КРИСТАЛЛОГРАФИЯ, 2024, том 69, № 6, с. 971–980

СТРУКТУРА МАКРОМОЛЕКУЛЯРНЫХ СОЕДИНЕНИЙ

УДК 539.219.1

ОСОБЕННОСТИ ТРЕХМЕРНОЙ РЕКОНСТРУКЦИИ СПИРАЛЕЙ ПО ДАННЫМ МАЛОУГЛОВОГО РЕНТГЕНОВСКОГО РАССЕЯНИЯ

© 2024 г. В. А. Григорьев[1,*], П. В. Конарев[1], В. В. Волков[1]

[1]*Институт кристаллографии им. А.В. Шубникова Курчатовского комплекса кристаллографии и фотоники НИЦ "Курчатовский институт", Москва, Россия*

**E-mail: vasiliy.grigorev.1996@mail.ru*

Поступила в редакцию 08.08.2024 г.
После доработки 12.09.2024 г.
Принята к публикации 18.09.2024 г.

Спиральные частицы представляют интерес из-за схожести с реальными наноструктурами, образующимися в процессах самоорганизации биополимеров (например каррагинанов, ДНК и пр.). С другой стороны, определение структурных параметров таких частиц по данным малоуглового рассеяния затруднено из-за плохой обусловленности обратной задачи. Это показывает практика применения известных программ моделирования шариковыми структурами. Рассмотрены модификация алгоритма поиска в ограниченной области пространства и поведение решений в зависимости от величин параметров целевой функции, отвечающих за связность структуры, типа взвешивания кривой интенсивности рассеяния, ширины углового диапазона данных. Для статистической оценки устойчивости решений применен режим последовательного поиска модели при варьировании величин вкладов штрафных членов. Определены эмпирические зависимости оптимальных значений параметров поиска в зависимости от параметров кривой распределения парных расстояний.

DOI: 10.31857/S0023476124060063, **EDN:** YHUDKD

ВВЕДЕНИЕ

Одним из незаменимых методов для получения информации о строении наночастиц (**НЧ**), размеры которых составляют от 10 до 5000 Å, является малоугловое рентгеновское рассеяние (**МУРР**) [1]. Этот неразрушающий метод позволяет получить такую информацию о НЧ, как форма, масса, радиус инерции, объем и максимальный размер. Расчет большинства характеристик НЧ доведен до автоматизма, однако до сих пор представляет интерес поиск трехмерной формы НЧ монодисперсной системы. Существует несколько способов представления трехмерной формы. Так, в [2, 3] использовали разложение плотности частицы на сферические гармоники с последующим поиском коэффициентов разложения с помощью численных методов. В [4] использовали метод машинного обучения для построения модели по данным МУРР. В большинстве методов поиска трехмерной формы используют шариковую модель НЧ, в которой структура представлена в виде набора однородных шариков, находящихся в узлах гексагональной решетки. Эти методы реализованы в наиболее популярных программах DAMMIN [5] и DAMMIF [6]. Основное различие программ в том, что в первой шариковая модель ограничена заданной областью, обычно сферой с диаметром, равным максимальному размеру частицы, а во второй – область поиска не ограничена.

В упомянутых программах, использующих шариковые модели, применяется алгоритм поиска решения, известный как "имитация отжига" [7]. Алгоритм позволяет восстановить трехмерную форму большинства типов НЧ, однако не является универсальным, и остаются задачи, которые нельзя решить, используя только шариковые модели. Так, например, в [8] была попытка восстановления трехмерной формы спиральной частицы, которая не восстанавливалась с помощью DAMMIN и DAMMIF, с помощью модифицированного алгоритма. Этот подход подробно описан в [9]. Частично в [10] был проведен анализ алгоритма на предмет эффективности на примере спиралей с разным относительным шагом спирали.

Спирали как модели формы рассеивающих тел представляют собой интерес в связи со схожестью с реальными НЧ, например комплексом ι-каррагинана с катионным поверхностно-активным веществом (цетилпиридин хлоридом) [11, 12]. Спираль представляет собой пример частицы, форма которой восстанавливается крайне неустойчиво. Одна из причин этого – сильная корреляция парных

расстояний между точками в пространственной модели, соответствующих разным структурным элементам. Функция парных расстояний $p(r)$ представляет собой фурье-образ кривой интенсивности МУРР $I(s)$, определенный в прямом пространстве:

$$p(r) = \frac{1}{2\pi^2} \int_{s=0}^{\infty} sr \cdot I(s) \cdot \sin(sr) \cdot ds ,$$

где s – модуль вектора рассеяния $|s| = \frac{4\pi \cdot \sin\theta}{\lambda}$, λ – длина волны излучения, θ – половина угла рассеяния, r – модуль расстояния, или длина отрезка, соединяющего два рассеивающих центра в структуре. Заметим, что в этой формуле для простоты не указан контраст рассеивающей плотности (в случае рентгеновских лучей – разность между электронной плотностью объекта и плотностью окружения – растворителя), который полагаем здесь равным единице. В случае спиралей одна и та же величина расстояния может описывать диаметр нити спирали, шаг спирали, внутренний диаметр. Соответственно, программе поиска сложно адекватно распределить такие отрезки в структуре, если их длины слабо различаются. Но неустойчивость восстановления модели остается довольно высокой и в случае существенного различия таких характеристических парных расстояний. Это связано с относительно низким пространственным разрешением модели из-за ограниченности доступного углового диапазона МУРР и сферическим усреднением интенсивности по ориентациям.

Тот факт, что форма спиралей с определенными геометрическими характеристиками плохо реконструируется с использованием базовых методов, подталкивает на поиск способа решения проблемы неустойчивости, которая связана не только с плохой обусловленностью задачи, но и с математической неоднозначностью (в пределах структурных инвариантов модели частиц должны иметь одинаковый максимальный размер, радиус инерции, площадь поверхности, объем). Исследование новых алгоритмов трехмерной реконструкции НЧ должно позволить избежать неправильной интерпретации результатов расчетов. Для этой цели в настоящей работе были восстановлены модели спиралей с разным шагом спирали при большом количестве различных значений параметров модифицированного алгоритма “имитации отжига” [9] для установления эмпирической зависимости результатов восстановления от шага спирали.

МОДИФИЦИРОВАННЫЙ АЛГОРИТМ “ИМИТАЦИИ ОТЖИГА”

Основное отличие модифицированного алгоритма “имитации отжига”, заложенного в программе DAMMINV [9], от базового (DAMMIN [5]) в наличии режима перемежающихся в процессе поиска относительных весов штрафных членов целевой функции, отвечающих за компактность и связность модели, а также вклада невязки эксперимент–модель. Полная целевая функция имеет вид

$$\Phi(\mathbf{X}) = w_R R + w_D P_D + w_L P_L + w_C P_C, \qquad (1)$$

где w_i и P_i – веса и штрафные коэффициенты, отвечающие за разрывность частицы (D), ее рыхлость (L) и отклонение центра масс частицы от центра области поиска (C). R – невязка между экспериментальной и модельной кривыми интенсивности рассеяния:

$$R = \frac{\sum_{i=1}^{N}\left[\left(\mathbf{I}_{\exp}(s_i) - \xi \mathbf{I}_{\mathrm{mod}}(s_i)\right) W(s_i)\right]^2}{\sum_{i=1}^{N}\left[\mathbf{I}_{\exp}^2(s_i) W^2(s_i)\right]}, \qquad (2)$$

где $\xi = (\mathbf{I}_{\exp}\mathbf{I}_{\mathrm{mod}})/\|\mathbf{I}_{\exp}\|^2$ – шкалирующий множитель, $W(s)$ – весовая функция, применяемая к интенсивности рассеяния:

$$W(s) = \begin{cases} s^n, n = 0:4.0 \text{ если } s > s\big|_{\max\left[I_{\exp}(s)\cdot s^n\right]} \\ 0.5 \cdot \left\{ s^n + \dfrac{\max\left[\mathbf{I}_{\exp}(s)\cdot s^n\right]}{\mathbf{I}_{\exp}(s)} \right\} \text{ если } s \le s\big|_{\max\left[I_{\exp}(s)\cdot s^n\right]} \end{cases}. \qquad (3)$$

Взвешивание интенсивности необходимо для выравнивания вкладов в невязку от высоко- и малоинтенсивной частей кривой интенсивности независимо от оценки стандартного отклонения шумов измерений.

Режим перемежающихся весовых коэффициентов работает следующим образом. Целевая функция минимизируется с помощью алгоритма “имитации отжига”, как в программах DAMMIN [5] и DAMMIF [6]. После достижения минимума (1) полученная модель сохраняется, затем температура (которая отвечает за “варьируемость” модели в процессе поиска) повышается в 5–10 раз. Вес, отвечающий за невязку, уменьшается в 10 раз. Последнее меняет рельеф целевой функции, и, как следствие, возникает ненулевой градиент функции, что провоцирует поиск нового глобального минимума. Кроме того, меняется акцент в сторону минимизации штрафных коэффициентов, отвечающих за разрывность, рыхлость частицы и т.д.

(формула (1)). При достижении нового минимума получается модель с лучшими “физическими” параметрами, но худшим значением невязки R (2) (обычно в 1.5–2.5 раза). На следующей итерации вес, отвечающий за невязку, устанавливают таким же, как на первой итерации, и процедура повторяется. Всего получается 10–15 моделей, каждая из которых соответствует минимуму целевой функции (1) и может рассматриваться как решение.

В алгоритме DAMMINV важен выбор значений весов, отвечающих за структурные параметры модели, поскольку в ходе работы алгоритма меняется вес невязки, а результат минимизации зависит от соотношения весов, отвечающих за физический смысл модели, и веса невязки w_R. Значительное влияние оказывает изменение весов, отвечающих за разрывность и рыхлость частицы. Эти параметры определяют вероятность качественного перехода из одной формы частицы в другую в процессе минимизации и, соответственно, возможность получения новой конфигурации частицы, отвечающей экспериментальным данным.

Для понимания того, как меняется результат восстановления формы частицы от веса ее разрывности и рыхлости, веса в настоящей работе меняли в широких пределах: $w_D = 0.0003$–0.96, $w_L = 0.006$–3.84. По умолчанию значения весов определяются в программе по эмпирической формуле, выведенной автором программы DAMMINV опытным путем:

$$w_i = 0.03 \cdot \left(\frac{a_i \cdot 10^4}{N} + 1.5 \right), \qquad (4)$$

где $a_i = 1$ и 2 для $i = L$ и D соответственно, N – число узлов в области поиска.

Для тестирования алгоритма выбраны трехвитковые спирали с толщиной витков $d = 30$ Å, внешним диаметром $D = 100$ Å и различным шагом спирали $h = 45$, 50, 55 и 60 Å (модели I, II, III, IV на рис. 1). Спирали с заданными толщиной и количеством витков были предварительно отобраны опытным путем как наиболее сложные для восстановления с помощью базового алгоритма “имитации отжига”.

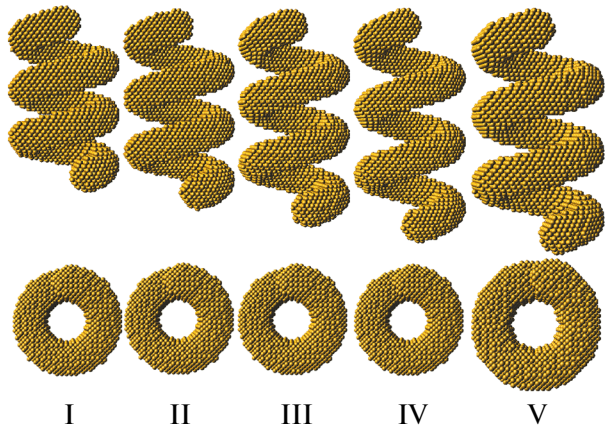

Рис. 1. Теоретические модели спиралей: сверху – вид сбоку, снизу – вид сверху. Шаг спирали (слева направо): 45, 50, 55 и 60 Å при диаметре 100 и 62 Å при диаметре 120 Å.

Кроме весов варьировали степень весовой функции (формула (3)): $n = 1.5$ и 2.0, а также угловой диапазон данных, представленный числом шенноновских каналов [13]: $N_{sh} = 20$ и 28 ($N_{sh} = s_{\max} D_{\max}/\pi$, $s_{\max}$ – максимальное значение модуля вектора рассеяния, $D_{\max}$ – максимальный размер частицы). Эти значения отобраны как наиболее оптимальные согласно [10].

Качество восстанавливаемых частиц оценивали с помощью параметра пространственной корреляции NSD, вычисленного с помощью программы SUPCOMB [14], который показывает сходство двух моделей, предварительно ориентированных до достижения максимального перекрытия объемов. При NSD = 0 частицы идентичны, при NSD $\geqslant 1$ – существенно различаются по структуре. В настоящей работе восстановленные модели сравнивали с теоретическими, которые также представляют собой шариковые модели, но с меньшим (~ в 2 раза) радиусом шариков. В этом случае для идеально восстановленной спирали NSD отлично от нуля и приблизительно равно 0.45.

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

В настоящей работе программа DAMMINV выдавала пять–семь структур в одном расчете, для каждой модели рассчитан параметр NSD. Всего было проведено 2240 расчетов, в каждом из которых получено от пяти до семи моделей. Термин “расчет” далее рассматривается как результат запуска программы DAMMINV для конкретной уникальной комбинации значений параметров N_{sh}, n, w_D, w_L и шага спирали. Для того чтобы определить, какую оценку качества расчета использовать, сначала значения NSD в группах расчетов из пяти–семи моделей были отсортированы в порядке возрастания, затем внутри каждой группы NSD было усреднено между моделями в разных расчетах так, чтобы получить общее представление о возрастании NSD в группе. Для наглядности на рис. 2 показано, как возрастает NSD в среднем в каждой группе расчетов из пяти и семи моделей. Поскольку NSD возрастает от модели к модели (которые в каждом расчете были отсортированы в порядке возрастания невязки R), то среднее значение NSD не может быть взято в качестве оценки расчета. Средние значения (горизонтальные линии) также достаточно близки к медианному (разница 1%), поэтому медианное значение также не может рассматриваться как оценка качества расчета. Минимальное значение не было взято из-за его вероятной принадлежности к статистическому выбросу. Таким образом, в качестве оценки качества расчета был выбран

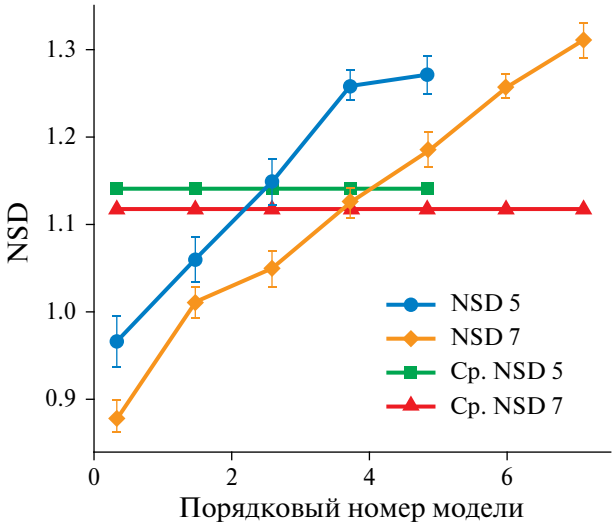

Рис. 2. Усредненные по расчетам отсортированные значения NSD для групп расчетов, состоящих из пяти и семи моделей, и их средние значения (горизонтальные линии).

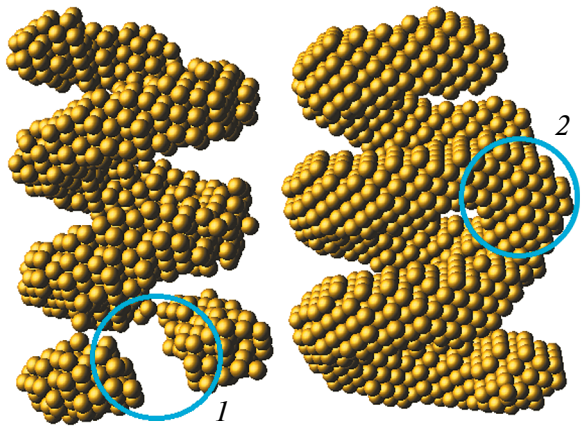

Рис. 3. Примеры найденных структур с дефектом типа “разрыв” (*1*) и типа “соединение” (*2*).

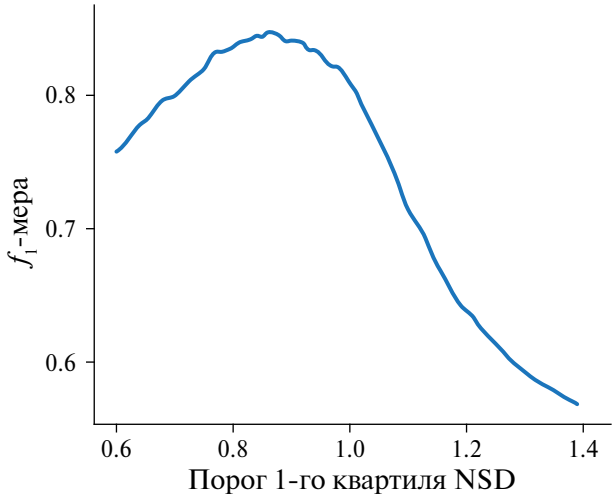

Рис. 4. Оценка f_1-мера в зависимости от порогового значения 1-го квартиля NSD.

1-й квартиль (0.25-квантиль) NSD как наиболее компромиссный.

Для простоты анализа оценки качества расчета были дополнительно конвертированы в бинарную форму на основе значений 1-го квартиля NSD. Для определения разделяющего (порогового) значения 1-го квартиля был принят следующий алгоритм. Условно примем за хорошую модель такую, которая имеет не более одного дефекта. В случае спирали дефектами назовем разрыв в спиральном витке и межвитковое соединение (рис. 3). При одном подобном дефекте спираль все еще удается идентифицировать, и это не накладывает жестких ограничений на значение NSD. Анализ форм полученных спиралей показывает, что при значении NSD < 0.54 решения имеют не более одного дефекта. Очевидно, что если 1-й квартиль NSD также меньше 0.54, то в расчете есть хорошие модели. Однако при таком пороговом значении не учитывается некоторая часть расчетов с хорошими решениями. В то же время при увеличении порогового значения будут учитываться расчеты, в которых все модели плохие (минимальное NSD > 0.54). Таким образом, необходимо выбрать такое пороговое значение 1-го квартиля Q, которое оптимально разделяет расчеты с хорошими и плохими моделями. Для данной задачи бинарной классификации выбрана оценка f_1-мера [15]:

$$f_1 = \frac{2 \cdot TP}{2 \cdot TP + FP + FN},$$

где TP, FP, FN – количество истинно положительных, ложноположительных и ложноотрицательных результатов соответственно. В настоящей работе положительные расчеты (результаты запусков) содержат не менее одной хорошей модели, отрицательные – ни одной хорошей модели. Введение порогового значения Q разбивает положительные расчеты на истинно положительные и ложноотрицательные, а отрицательные расчеты – на истинно отрицательные и ложноположительные. В данном случае TP – количество расчетов, в которых минимальное NSD < 0.54 и 1-й квартиль NSD < Q, FP – количество расчетов, в которых минимальное NSD > 0.54 и 1-й квартиль NSD < Q, FN – количество расчетов, в которых минимальное NSD < 0.54 и 1-й квартиль NSD > Q. Метрика f_1-мера используется для оценки качества классификации данных, классы которых не сбалансированы и ошибки первого и второго рода имеют схожую важность. В этом случае важно минимизировать количество как ложноположительных результатов, так и ложноотрицательных. При идеальной классификации $f_1 = 1$, при невозможности классифицировать данные $f_1 = 0$. При оптимальном разделении расчетов f_1-мера максимальна и равна 0.85 при пороговом значении $Q = 0.86$ (рис. 4). При таком пороговом значении точность, определяемая

как $TP/(TP + FP)$, равна 0.9. Это означает, что случайно выбранный расчет из отобранных с помощью 1-го квартиля как положительный с вероятностью 90% действительно окажется положительным, что является приемлемым результатом. Таким образом, расчеты с 1-м квартилем NSD < 0.86 приняты в качестве положительных, все остальные – в качестве отрицательных. На рис. 5 показаны двумерные графики оценки расчетов P в зависимости от различных параметров алгоритма. Положительные оценки расчетов помечены единицами, отрицательные – нулями. Значения оценки в промежуточных координатах получены с помощью линейной интерполяции по обеим осям. Из-за широкого диапазона значений весов графики построены по логарифмической шкале. Веса дополнительно увеличены для смещения значений логарифма в сторону положительных для удобства восприятия графиков.

Из рис. 5 видно, что чем меньше шаг спирали, тем меньше области положительных расчетов (черный цвет). Для структуры I при $N_{sh} = 20$ отсутствуют положительные расчеты в рамках принятой обработки расчетов. Для этой же структуры при $N_{sh} = 28$ есть несколько обособленных областей положительных расчетов, которые, однако, могут являться результатом случайных выбросов в работе алгоритма.

Выбор оптимального значения степени весовой функции (3) и числа шенноновских каналов (т.е. максимального значения угла измерения)

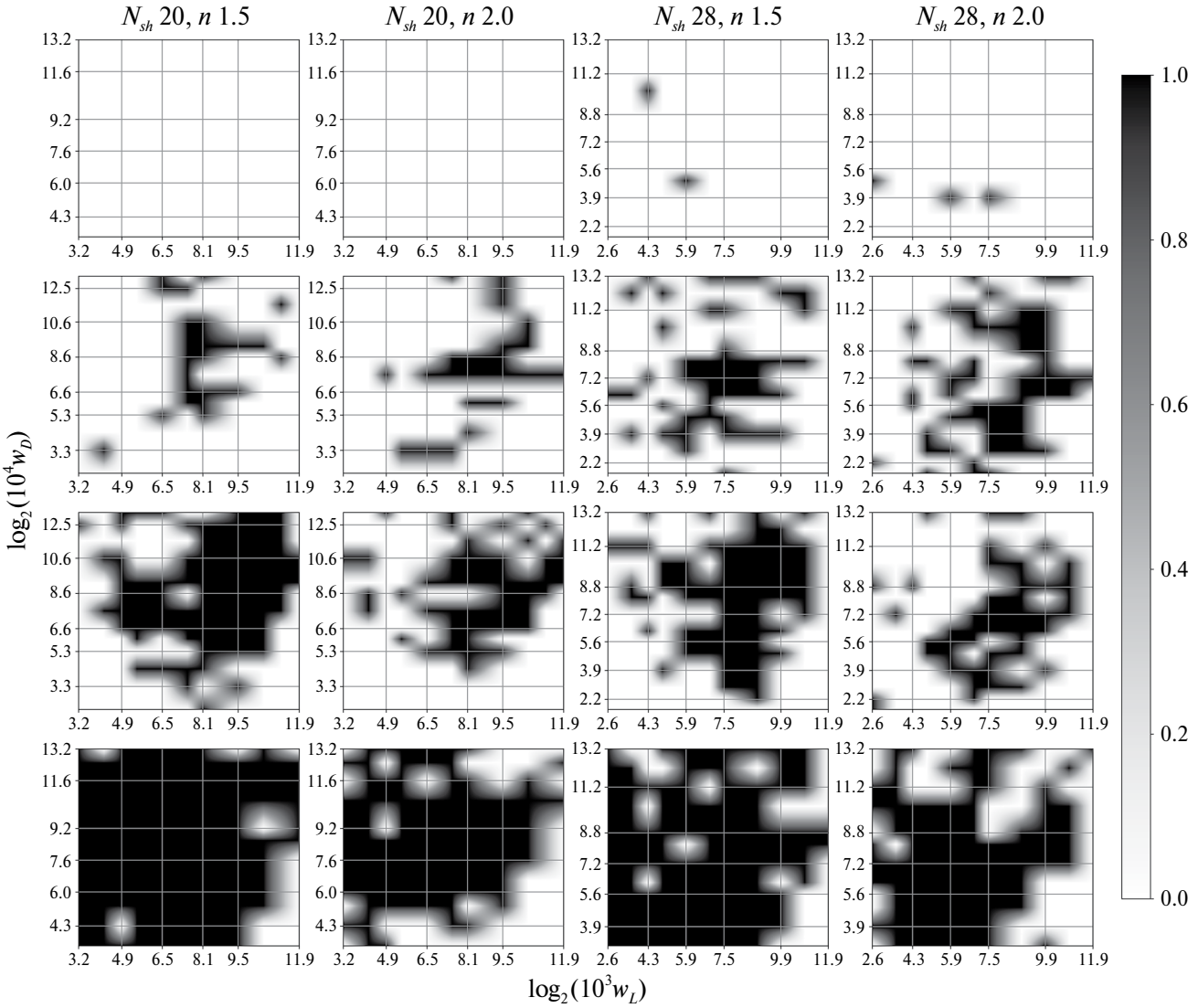

Рис. 5. Значения оценки расчетов в зависимости от различных значений параметров алгоритма для моделей I, II, III и IV (сверху вниз). По оси абсцисс отложены логарифмы весов штрафов за разрыв структуры, по горизонтали – за ее рыхлость. Черный цвет отмечает области успешного поиска.

неоднозначен в рамках используемых в настоящей работе значений, и зависит от выбора веса штрафов за разрывность и рыхлость модели.

Для того чтобы определить оптимальные значения штрафных параметров w_D, w_L для каждой спирали, принято допущение, что эти значения находятся в одном конечном диапазоне. Оптимальные значения определяли путем аппроксимации оценки расчетов прямоугольным двумерным окном сканирования вида

$$P = \begin{cases} 1, \text{если} \left| w_D - w_D^0 \right| \le \Delta w_D \, u \left| w_L - w_L^0 \right| \le \Delta w_L \\ 0 \quad \text{в противном случае,} \end{cases}$$

где $(w_L{}^0, w_D{}^0)$ – координаты центра окна, Δw_D, Δw_L – половины сторон окна. На рис. 6 показаны результаты аппроксимации.

Как было отмечено, для функции парных расстояний $p(r)$ спиралей характерна сильная корреляция, что приводит к появлению нескольких пиков, соответствующих различным структурным элементам спирали (рис. 7). Поэтому для обобщения данных статистического анализа расчетов были построены зависимости оптимальных значений весов, определенных в результате аппроксимации, от разрешения пиков кривой $p(r)$. Функции парных расстояний были вычислены в программе GNOM [16]. В настоящей работе взято разрешение R второго и третьего пиков ($r \approx 68$ и 107 Å):

$$R = \frac{\Delta r}{a_l + a_r}, \tag{5}$$

где Δr – расстояние между пиками, a_l, a_r – полуширины второго пика с левой стороны и третьего пика с правой стороны соответственно. Для структуры I

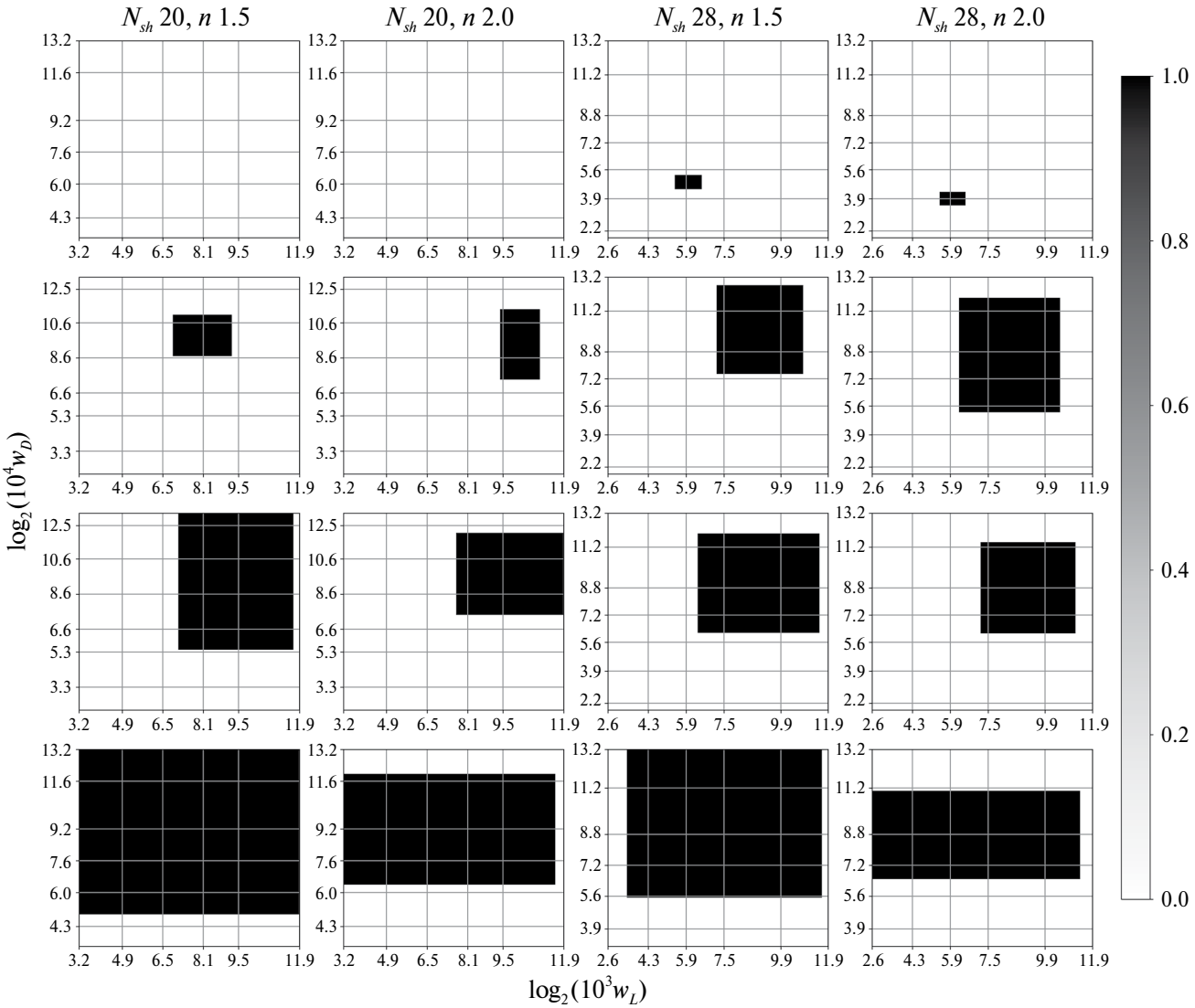

Рис. 6. Аппроксимации оценки расчетов для моделей I, II, III и IV (сверху вниз).

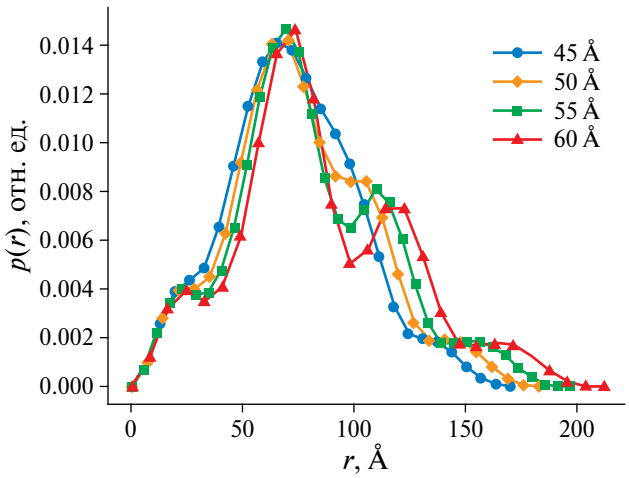

Рис. 7. Графики функций парных расстояний для спиралей с шагом h = 45, 50, 55 и 60 Å (модели I, II, III, IV рис. 1).

(рис. 1) положение и полуширина третьего пика определены из линейной аппроксимации положений и полуширин для других спиралей. Это оправдано, поскольку коэффициент линейной корреляции между шагом спирали и положением третьего пика составляет 0.998, а между шагом спирали и полушириной третьего пика – 0.996. На рис. 8 отображены зависимости координат центра окна сканирования от величины разрешения пиков R для разных значений степени весовой функции и числа шенноновских каналов. Более полная статистика результатов представлена в табл. 1.

Рассмотрим результаты для структур II, III и IV. Оптимальный диапазон веса штрафа за разрывность модели меняется немонотонно в зависимости

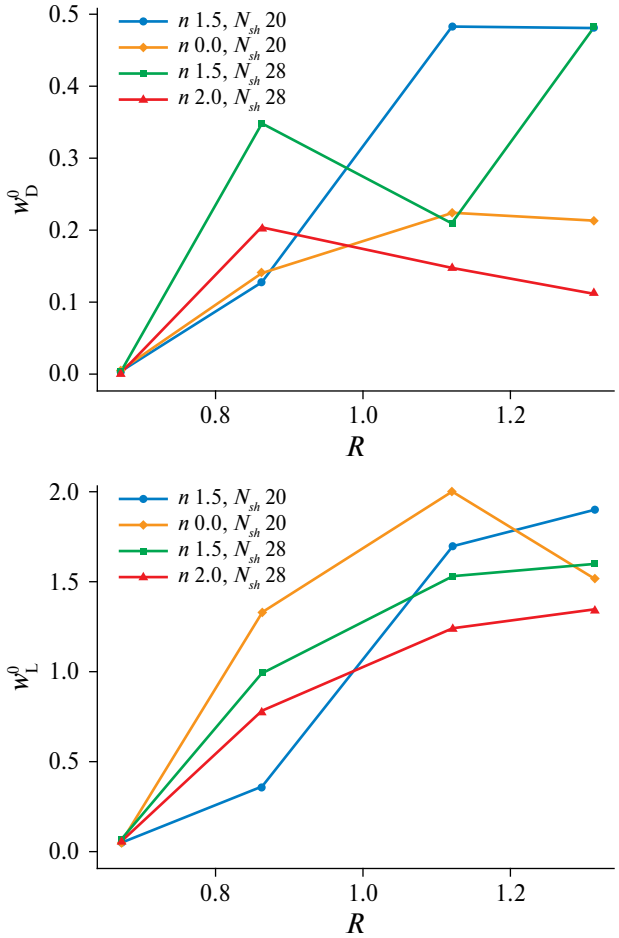

Рис. 8. Оптимальные значения весов штрафов за рыхлость (w_L) и разрывность (w_D) частицы в зависимости от разрешения пиков R на кривой $p(r)$ для различных пар значений степени весовой функции (3) n и числа шенноновских каналов N_{sh}.

Таблица 1. Статистика результатов аппроксимации оценки расчетов из рис. 6

Разрешение 2-го и 3-го пиков на кривой $p(r)$	Степень весовой функции n (уравнение 2)	Число шенноновских каналов, N_{sh}	Оптимальный интервал веса разрывности частицы Δw_D	Оптимальный интервал веса рыхлости частицы Δw_L
0.86	1.5	20	0.128 ± 0.086	0.36 ± 0.24
1.12	1.5	20	0.483 ± 0.478	1.69 ± 1.55
1.31	1.5	20	0.481 ± 0.478	1.89 ± 1.88
0.86	2	20	0.141 ± 0.124	1.32 ± 0.66
1.12	2	20	0.225 ± 0.208	2.00 ± 1.79
1.31	2	20	0.213 ± 0.204	1.52 ± 1.51
0.67	1.5	28	0.003 ± 0.001	0.07 ± 0.03
0.86	1.5	28	0.348 ± 0.330	0.99 ± 0.84
1.12	1.5	28	0.210 ± 0.203	1.53 ± 1.44
1.31	1.5	28	0.483 ± 0.478	1.59 ± 1.58
0.67	2	28	0.002 ± 0.001	0.07 ± 0.02
0.86	2	28	0.205 ± 0.201	0.78 ± 0.70
1.12	2	28	0.148 ± 0.141	1.24 ± 1.09
1.31	2	28	0.111 ± 0.103	1.34 ± 1.33

от разрешения R и зависит от конкретной пары значений n и N_{sh}. При $n = 2.0$ изменения веса более плавные, при $n = 1.5$ – резкие. Это может указывать на взаимосвязь веса штрафа за разрывность частицы и степени весовой функции в отношении данных структур – при $n = 1.5$ оптимальное значение веса становится более чувствительным к изменениям в структуре. Оптимальный интервал значений веса штрафа за рыхлость частицы постепенно увеличивается с ростом разрешения R для всех пар значений n и N_{sh}. Это может быть связано с тем, что при меньшем шаге спирали компактификация формы частицы при поиске может происходить путем наикратчайшего соединения соседних витков, что приводит к неправильной структуре. По этой причине при меньших R ограничения на оптимальный диапазон веса штрафа за рыхлость частицы оказываются более жесткими.

Частичное или полное отсутствие оптимального диапазона весов для структуры I свидетельствует о нестабильности структуры, которая, вероятно, связана с близостью длин отрезков, соединяющих различные структурные части спирали. Поскольку веса варьировались в диапазоне, гораздо большем, чем обычно используется на практике, вероятно, оптимальный диапазон весов может быть найден при других значениях n и N_{sh}, не использованных в настоящей работе, либо при использовании иных подходов. Другой причиной разброса решений при восстановлении формы спирали может являться излишняя плотность расположения шариков. Для структуры I кратчайшее расстояние между витками равно 15 Å, что приблизительно равно двум–трем шарикам при радиусе шариков 3.5 Å. В таких условиях минимизация рыхлости вряд ли сможет способствовать реконструкции спирали. Причина в том, что поверхностные атомы, имея небольшое число контактов, вносят вклад в оценку рыхлости и поэтому наиболее выгодной с точки зрения штрафа за рыхлость может являться структура со слипшимися витками, которая имеет меньшую удельную площадь поверхности. Кроме того, слипание витков наиболее выгодно с точки зрения штрафа за разрывность частицы. В случае истинности данной причины необходимо уменьшение радиуса шариков до минимально возможного, при котором структурный фактор их упаковки еще не влияет на форму кривой рассеяния.

Хотя приведенные выше рассуждения характерны для спиралей, подобные рассуждения можно провести для любых частиц, обладающих следующим свойством: отрезки одной и той же длины, концы которых принадлежат объему частицы, лежат как внутри частицы, так и могут пересекать пустое пространство. В случае со спиралями это отрезки, лежащие внутри одного витка, и отрезки, соединяющие соседние витки. В зависимости от величины весов штрафных членов в (1) в формируемой структуре частицы может преобладать количество тех или иных отрезков. Другим необходимым условием является наличие двух и более пиков на кривой $p(r)$, указывающих на преобладание двух и более средних длин отрезков.

Для проверки эффективности полученной эмпирической зависимости весов штрафов за разрывность и рыхлость частицы от разрешения пиков на кривой $p(r)$ была выбрана трехвитковая спираль с шагом 62 Å, толщиной витков 35 Å и внешним диаметров 120 Å (структура V на рис. 1). Разрешение пиков на кривой $p(r)$ для этой спирали $R = 1.04$. Это значение находится приблизительно посередине между значениями разрешения для структур II и III. Было выбрано по три значения весов рыхлости и разрывности модели: среднее оптимальное, больше и меньше среднего оптимального на величину разброса оптимальных значений. Расчеты повторены $k = 8$ раз, обработаны аналогично основной выборке и усреднены. Средние значения оценки расчетов $P_{\text{ср}} = \sum_{i=1}^{k} P_i / k$ отображены на рис. 9. Точность валидации оценивали по формулам

$$A = \frac{\sum_i E\left(P_i, P_i^0\right) w_i}{\sum_i w_i} \tag{6а}$$

$$E\left(P_i, P_i^0\right) = \begin{cases} 1, & \text{если } P_i = P_i^0 \\ 0, & \text{если } P_i \neq P_i^0 \end{cases}, \tag{6б}$$

где P_i, P_i^0 – реальное и идеальное значения оценки i-го расчета, w_i – вес i-го расчета ($w_i = 8$ для центральной ячейки, $w_i = 1$ для остальных) (рис. 9). Вес для центральной ячейки подобран так, чтобы сбалансировать вклад отклонений результатов каждой ячейки. Точность A фактически отражает взвешенную долю правильных ответов. Для степени весовой функции (3) $n = 1.5$ и $N_{sh} = 20$ точность равна 0.844, для $n = 2.0$ и $N_{sh} = 20$ – 0.766, для $n = 1.5$ и $N_{sh} = 28$ – 0.680, для $n = 2.0$ и $N_{sh} = 28$ – 0.930, что является хорошим результатом. Усредненные оценки в центральных ячейках на рис. 9 оказываются меньше единицы и в нецентральных – больше нуля, так как алгоритм минимизации (“имитации отжига”) имеет стохастический характер. На рис. 10 показаны типичные восстановленные модели спирали V для центральной и правой верхней ячеек из рис. 9. Типичные модели из правых верхних ячеек оказываются плохо восстановленными, что является ожидаемым, поскольку эти ячейки соответствуют неоптимальным весам, тогда как для центральных ячеек восстановленные модели близки к модели идеальной спирали.

ЗАКЛЮЧЕНИЕ

В работе проведен анализ результатов восстановления трехмерной формы спиральной

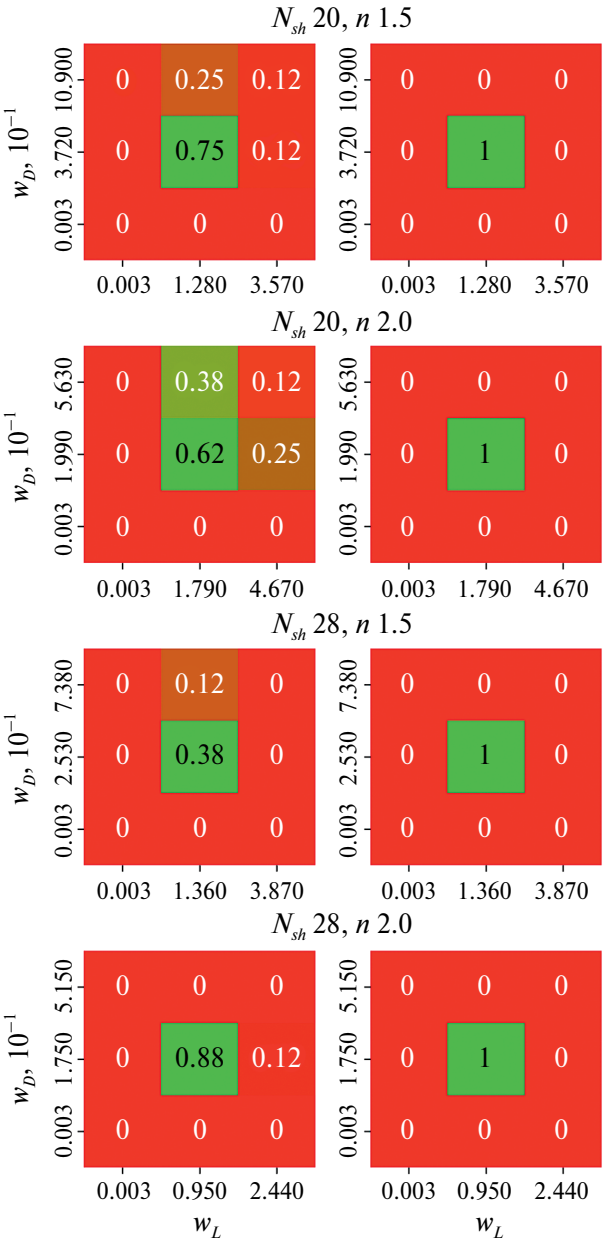

Рис. 9. Усредненные значения оценки расчетов для тестовой модели V. Слева – реальный результат, справа – идеальный.

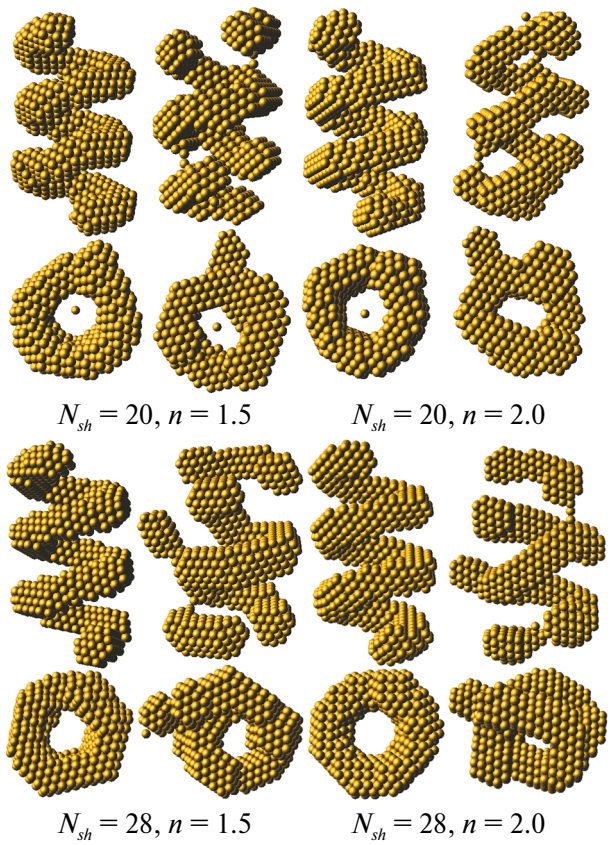

Рис. 10. Типичные восстановленные тестовые модели V для центральной (столбцы 1, 3) и правой верхней (столбцы 2, 4) ячеек из рис. 9.

трехвитковой частицы по данным МУРР с помощью модифицированного алгоритма “имитации отжига”. По результатам статистической обработки наборов решений определена взаимосвязь между качеством восстановления структуры, параметрами алгоритма поиска и разрешением двух пиков на кривой функции парных расстояний $p(r)$. Установлены эмпирические зависимости оптимальных значений весов штрафов за разрывность и рыхлость моделей от параметров кривой распределения $p(r)$ для двух разных значений степени весовой функции, применяемой к кривой интенсивности рассеяния, и двух значений числа шенноновских каналов. Предсказаны оптимальные параметры алгоритма поиска, на основании которых реконструирована трехмерная форма спирали с характеристиками, отличными от характеристик спиралей, использованных для статистического исследования. Результаты реконструкции тестовой спирали соответствуют результатам статистического исследования, что указывает на корректность полученных эмпирических зависимостей и возможность их практического применения.

Предложенная стратегия для алгоритма поиска может быть применена и к другим структурам (не обязательно спиралям), если функции парных расстояний для них имеют два и более максимумов либо один максимум с наличием точки перегиба. В последнем случае можно провести аппроксимацию функции парных расстояний суммой двух гауссовских функций для оценки их относительной полуширины. Далее необходимо выбрать значения степени весовой функции для интенсивности рассеяния и числа шенноновских каналов, при этом можно руководствоваться значениями, приведенным в табл. 1. После этого следует, также руководствуясь значениями в таблице,

оценить оптимальный диапазон штрафных весов за рыхлость и разрывность частицы. Если используемый экспериментальный угловой диапазон кривой рассеяния меньше 20 или больше 28 шенноновских каналов, то потребуется дополнительно экстраполировать значения весов на имеющееся число каналов. Из полученных диапазонов значений весов штрафов за рыхлость и разрывность частицы при проведении восстановления формы рекомендуется выбирать их центральные значения и значения на расстоянии четверти длины диапазона от границ интервала.

Работа выполнена в рамках государственного задания НИЦ “Курчатовский институт”.

СПИСОК ЛИТЕРАТУРЫ

1. *Свергун Д.И., Фейгин Л.А.* Рентгеновское и малоугловое рассеяние. М.: Наука, 1986. 280 с.
2. *Svergun D.I., Stuhrmann H.B.* // Acta Cryst. A. 1991. V. 47. P. 736. https://doi.org/10.1107/S0108767391006414
3. *Svergun D.I., Volkov V.V., Kozin M.B. et al.* // Acta Cryst. A. 1996. V. 52. P. 419. https://doi.org/10.1107/S0108767396000177
4. *He H., Liu C., Liu H.* // iScience. 2020. V. 23. 100906.
5. *Svergun D.I.* // Biophys J. 1999. V. 78. P. 2879. https://doi.org/10.1016/S00063495(99)77443-6
6. *Franke D., Svergun D.I.* // J. Appl. Cryst. 2009. V. 42. P. 342. https://doi.org/10.1107/S0021889809000338
7. *Kirkpatrick S., Gelatt C.D., Vecchi M.P.* // Science. 1983. V. 220. P. 671. https://doi.org/10.1126/science.220.4598.671
8. *Григорьев В.А., Конарев П.В., Волков В.В.* // Кристаллография. 2023. Т. 68. С. 941. https://doi.org/10.31857/S0023476123600295
9. *Волков В.В.* // Кристаллография. 2021. Т. 66. С. 793. https://doi.org/10.31857/S0023476121050234
10. *Григорьев В.А., Конарев П.В., Волков В.В.* // Успехи в химии и химической технологии. 2022. Т. 36. С. 53
11. *Rees D.A.* Polysaccharides Shapes. London: Chapman and Hall, 1977. 80 p.
12. *Shtykova E.V., Volkov V.V., Konarev P.V. et al.* // J. Appl. Cryst. 2003. V. 36. P. 669. https://doi.org/10.1107/S0021889803006198
13. *Shannon C.E., Weaver W.* The Mathematical Theory of Communication. University of Illinois Press, 1949. 125 p.
14. *Kozin M., Svergun D.* // J. Appl. Cryst. 2001. V. 34. P. 33. https://doi.org/10.1107/S0021889800014126
15. *Taha AA., Hanbury A.* // BMC Med Imaging. 2015. V. 15. P. 29. https://doi.org/10.1186/s12880-015-0068-x
16. *Svergun D.I.* // J. Appl. Cryst. 1992. V. 25. P. 495.

FEATURES OF THREE-DIMENSIONAL RECONSTRUCTION OF SPIRALS BASED ON SMALL-ANGLE X-RAY SCATTERING DATA

V. A. Grigorev*, P. V. Konarev, V. V. Volkov

Shubnikov Institute of Crystallography of Kurchatov Complex of Crystallography and Photonics of NRC “Kurchatov Institute”, Moscow 119333, Russia

**e-mail: vasiliy.grigorev.1996@mail.ru*

Abstract. The interest in spiral particles lies in their resemblance to authentic nanostructures that emerge through the self-organisation of biopolymers (such as carrageenans, DNA, and so forth). Conversely, the determination of the structural parameters of such particles based on small-angle scattering data is challenging due to the lack of conditioning in the inverse problem. This is demonstrated by the utilisation of established bead structure modelling software. This paper considers a modification of the search algorithm in a limited area of space and the behaviour of solutions depending on the values of the parameters of the objective function responsible for the connectivity and looseness of the structure, the type of weighing of the scattering intensity curve, and the width of the angular range of data. In order to statistically assess the stability of the solutions, a sequential model search mode was applied, with varying amounts of contributions of penalty terms. The empirical dependences of the optimal values of the search parameters with respect to the parameters of the distribution curve of paired distances were determined.