминокислотных последовательностей с использованием формфакторов аминокислоты и несимметричного двойного эллиптического тора представлены на рис. 21. Величины R-факторов, вычисленные по формуле (5.2), составляют R = 0.01410 для белка в исходном буфере и R = 0.01195 для белка в буфере с АДФ.

7. ВЫВОДЫ

1. С целью исследования возможностей практического применения ранее вычисленного в работе [4] формфактора двойного эллиптического тора был проведен эксперимент по малоугловому рассеянию рентгеновских лучей в растворах бактериального (GroEL) и вирусного (gp146) шаперонинов. Представлен сравнительный анализ размеров и формы белков в растворе. Для бактериального белка GroEL также проведено сравнение экспериментальных кривых МУР с теоретически рассчитанными на основе известной кристаллической структуры.

2. Как для индивидуального GroEL, так и для шаперонина в составе комплекса с ко-шаперонином (GroEL/GroES) максимальный размер частицы в растворе, вычисленный через функцию парных расстояний, оказался на 20–30 Å больше, чем в кристалле (см. табл. 2) что согласуется с результатами работы [28].



Рис.21. Моделирование аминокислотными остатками с осью симметрии *P*7 и модель макромолекулы в виде двойного эллиптического тора: *a*, *б* − gp146 в исходном буфере с АДФ; *в* − модель симметричных торов, численные значения параметров см. в табл. 2 (рассчитано для кривой рассеяния, представленной на рис. 20*a*, кривая 2); *e*, *д* − gp146 в исходном буфере; *e* − модель несимметричных торов с параметрами *a*₁ = 5 Å, *c*₁ = 40 Å, *a*₂ = *c*₂ = 30 Å, *b* = 50 Å (рассчитано по кривой рассеяния, представленной на рис. 20*a*, кривая 1)

3. По результатам подгонки формы макромолекулы моделями цилиндра и двойного эллиптического тора оказалось, что последняя модель по *R*-фактору более точно подгоняется к экспериментальной кривой рассеяния. Основным недостатком модели полого цилиндра является наличие резкой границы между торцом и образующей. Ясно, что поверхность макромолекулы не может содержать таких резких границ. Иными словами, наличие этой грани между торцевой поверхностью цилиндра и образующей может давать существенный вклад в кривую рассеяния для этой модели.

4. Из качественного анализа кривых рассеяния по асимптотике Порода (см. рис. 19) следует, что фаговый шаперонин gp146, так же как и бактериальный белок GroEL, имеет двусвязную поверхность.

5. На основе кривых малоуглового рассеяния вычислены молекулярные структурные параметры — радиус инерции R_g и максимальный размер макромолекулы D_{max} — для бактериального и вирусного шаперонинов, соответственно $R_g = 62 \pm 2$ Å, $D_{max} = 170-175$ Å и $R_g = 67 \pm 2$ Å, $D_{max} = 175-185$ Å.

12 ЖЭТФ, вып. 2 (8)

6. Было показано, что модель двойного эллиптического тора для исследуемых шаперонинов согласуется с моделью, полученной вращением аминокислотных остатков. Используя полученные в предыдущей работе [4] точное выражение для формфактора модели двойного эллиптического тора, одним из методов квазиньютоновской минимизации мы нашли геометрические параметры *a*, *b*, *c* этой модели для каждого белка (см. табл. 2).

7. Показано, что в зависимости от внешних условий для вирусного шаперонина при сохранении габаритных размеров частицы и радиуса инерции возможен конформационный переход от симметричной формы, описываемой моделью двойного эллиптического тора с параметрами $a = a_1 = a_2$, b, $c = c_1 = c_2$ к несимметричной форме с параметрами $a_1 \neq a_2$, b, $c_1 \neq c_2$.

Работа выполнена при финансовой поддержке РФФИ (грант № 08-08-00540-а) и частично Президиума РАН (грант № 27/3.5.2).

ЛИТЕРАТУРА

- 1. O. Glatter and O. Kratky, *Small-Angle X-ray Scattering*, Acad. Press, London-New York (1982).
- A. Guinier, Ann. Phys. 12, 161 (1939); G. Porod, Acta Phys. Austr. 2, 255 (1948).
- Д. И. Свергун, Л. А. Фейгин, Рентгеновское и нейтронное малоугловое рассеяние, Наука, Москва (1986).
- **4**. С. В. Амарантов, ЖЭТФ **135**, 72 (2009).
- W. A. Fenton and A. L. Horwich, Q. Rev. Biophys. 36, 229 (2003).
- F. U. Hartl and M. Hayer-Hartl, Science 295, 1852 (2002).
- J. C. Young, V. R. Agashe, K. Siegers, and F. U. Hartl, Nat. Rev. Mol. Cell Biol. 5, 781 (2004).
- D. Thirumalai and G. H. Lorimer, Ann. Rev. Biophys. Biomol. Struct. 30, 245 (2001).
- K. Braig, Z. Otwinowski, R. Hegde et al., Nature 371, 578 (1994).
- Z. H. Xu, A. L. Horwich, and P. B. Sigler, Nature 388, 741 (1997).

- J. S. Weissman, C. M. Hohl, O. Kovalenko et al., Cell 83, 577 (1995).
- M. A. Mayhew, C. R. Da Silva, J. Martin et al., Nature 379, 420 (1996).
- K. Hertveldt, R. Lavigne, E. Pleteneva et al., J. Mol. Biol. 354, 536 (2005).
- 14. F. J. Corrales and A. R. Fersht, Fold. Des. 1, 265 (1996).
- I. N. Naletova, V. I. Muronetz, and E. V. Schmalhausen, Biochim. Biophys. Acta 1764, 831 (2006).
- 16. C. Walters, N. Errington, A. J. Rowe, and S. E. Harding, Biochem. J. 364, 849 (2002).
- 17. В. Де Альфаро, Т. Редже, Потенциальное рассеяние, Мир, Москва (1966).
- **18**. Л. Д. Ландау, Е. М. Лифшиц, Квантовая механика (нерелятивистская теория), Наука, Москва (1989).
- **19**. А. Н. Тихонов, В. Я. Арсенин, Методы решения некорректных задач, Наука, Москва (1986).
- 20. D. I. Svergun, J. Appl. Cryst. 25, 495 (1992).
- 21. A. J. Jerri, Proc. IEEE 65, 1565 (1977).
- 22. D. I. Svergun, V. V. Volkov, M. B. Kozin, and H. B. Stuhrmann, Acta Crystallogr. A 52, 419 (1996).
- 23. M. Petoukhov and D. Svergun, J. Appl. Cryst. 28, 768 (1995).
- 24. R. D. B. Fraser, T. P. Macrae, and E. J. Suzuki, J. Appl. Cryst. 11, 693 (1978).
- 25. L. Ingber, Math. Comp. Mod. 18, 29 (1993).
- 26. D. I. Svergun, Biophys. J. 76, 2879 (1999).
- 27. T. C. Huang, H. Toraya, T. N. Blanton, and Y. Wu, J. Appl. Cryst. 26, 180 (1993).
- 28. T. Inobe, K. Takahashi, K. Maki et al., Biophys. J. 94, 1392 (2008).
- 29. Ф. Гилл, У. Мюррей, М. Райт, *Практическая оптимизация*, Мир, Москва (1985).
- 30. Национальный стандарт Российской Федерации. Государственная система обеспечения единства измерений. Размерные параметры наночастиц и тонких пленок. Методика выполнения измерений с помощью малоуглового рентгеновского дифрактометра. ГОСТ Р 8.698-2010, Стандартинформ, Москва (2010).