ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ УЧРЕЖДЕНИЕ «ФЕДЕРАЛЬНЫЙ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР «КРИСТАЛЛОГРАФИЯ И ФОТОНИКА» РОССИЙСКОЙ АКАДЕМИИ НАУК»

ИНСТИТУТ КРИСТАЛЛОГРАФИИ ИМ. А.В. ШУБНИКОВА РАН

На правах рукописи

КРЮКОВА АЛЁНА ЕВГЕНЬЕВНА

КОМБИНИРОВАННЫЙ ПОДХОД К ПОИСКУ РАСПРЕДЕЛЕНИЙ РАЗМЕРОВ СФЕРИЧЕСКИХ НАНОЧАСТИЦ ПО ДАННЫМ МАЛОУГЛОВОГО РЕНТГЕНОВСКОГО РАССЕЯНИЯ

Диссертация

на соискание ученой степени кандидата физико-математических наук по специальности 1.3.8 - физика конденсированного состояния

> Научный руководитель: Кандидат физико-математических наук Конарев Петр Валерьевич

Москва 2022

оглавление

ВВЕДЕНИЕ	5
<u>ГЛАВА 1.</u> ЛИТЕРАТУРНЫЙ ОБЗОР	13
1.1. Краткая история развития метода малоуглового рассеяния	13
1.2. Физические основы малоуглового рентгеновского рассеяния	20
1.3. Основные алгоритмы для анализа полидисперсных систем	23
1.3.1. Поиск функции распределения частиц по размерам с	
использованием регуляризации по Тихонову	27
1.3.2. Прямой поиск гистограммы распределения по разме-	
рам. Алгоритм Левенберга-Марквардта	29
1.3.3. Параметрическое задание функции распределения по	
размерам	31
1.3.4. Преобразование Титчмарша-Фокса для вычисления	
распределения частиц по размерам	33
1.4. Методика измерений малоуглового рассеяния и эксперимен-	
тальные установки	36
1.5. Возможности метода малоуглового рассеяния. Преимущества	
и ограничения	39
1.6. Информативность данных и проблема неоднозначности реше-	
ния в малоугловом рассеянии	41
1.7. К понятию устойчивости решений	45
1.8. Методы оптимизации, рассмотренные в исследовании	46
1.8.1. Градиентный метод Бройдена-Флетчера-Гольдфарба-	
Шанно	47
1.8.2. Метод многогранника Нелдера-Мида	48
1.8.3. Метод моделирования отжига	50
1.9. Примеры использования функции распределения по размерам	
для анализа олигомерного состава растворов биомакромолекул	51
Заключение к главе 1	55

<u>ГЛАВА 2.</u> ИССЛЕДОВАНИЕ УСТОЙЧИВОСТИ РЕШЕНИЙ	
ПРИ АНАЛИЗЕ ТЕОРЕТИЧЕСКИХ ПОЛИДИСПЕРСНЫХ СИ-	
СТЕМ МЕТОДОМ МАЛОУГЛОВОГО РАССЕЯНИЯ. ЗАВИСИ-	
мость решения от параметров модели и шумов	
ДАННЫХ	57
2.1. Системы с неперекрывающимися распределениями сфериче-	
ских частиц по размерам	57
2.2. Системы с частично перекрывающимися распределениями	
сферических частиц по размерам	72
Заключение к главе 2	75

ГЛАВА З. РАЗРАБОТКА ЭФФЕКТИВНОГО АЛГОРИТМА ПО-ИСКА РЕШЕНИЙ ПРИ АНАЛИЗЕ ДАННЫХ МАЛОУГЛОВОГО 76 РАССЕЯНИЯ 76 3.1. Оценка эффективности работы алгоритмов поиска функции 76 распределения на примере двух- и трехкомпонентных полидисперсных систем. 77 3.2. Новая схема комбинированного использования алгоритмов поиска функции распределения. 83 Заключение к главе 3. 83

4.1.1. Кремнезоли. Полидисперсная система сферических ча-	
стиц	85
4.2. Результаты применения методов оптимизации, рассмотрен-	
ных в исследовании	93
4.2.1. Комбинирование методов оптимизации. Градиентный	
метод Бройдена-Флетчера-Гольдфарба-Шанно и метод моде-	
лирования отжига	93
4.2.2. Исследование устойчивости решений при анализе дан-	
ных малоуглового рентгеновского рассеяния от раствора	
кремнезоля с использованием ряда оптимизационных схем	94
Заключение к главе 4	99
Основные результаты и выводы	102
<u>Приложение А.</u> Анализ экспериментальных данных малоуглового	
рентгеновского рассеяния от раствора кремнезоля	105
<u>Приложение Б.</u> Распределения размеров наночастиц оксида цинка	
по данным просвечивающей электронной микроскопии и малоуг-	
лового рентгеновского рассеяния	109
Список сокращений и условных обозначений	111

Список работ автора по теме диссертации	112
Список цитируемой литературы	113

ВВЕДЕНИЕ

Актуальность темы.

Многие современные функциональные материалы представляют собой многокомпонентные системы, свойства которых определяются структурными особенностями и/или включениями в наноразмерном диапазоне. Метод малоуглового рассеяния (МУР) является эффективным, и часто единственным методом, позволяющим качественно и количественно исследовать строение таких материалов без специальной подготовки образцов. МУР [1-3] – является одним из самых востребованных структурных методов, он позволяет исследовать упорядоченные, частично упорядоченные и неупорядоченные объекты размерами 1-200 нм, такие как монодисперсные системы (растворы идентичных частиц), полидисперсные системы (в том числе нанокомпозиты, гели, аэрозоли, полимеры, сплавы, мезопористые материалы и др.). В случае монодисперсных систем с помощью МУР можно восстанавливать форму и структуру частиц с низким разрешением 1-2 нм. Для полимерных систем можно определять размер макромолекулярных клубков, степень кристалличности, ориентацию цепей. МУР широко применяют для определения удельной пористости. Так как метод не требует специальной подготовки образцов, его используют для анализа динамики структур при изменении внешних условий, таких как pH, температуры, освещения и т.д. Метод МУР активно применяется в физике, биологии, медицине, материаловедении, различных отраслях промышленности.

МУР основан на использовании эффекта когерентного упругого рассеяния монохроматического излучения на неоднородностях вещества [1]. Обычно эксперименты по МУР проводят с использованием рентгеновских или нейтронных пучков с длинами волн в диапазоне от долей ангстрема до нескольких ангстремов. В малоугловом рентгеновском рассеянии (**МУРР**) рассеяние происходит на электронах, а в малоугловом рассеянии нейтронов (**МУРН**) на атомных ядрах. Теория дифракции одинакова и для рентгеновских лучей, и для нейтронов, отличия заключаются лишь в особенностях взаимодействия излучения с веществом. Поэтому все полученные в данной работе результаты и разработки могут быть использованы как для анализа данных МУРР, так и МУРН.

Для большинства изучаемых объектов методом МУР задача состоит не только в реконструкции структуры рассеивающих объектов, но также в определении функций распределений их по размерам. В основном это относится к полидисперсным системам, исследование которых является очень важной задачей для развития нанотехнологий. Несмотря на наличие разработанных методов анализа для монодисперсных систем, для полидисперсных систем задача становится более неоднозначной и трудной. При восстановлении профилей распределений часто можно найти несколько различных решений при почти одинаковом критерии качества соответствия экспериментальным данным, что свойственно некорректно поставленным задачам и существенно усложняет получение решения (задача считается корректно поставленной, когда ее решение единственно и устойчиво относительно возмущений исходных данных). Цель работы состояла в поиске таких оптимизационных схем, чтобы невыполнение условия однозначности не приводило к серьезным нарушениям условия устойчивости.

Дополнительной проблемой при поиске распределений является плохая обусловленность обратной задачи: число обусловленности матрицы вторых производных минимизируемой целевой функции может достигать десятков и сотен миллионов (верхняя граница возможной относительной ошибки решения по сравнению с ошибками в данных). При этом решение зависит не только от ошибок измерений, но и от реализации алгоритма поиска, и от величин стартовых значений параметров модели. Поэтому анализ устойчивости нахождения распределений частиц по размерам по данным МУР приобретает исключительно важный характер. В работе предложен подход, позволяющий повысить устойчивость решений обратной задачи.

В настоящее время существует несколько методов определения функции распределения частиц по размерам, в частности реализованные в программах

GNOM [4] и MIXTURE [5], они применяются во всем мире, так как входят в пакет программ ATSAS [6].

В программе прямого поиска распределения методом регуляризации GNOM в качестве формфактора используются только один тип частиц, чаще всего – сферы. Но если изменить тип частиц, например, на цилиндры, то задача становится более неопределенной, так увеличивается число параметров модели – к радиусу добавляется длина частицы. Кроме того, в области малых размеров возможны артефакты из-за ограничений на кривизну контура распределения. Также решению свойственны артефакты в виде паразитных осцилляций из-за эффектов обрыва как кривой рассеяния, так и ограничения на максимальный размер частиц. С другой стороны, в программе MIXTURE, которая аппроксимирует распределение суперпозицией априори гладких функций, есть возможность задавать до 10 различных компонент (форму частиц в которых описывают простыми геометрическими телами) и учитывать межчастичное взаимодействие. Однако неоднозначность решения при этом возрастает. И если об исследуемой системе заранее ничего не известно, для каждой компоненты приходится эмпирически подбирать коридоры значений параметров модели, что зачастую затрудняет поиск правильного решения. Понятия "правильности", "точности" и "надежности" решений будут рассмотрены далее.

В работе представлен новый комбинированный подход, позволяющий с использованием этих программ существенно расширить область сходимости к правильному решению в многомерном пространстве стартовых параметров модели и, таким образом, улучшить надежность анализа сложных полидисперсных систем по данным малоуглового рассеяния.

Работа была выполнена в приближении сферических частиц, поскольку оно применяется для широкого круга объектов и для них неоднозначность восстановления формы обычно наименьшая. Согласно работе [7] для анизометричных частиц степень неоднозначности становится гораздо выше, даже в случае монодисперсных систем.

Цель работы.

Систематическое исследование устойчивости восстановления распределений частиц по размерам для смесей полидисперсных сферических частиц и разработка новых оптимизационных схем, позволяющих существенно расширить диапазон сходимости к правильному решению в многомерном пространстве параметров.

В работе поставлены следующие задачи.

1. Разработка эффективных методов поиска распределений по размерам наночастиц по данным малоуглового рассеяния в многокомпонентных системах.

2. Сравнение эффективности работы наиболее результативных минимизационных алгоритмов и нахождение оптимальной схемы их комбинированного использования.

3. Исследование устойчивости решений обратной задачи определения размерных параметров наночастиц по данным малоуглового рассеяния для модельных и реальных полидисперсных систем.

4. Исследование влияния шумовой составляющей данных малоуглового рассеяния на результаты поиска распределения частиц по размерам.

Научная новизна и практическая значимость работы.

В работе впервые:

- проведено систематическое исследование устойчивости решений задачи поиска распределений частиц по размерам для смесей полидисперсных сферических частиц,
- разработана новая оптимизационная схема, основанная на комбинации взаимодополняющих методов высокоэффективного градиентного поиска и моделирования отжига,
- предложена схема комбинированного использования алгоритмов, сочетающая методы линейных и нелинейных наименьших квадратов.

Разработанные подходы позволили существенно расширить диапазон сходимости к правильному решению в многомерном пространстве параметров и, таким образом, улучшить эффективность (и, как следствие, надежность) анализа сложных полидисперсных систем по данным малоуглового рассеяния.

На основании проведенных исследований даны рекомендации, позволяющие находить устойчивые решения при широком диапазоне возможных начальных значений параметров модели.

Положения, выносимые на защиту:

- Новая схема комбинированного использования алгоритмов поиска распределений наночастиц по размерам, включающая в себя анализ методами регуляризации, прямого поиска гистограммы и в виде суперпозиции гладких аналитических функций, позволяющая расширить диапазон сходимости к точному решению обратной задачи в пространстве параметров, описывающих структурную модель исследуемой системы во всех рассмотренных случаях с точностью до 1% по параметрам распределения.
- 2. Программная реализация предложенного подхода с использованием поэтапного сочетания квазиньютоновского градиентного метода в варианте Бройдена-Флетчера-Гольдфарба-Шанно (BFGS) и метода моделирования отжига (SA). Работоспособность программы опробована в ходе проведенного в работе решения большой серии модельных задач.
- 3. Результаты анализа эффективности четырех различных оптимизационных схем: (метода BFGS, метода многогранника Нелдера-Мида, метода SA и комбинации методов BFGS и SA) на примере данных МУР от раствора кремнезоля, подтверждающие более высокую эффективность предложенной комбинированной схемы.
- Результаты восстановления распределений для модельных полидисперсных систем, состоящих из сферических частиц как с неперекрывающимися, так и с частично перекрытыми распределениями

компонентов, а также для экспериментальных данных малоуглового рассеяния от раствора кремнезоля, состоящего из двух типов сферических частиц.

5. Результаты исследования влияния шумовой составляющей в данных малоуглового рассеяния на результаты поиска распределения частиц по размерам, выявившие факт расширения области допустимых стартовых значений параметров минимум на 10% при наличии пуассоновского шума в данных рассеяния с относительным стандартным отклонением в области малых значений интенсивностей до 25%. Показано, что наличие более равномерно распределенного по угловому диапазону гауссовского шума приводит к уменьшению диапазонов сходимости.

Личный вклад автора:

Все результаты, представленные в диссертационной работе, получены лично автором или при его непосредственном участии.

Автором проведена вся обработка данных МУР от многокомпонентных систем с использованием линейных и нелинейных методов наименьших квадратов, а также прямого поиска распределений частиц по размерам в виде гистограммы, представленная в работе.

Автор выполнил систематическое исследование устойчивости методов минимизации, предварительно рассчитав теоретические кривые малоуглового рассеяния от модельных систем сферических частиц.

Автор непосредственно участвовал в процессе разработки модифицированного алгоритма поиска распределения частиц по размерам в программе MIXTURE и её варианте POLYMIX в части автоматизации подбора стартовых значений параметров.

Постановка задачи, обсуждение полученных результатов осуществлялись совместно с научным руководителем, соавторами публикаций и сотрудниками лаборатории.

Апробация результатов работы:

Материалы по результатам диссертационной работы были представлены на следующих конференциях:

- 1. Первый Российский кристаллографический конгресс (РКК-2016, 21-26 ноября 2016, Москва).
- Лазерные и плазменные исследования и технологии (ЛаПлаз-2017, 24-27 января 2017, Москва).
- 51-я Зимняя Школа по физике конденсированного состояния (ФКС 2017, 11-16 марта 2017, Санкт-Петербург).
- 4. Белки и пептиды (18-22 сентября 2017, Москва).
- 5. 52-я Зимняя Школа по физике конденсированного состояния (ФКС 2018, 12-17 марта 2018, Санкт-Петербург).
- Проблемы сольватации и комплексообразования в растворах (1-6 июля 2018, Суздаль).
- Лазерные и плазменные исследования и технологии (ЛаПлаз-2019, 12-15 февраля 2019, Москва).
- 53-я Зимняя Школа по физике конденсированного состояния (ФКС 2019, 11-16 марта 2019, Санкт-Петербург).
- Лазерные и плазменные исследования и технологии (ЛаПлаз-2020, 11-14 февраля 2020, Москва).
- 10.54-я Зимняя Школа по физике конденсированного состояния (ФКС 2020, 16-21 марта 2020, Санкт-Петербург).
- 11.Современные методы анализа дифракционных данных и актуальные проблемы рентгеновской оптики (01-11.07.2020, Москва, Санкт-Петербург).
- 12. Лазерные и плазменные исследования и технологии (ЛаПлаз-2021, 23-26 марта 2021, Москва).
- 13. Проблемы сольватации и комплексообразования в растворах (20-24 сентября 2021, Иваново).

Публикации по теме диссертации:

Результаты исследований по теме диссертации опубликованы в 8 работах [A1-A8].

Структура и объем диссертации:

Диссертация состоит из введения, четырех глав, выводов, двух приложений, списка сокращений и условных обозначений, списка цитируемой литературы и списка публикаций по материалам работы. Общий объем диссертации составляет 125 страниц, включая 16 рисунков и 8 таблиц и список литературы из 112 цитируемых работ, а также 8 публикаций автора по теме диссертации.

Благодарности:

Автор приносит глубокую благодарность научному руководителю к.ф.м.н. **П. В. Конареву** за чуткое руководство, неоценимую помощь и поддержку в процессе подготовки работы.

Автор выражает благодарность д.х.н. В. В. Волкову, д.ф.-м.н. В. Е. Асадчикову за научные консультации и ценные советы.

Автор также признателен д.х.н. Э.В. Штыковой, к.ф.-м.н. Янусовой Л.Г., всем сотрудникам лаборатории за обсуждение результатов, полезные замечания и внимание к работе.

ГЛАВА 1. ЛИТЕРАТУРНЫЙ ОБЗОР

1.1 Краткая история развития метода малоуглового рассеяния

Метод МУР не является новым, он имеет долгую историю, которая начинается в начале XX века после открытия рентгеновских лучей немецким ученым В. К. Рентгеном в 1895 году (за открытие дифракции рентгеновских лучей на трехмерной кристаллической решетке он получил первую в истории науки Нобелевскую премию по физике в 1901 году). После этого потребовалось не более двух десятилетий научных экспериментов и дискуссий о природе рентгеновского излучения, пока В. Фридрих и П. Книппинг по предложению М. Лауэ провели опыт, в результате которого они наблюдали первую картину дифракции рентгеновских лучей в кристалле [8]. Ученые обнаружили, что лучи, описанные В. Рентгеном, представляют собой особый вид электромагнитных волн с очень короткими длинами волн. М. Лауэ в 1914 году получил Нобелевскую премию за это открытие, которое А. Эйнштейн считал самым красивым экспериментом XX века.

Вскоре после этого в 1913 году теоретическое обоснование дифракции рентгеновских лучей в кристалле дали ученые из Великобритании У. Г. Брэгг и У. Л. Брэгг, отец и сын [9]. В 1915 году они удостоились Нобелевской премии за свои труды. На тот момент У.Л. Брэггу было 25 лет, он стал самым молодым лауреатом Нобелевской премии по физике, и остается таковым до сих пор. Независимо от Брэггов в 1913 году российский ученый Г. В. Вульф вывел условия интерференционного отражения рентгеновских лучей от кристаллов, за что также заслужил Нобелевскую премию по физике 1915 года. Г. В. Вульф оставил свой след в развитии кристаллографии, положив начало рентгеновской спектроскопии, тем самым он был первым в России, кто начал проводить рентгеноструктурные исследования. Открыт закон Брэгга-Вульфа.

Позже, в 1929 году, в своих рентгеновских экспериментах индийские физики П. Кришнамурти и Ч. В. Раман исследовали графитовые порошки с помощью дифракции рентгеновских лучей и наблюдали сильные интенсивности рассеяния вблизи первичного луча, которые становились более выраженными с уменьшением размеров частиц порошка [10]. Аналогичный сигнал рассеяния вблизи первичного луча наблюдался от образцов аморфного углерода различного происхождения [11]. Несколькими годами позже Б. Э. Уоррен сообщил о том же явлении, которое он наблюдал при измерении интенсивности рассеяния рентгеновских лучей на образцах сажи [12]. В то время и Уоррен, и Кришнамурти уже знали и осознавали тот факт, что измеренные интенсивности диффузного малоуглового рассеяния вблизи начала координат связаны с размером мелких частиц, присутствующих в образце.

Но реальное развитие малоуглового рассеяния началось с работ французского ученого А. Гинье. В 1938 году во время работы над своей .докторской диссертацией он заметил, что в центре дифракционной картины от сплавов металлов содержится довольно сильное рассеяние при наличии в веществе частиц размерами в десятки нанометров, после чего стал более систематично исследовать рассеяние рентгеновских лучей на малых углах [13]. Гинье разработал количественную теоретическую основу для интерпретации и понимания наблюдаемого диффузного рассеяния рентгеновских лучей вблизи первичного луча. Гинье также обнаружил, что независимо от формы частиц рассеяние вблизи центральной части дифракционной картины может быть приближено экспоненциальной функцией, что послужило основанием для его известной теоремы, связывающей интенсивность рассеяния с радиусом инерции частицы, которая до настоящего времени остается одним из основных этапов при первичной обработке данных МУР. Кроме того, он понял, что белки, белковые комплексы и другие биологические макромолекулы в растворе будут представлять собой идеальные системы для изучения с помощью нового на тот момент метода. С одной стороны, потому что их типичные диапазоны размеров совпадают с размерами, доступными для метода МУР, с другой стороны, потому что они могут быть получены с очень высокой степенью чистоты, которая необходима для анализа.

В последующие годы вклад в развитие теоретической основы для интерпретации данных МУР внесли ряд ученых, в том числе Ж. Фурне, О. Кратки, Г. Пород. Гинье совместно с Фурне опубликовали первую монографию по МУР [1], в которой показали, что метод дает информацию не только о размере и форме рассеивающих объектов, но и о внутренней структуре разупорядоченных или частично упорядоченных систем.

Все вышеперечисленные теоретические разработки были основаны на экспериментах по рассеянию рентгеновских лучей, тогда как открытие нейтрона Д. Чедвиком [14] и первые демонстрации его волновых свойств путем дифракции [15, 16] произошли спустя четыре десятилетия после открытия рентгеновских лучей. Э. Воллан и К. Шулл провели систематические исследования дифракции нейтронов [17] и заложили основу для МУРН как эффективного инструмента для исследований биологических макромолекулярных растворов. Однако, низкая интенсивность потока и низкое энергетическое разрешение доступных источников нейтронов делали использование МУРН для исследования биологических растворов несколько более сложным, чем в случае рентгеновских лучей, и только в конце 1960-х годов впервые были представлены результаты исследований на белках в растворе с использование МУРН [18].

Несмотря на то, что метод МУР поначалу был признан физическим сообществом как полезный инструмент для исследования растворов наночастиц, технические ограничения делали его экспериментальное применение в 1950-х и 1960-х годах довольно трудным, поскольку измеренные данные демонстрировали низкое отношение сигнал/шум или требовали очень большое время накопления сигнала.

Ситуация начала постепенно меняться к 1970-1980-м годам с появлением высокоинтенсивных нейтронных источников, а также с ростом применения синхротронного излучения в качестве источника рентгеновского излучения. Новые источники излучения обладали достаточной яркостью и позволяли получать хорошо сколлимированные пучки с высокой интенсивностью потока (на несколько порядков превышающим возможности лабораторных рентгеновских источников). Дальнейшие технические достижения привели к разработке одно- и двумерных позиционно-чувствительных детекторов рентгеновского излучения [19], которые значительно повысили эффективность сбора данных и позволили значительно улучшить отношение сигнал/шум.

Параллельно с этими технологическими достижениями и прогрессом, достигнутым в области оборудования, были также сделаны значительные шаги вперед в разработке новых подходов в анализе данных МУР. Одним из таких шагов было введение сферических гармоник в обработку данных МУР [20-22] и их применение для восстановления формы частиц по измеренной интенсивности рассеяния [23]. Другим важным теоретическим достижением того времени было введение метода непрямого преобразования Фурье [2]. Это позволило надежно и точно определить функцию распределения парных расстояний изолированного рассеивающего объекта. Следующим ключевым достижением, особенно важным для МУРН, было совершенствование метода вариации контраста [24, 25] и его применение, в частности, в биологии. Хотя метод впервые был разработан и систематически применялся в МУРР, он оказался гораздо более эффективным и чаще применяемым в МУРН (в особенности для растворов комплексов белок-ДНК и белок-РНК) из-за большой разницы в типе рассеяния между водородом и дейтерием – водород рассеивает, в отличие от ядер дейтерия, в значительной степени некогерентно.

Но несмотря на значительный прогресс в развитии МУР (приборной части и методов анализа данных), а также на значительные успехи в исследовании крупных многокомпонентных комплексов, таких как субъединицы рибосом, применение метода МУР в 1970-1980х годах в биологии все еще оставалось ограниченным.

Эта ситуация начала меняться примерно с середины 1990-х годов. В этот момент применение МУР в исследованиях структурной биологии резко возросло, и его признание в качестве мощного и эффективного инструмента структурных исследований начало стремительно расширяться. В конечном

итоге это изменение было вызвано сочетанием нескольких ключевых факторов. С одной стороны, увеличился доступ к высококачественным станциям МУРР на растущем числе синхротронных и нейтронных источников по всему миру. Во-вторых, у биофизического сообщества возникла потребность в изучении больших многодоменных комплексов биомолекул и систем с гибкими частями или неструктурированными концами, имеющими функциональную значимость. В то время как кристаллографические подходы для таких систем были сложными, а иногда просто невозможными из-за отсутствия кристаллов, МУРР могло предоставить структурную информацию для этих макромолекул в растворе, хотя и с низким разрешением. Третий и, возможно, наиболее влиятельный фактор для такого резкого роста приложений МУР в структурной биологии — это разработка и доступность программного обеспечения для анализа данных с удобным интерфейсом вместе с легким доступом к вычислительным ресурсам, достаточно мощным для получения быстрого результата.

Особенно важным шагом в развитии МУР была возможность восстанавливать форму частицы непосредственно по кривой рассеяния, не ограничиваясь простой геометрической моделью. Этот *ab-initio* подход позволил восстанавливать трехмерные модели макромолекулярной структуры с нанометровым разрешением. Хотя разрешение, полученное этим методом, было довольно ограниченным (порядка 10-15 ангстрем), оно, тем не менее, позволяло изобразить общую форму частицы и проводить сравнение со структурными данными с более высоким разрешением, если они были доступны (например, кристаллографическими структурами или структурами ЯМР), и такое сравнение моделей в конечном итоге обеспечивало лучшее понимание функциональных аспектов исследуемых макромолекул и комплексов.

Первый такой *ab-initio* метод восстановления структуры по данным МУР использовал разложение формы частицы по сферическим гармоникам. Хотя этот метод уже был разработан в начале 1970-х годов [21, 22], он был доработан, реализован и стал доступным для биологического сообщества в 1990-х годах [26, 27].

Этот метод в принципе применим к частицам произвольной формы, но он не позволяет адекватно воспроизводить топологически более сложную геометрию частиц, например, для частиц с внутренними полостями. Впоследствии был предложен более широко применяемый метод, основанный на моделировании шариковыми структурами [28].

При таком подходе частица моделируется как набор большого количества шариков, описывающих структуру частицы. Затем расположение шариков в пространстве итеративно модифицируется, чтобы найти модель, которая наилучшим образом соответствует экспериментальным данным рассеяния. Различные версии этого общего алгоритма были реализованы множеством авторов [29-33].

В большинстве случаев эти программы могут применяться как к данным МУРР, так и к данным МУРН. Возможно, наиболее полным и популярным набором программ для анализа данных МУР является набор программ ATSAS, разработанный группой талантливых исследователей и программистов из разных стран под руководством Д. И. Свергуна. Развитие комплекса программ ATSAS [6] началось в 1990-х годах, при этом изначально это были отдельные программы, которые в 2004 объединили в один комплекс, который назвали ATSAS (All That Small-Angle Scattering), тем самым создав мощнейший инструмент МУР уже спустя почти столетие после начала развития этого метода. Тем не менее другие комплексы программ также доступны для пользователей.

В связи с растущим спросом на время проведения экспериментов на экспериментальных синхротронных станциях были предприняты усилия для повышения эффективности и ускорения процесса сбора и обработки данных. В частности, время экспозиции становилось значительно меньше, чем время, необходимое для ручной замены образца, вследствие чего была разработана измерительная ячейки с автоматическим заполнением образцом. Сегодня большинство биологических экспериментальных станций МУРР на синхротронных центрах имеют автоматическую роботизированную загрузку образцов [34-37]. Эти роботизированные системы не только сокращают общее время, затрачиваемое на замену образца, но также исключают человеческие ошибки в процессе загрузки и, таким образом, делают сбор данных более надежным, а данные - более согласованными. Благодаря им стало возможным осуществлять непрерывный сбор данных в течение продолжительного времени. В большинстве случаев эти роботизированные системы связаны с автоматизированным программным обеспечением по обработке данных, проводящим анализ данных в режиме реального времени во время измерения.

Несмотря на то, что подготовка образцов для МУР гораздо проще по сравнению с кристаллографией или электронной микроскопией, чистота и монодисперсность биологических образцов чрезвычайно важны. Белки в растворе склонны к агрегированию, и даже небольшое количество агрегатов может сделать данные непригодными для дальнейшего анализа. К сожалению, для некоторых систем даже короткий период между очисткой в биохимической лаборатории и измерением на экспериментальной станции может быть достаточным, чтобы начался процесс агрегации. Чтобы обойти это препятствие и обеспечить максимально возможную чистоту образца для эксперимента МУР, в настоящее время широко применяется схема сочетания эксклюзионной хроматографии с измерением МУР.

Таким образом, сегодня МУР — это один из широко используемых методов в экспериментальном арсенале структурных биологов. Универсальность, быстрота, умеренные требования к количеству образцов и относительная простота использования делают его привлекательным методом для рутинной структурной характеристики биомолекул. Сочетание метода МУР с другими структурными методами (такими как, кристаллография, ЯМР и электронная микроскопия) сделало возможным исследования с использованием гибридного подхода больших бимолекулярных ансамблей и комплексов [38, 39].

Для анализа сложных многокомпонентных систем методом МУР все более важную роль будут играть передовые вычислительные методы. В частности, моделирование методом молекулярной динамики позволит лучше уточнять структуры с высоким разрешением и использовать более широкий угловой диапазон данных МУР. Такое моделирование позволит изучать кинетику и динамику макромолекул и поможет интерпретировать наборы данных МУР с временным разрешением.

В целом, технический прогресс в ближайшие годы откроет новые захватывающие возможности для МУР в структурной биологии и междисциплинарных исследованиях.

1.2. Физические основы малоуглового рентгеновского рассеяния

Рассеяние плоской монохроматической волной на реальных объектах (схематически представлено на рис.1.1).

В некоторой точке наблюдения результирующая амплитуда электромагнитной волны будет иметь вид:

$$A = A_0 e^{-i\varphi} = A_0 e^{-i(k_0 r + \alpha_0)}, \qquad (1.1)$$

где *A*₀ – амплитуда источника;

r – вектор текущей координаты;

*k*₀ – волновой вектор, перпендикулярен плоскости волнового фронта, который задает направление распространения волны;

 φ – фаза, или фазовый сдвиг, равный в текущей точке $k_0r + \alpha_0$;

 $e^{i\varphi}$ – осциллирующая функция (по формуле Эйлера: $e^{i\varphi} = \cos \varphi + i \sin \varphi$);

*а*₀ – начальный фазовый сдвиг фронта, примем его равным 0.

Далее будет рассмотрено первое Борновское приближение (т.е. не будет учитываться повторное рассеяние уже рассеянных волн).

В результате интерференции лучей, регистрируется дифракционная картина, которая формируется путем сложения множества когерентно рассеянных волн (принцип суперпозиции), отличающихся друг от друга по фазе. Фазовые отличия и амплитуды слагаемых зависят от пространственного распределения электронной плотности $\rho(\mathbf{r})$, то есть от структуры объекта, и определяют форму экспериментальной кривой рассеяния I(s).



Рис. 1.1. Схема упругого рассеяния

Вектором рассеяния называется $s = k_1 - k_0$ и является разницей между волновыми векторами рассеянной k_1 и падающей и k_0 волн.

Модуль вектора рассеяния:

$$s = |s| = \frac{4\pi \cdot \sin \theta}{\lambda},\tag{1.2}$$

где 2*θ* – угол рассеяния (между падающим и рассеянным лучом); *λ* – длина волны.

Согласно принципу суперпозиции, амплитуда рассеяния от всего объекта принимает вид:

$$A_s(\mathbf{s}) = \sum_{i=1}^N f_i(\mathbf{s}) \cdot e^{i\mathbf{s}\mathbf{r}_i}, \qquad (1.3)$$

где $f_i(s)$ – амплитуда рассеяния *i*-го центра (атома, молекулы, ...); N – число центров;

Например, атомная амплитуда рассеяния выражается функцией (в силу близости симметрии электронной плотности атома к сферической, *s* и *r* здесь скалярные величины):

$$f(s) = 4\pi \int_0^\infty r^2 \rho_a(\mathbf{r}) \frac{\sin sr}{sr} d\mathbf{r}, \qquad (1.4)$$

где $\rho_a(\mathbf{r})$ – радиальная функция электронной плотности атома.

Будем работать с непрерывной функцией распределения электронной плотности $\rho(\mathbf{r})$, так как число электронов (атомов) в образце велико. После интегрирования по всему облученному объему получаем (с точностью до константы):

$$A(\mathbf{s}) = \int \rho(\mathbf{r}) e^{i\mathbf{s}\mathbf{r}} d\mathbf{r}$$
(1.5)

После обратного Фурье-преобразования:

$$\rho(\mathbf{r}) = \int A(\mathbf{s})e^{-i\mathbf{s}\mathbf{r}}d\mathbf{s}$$
(1.6)

Таким образом можно по известной функции амплитуды рассеяния A(s) находить структуру $\rho(\mathbf{r})$. Но в эксперименте измеряют не A(s), а интенсивность рассеяния $I(\mathbf{s})$

$$I(\mathbf{s}) = A(\mathbf{s})A^*(\mathbf{s}) \tag{1.7}$$

Функция I(s) задана в пространстве вектора *s*, т.е. в обратном пространстве. Так как детекторы регистрируют интенсивность излучения, а не его амплитуду, решение обратной задачи поиска структуры проводят с помощью последовательности решения прямых задач, варьируя параметры структуры ($\rho(\mathbf{r})$ или упрощенные варианты структур, заданных простыми геометрическими формами). При этом расчеты интенсивности рассеяния основаны на формулах (1.5) и (1.7), или на известных аналитических зависимостях, примеры которых можно найти, например, в [3].

Метод МУР получил такое название в связи с тем, что из-за быстрого спада интенсивности с ростом угла рассеяния и связанного с этим ростом относительной ошибки на больших углах, для расшифровки структуры объектов приходится регистрировать интерференционную картину только в области малых углов рассеяния (типично не более 3-5°). Причем с увеличением размера рассеивающего объекта уменьшается угловой диапазон, в котором сосредоточена основная часть рассеяния *s*, определяется соотношением $d = 2\pi/s$. То есть рассеяние на малые углы (0.1-2°) несет в себе информацию об объектах с размерами, многократно превышающими длину волны излучения, и дает структурную информацию с разрешение 1-2 нм.

Рассеяние в МУРР происходит на неоднородностях электронной плотности, то есть рассеивающая способность и, соответственно, структура вещества могут быть описаны распределением электронной плотности $\rho(\mathbf{r})$. Эффективная рассеивающая способность вещества определяется разницей (контрастом) между средней электронной плотностью пространства ρ_s , в котором находится рассеивающая частица, и электронной плотностью самой частицы $\Delta \rho(\mathbf{r}) = \rho(\mathbf{r})$ - ρ_s .

При анализе данных МУР следует разделять случаи монодисперсных и полидисперсных систем. Если все частицы идентичны, то система называется монодисперсной. Для монодисперсных систем интенсивность малоуглового рассеяния с точностью до постоянного множителя равна усредненной по всем ориентациям интенсивности рассеяния одной частицей. Если же частицы различаются между собой как по форме, так и по размерам, то систему следует отнести к полидисперсной. Анализ таких систем представляет собой нетривиальную задачу, так как в регистрируемую картину рассеяния вносят вклад все типы частиц, при этом их вклад пропорционален объемной доле каждого типа частиц. На практике одновременное определение формы и размеров частиц оказывается невозможным в силу неоднозначности решений. Поэтому в большинстве случаев форму частиц аппроксимируют простыми геометрическими телами (сферой, цилиндром, эллипсом), и задача анализа данных МУР сводится к нахождению функции распределения по размерам для частиц заданной формы. Ниже будут представлены имеющиеся в настоящее время подходы к решению данной задачи.

1.3. Основные алгоритмы для анализа полидисперсных систем

Коллоидная дисперсия может содержать до 10²¹ частиц на килограмм материала, и все эти частицы могут быть разными по размеру и форме. В некоторых случаях желательно иметь набор частиц, которые почти идентичны, однако в ряде других случаев требуется иметь широкое распределение частиц по размерам. Например, в составе бетона или в керамическом производстве, стремятся получать дисперсию частиц, которые максимально заполняют пространство, когда частицы меньшего размера заполняют пустоты между большими частицами во всем масштабном диапазоне материала.

Довольно часто имеет место ситуация, когда широкое или мультимодальное распределение частиц является результатом процесса, с помощью которого была создана коллоидная дисперсия, например это может быть фрагментация или эрозия более крупных частиц. Механические, гидродинамические и оптические свойства такой дисперсии становятся зависимыми от функции распределения частиц по размерам, поэтому надежный метод определения этого распределения представляет собой актуальную и важную задачу.

В настоящее время большинство методов определения распределений частиц по размерам обеспечивают надежные результаты для относительно узких распределений [40]. Однако, методы могут становится нестабильными при применении к распределениям, которые являются очень широкими и сложными, даже в случае сферических частиц, о которых будет идти речь в данном разделе.

Объем информации, который требуется для описания сложного, широкого или мультимодального распределения частиц по размерам, может оказаться за пределами возможностей большинства методов доступных в настоящее время.

Стабильность методов может быть оценена путем выбора очень широкого распределения по размерам, вычисления рассеяния от него, и затем попыткой восстановления исходного распределения. Это сравнение обычно показывает хорошее соответствие общих характеристик размерного распределения, но обнаруживает сильные отклонения для распределений в области малых размерах частиц, которые вносят основной вклад в интенсивность рассеяний при высоких значениях вектора рассеяния *s* [41].

Это реальная проблема при применении малоуглового рассеяния к реальным коллоидным системам, потому что большинство коллоидных

дисперсий имеют распределение размеров, которое намного шире и намного более сложное, чем обычно предполагается.

Действительно, есть только несколько типов синтезов, которые дают узкие распределения частиц (например, эмульсионная полимеризация частиц латекса, синтез минеральных частиц в растворах посредством гидротермических процессов, осаждение наночастиц диоксида кремния), тогда как многие другие процессы могут давать очень широкие распределения частиц (например, процессы дробления, измельчения и эрозии частиц). Кроме того, маленькие частицы могут обладать важными функциями, такими как контроль осмотического давления дисперсии [42] и ее поведение в процессе спекания.

Метод малоуглового рассеяния оказывается достаточно эффективным при анализе такого рода систем. Поэтому для анализа полидисперсных объектов в растворе по малоугловым данным разработан ряд подходов [4, 5, 43-46], среди которых можно выделить несколько основных для получения распределения частиц по размерам:

- (1) методы расчета контура распределения в рамках модели линейных наименьших квадратов, в том числе с регуляризацией решения по Тихонову (программа GNOM [4] из пакета ATSAS [6] и программа GIFT [47]). Эти программы относятся к методам прямого поиска гистограммы распределения со сглаживанием искомого профиля с помощью регуляризации решения по производным;
- (2) прямой поиск распределения частиц по размерам в виде гистограммы методами нелинейных наименьших квадратов (программа VOLDIS [A6]) и случайного поиска McSAS [46];
- (3) постулирование вида распределения в аналитическом виде (например, нормального распределения или распределения Шульца [48]) и проведение приближения данных методами нелинейных наименьших квадратов (программа MIXTURE (и ее модифицированный вариант POLYMIX) [5] из пакета ATSAS [6] и программа SASFIT [45];
- (4) Использование преобразования Титчмарша [49].

В работе использованы программы MIXTURE и GNOM, которые входят в комплексный пакет программ ATSAS [6] – один из широко используемых в мире (в настоящее время пользователями которого являются более 10000 человек из более, чем 2000 лабораторий по всему миру), разработан с целью обработки и интерпретации данных МУР позволяет восстанавливать структуру как растворов биомакромолекул, так и небиологических образцов, анализировать белковые смеси и строение наночастиц. Для академических пользователей находится в открытом доступе по URL-адресу:

https://www.embl-hamburg.de/biosaxs/software.html.

Программный пакет использует сложные статистические методы и методы оптимизации (такие как метод моделирования отжига, генетический алгоритм, конформационный анализ с помощью нормальных колебаний). В настоящее время пакет программ ATSAS имеет широкий функционал, состоит из подпрограмм, каждая из которых выполняет свою функцию. Он позволяет проводить первичную обработку данных и оценивать структурные параметры [4, 5, 50], определять трехмерные *ab-initio* формы исследуемых объектов с разрешением 1-2 нм [29, 33, 51], моделировать многодоменные белки и комплексы методом молекулярной тектоники [52-54], количественно характеризовать белковые смеси и гибкие системы [5, 55-58]. Также разработаны программы, ориентированные на расчет интенсивности от моделей высокого разрешения (PDB структур) и их суперпозицию [51, 59-62], анализ липидных систем [63, 64], разделение белковых компонент для малоуглового эксперимента с использованием хроматографической колонки [50, 65]. Кроме того есть возможность совместного использования МУР с такими методами как ядерный магнитный резонанс (ЯМР) [66-68], криоэлектронная микроскопия [69-71], и рентгеновская дифракция [72-74].

Сразу стоит отметить, что ниже пойдет также речь о программах VOLDIS и POLYMIX, которые еще не опубликованы. Но тем не менее, уже вышли статьи с упоминанием об этих программах [A6, A7], а в диссертации они используются с разрешения автора В.В. Волкова. Программа POLYMIX

отличается от MIXTURE применением более эффективного поиска минимума целевой функции.

1.3.1. Поиск функции распределения частиц по размерам с использованием регуляризации по Тихонову

Для вычисления функции распределения по размерам частиц в полидисперсных системах по экспериментальным данным интенсивности рассеяния *I*(*s*), в программах GNOM [4] и GIFT [47] проводится численное решение интегрального уравнения Фредгольма первого рода:

$$I(s) = \int_{Rmin}^{Rmax} I_0(s, r) D_V(r) dr$$
(1.8)

где $I_0(s, r)$ – заданный форм-фактор частицы (в данной работе использовались сферы), $D_V(r)$ – искомая функция распределения частиц по размерам, определенная в диапазоне [R_{min} , R_{max}].

Данное интегральное уравнение, представленное в конечных разностях в виде системы линейных уравнений с матрицей коэффициентов, составленной из интенсивностей рассеяния сферическими частицами, рассчитанными на экспериментальной сетке угловых отсчетов интенсивности и искусственной сетке радиусов от R_{\min} (обычно от 1Å) до предполагаемого максимального значения R_{\max} , представляет собой плохо обусловленную задачу. Для получения устойчивых решений, имеющих физический смысл, в программе GNOM применена регуляризация решения по Тихонову [75], т.е. минимизируется функционал $T_{\alpha}[D]$ (где $D=D_{\nu}(r)$, I=I(s), $I_0=I_0(s,r)$):

$$T_{\alpha}(D) = ||I - I_0 D||^2 + \alpha \Omega(D)$$
(1.9)

в который входит как квадрат невязки между экспериментальной и рассчитанной кривой рассеяния $||I - I_0 D||^2$, так и квадрат нормы первой производной функции распределения по размерам, умноженный на параметр регуляризации α , обеспечивающий гладкость контура [4]

$$\Omega(D) = \int_{Rmin}^{Rmax} \left[\frac{dDv(r)}{dr}\right]^2 dr$$
(1.10)

Параметр регуляризации α в данном подходе фактически представляет собой степень сглаживания решения. При этом α позволяет повысить устойчивость решения, но остается проблема подбора области определения [R_{\min} , R_{\max}] функции $D_V(r)$.

Следует отметить, что программа GNOM имеет более широкие возможности, и может применяться не только для восстановления распределения по размерам сферических частиц, но и для частиц с произвольным форм-фактором, а также для определения функции распределения по расстояниям p(r) в случае монодисперсных систем, как внутри трехмерной структуры частицы, так и в поперечном сечении вытянутых или по толщине для сплюснутых частиц. Во всех случаях в программе GNOM применяется один и тот же алгоритм для решения интегрального уравнения с использованием регуляризации решения по Тихонову, различие состоит в используемой функции ядра интегрального уравнения. Кроме того, программа GNOM позволяет учитывать факторы размытия сигнала ('smearing effects'), которые оказывают существенное влияние в рентгеновских экспериментах с неточечным источником (например, при использовании камеры Кратки) и в случае нейтронных малоугловых данных.

При автоматической оценке качества решения GNOM использует так называемые "perceptual" критерии, которые складываются из следующих составляющих: значения невязки между экспериментальными данными и малоугловой кривой, вычисленной по найденной функции распределения с помощью обратного преобразования, наличия систематических отклонений между экспериментальной и вычисленной кривыми, степени гладкости найденной функции распределения по размерам, стабильности получаемого решения при небольшом варьировании параметра регуляризации, а также положительности функции распределения по размерам. Найденные значения критериев сравниваются с ожидаемыми ("идеальными") значениями, и на основе такого сравнения вычисляется "нормированный" показатель качества решения ('Total

estimate'), который изменяется от 0 до 1 (чем ближе значение показателя к единице, тем выше качество решения).

Решение $D_{\nu}(r)$ чувствительно к эффектам обрыва кривой рассеяния и ограничению на максимальный размер частиц и может содержать осцилляции, существенно искажающие форму распределения. Поэтому, необходимо критично оценивать полученные программой GNOM результаты и разделять реальные особенности функции распределения от искусственных "артефактов", вызванных ограничениями используемого алгоритма программы.

1.3.2. Прямой поиск гистограммы распределения по размерам. Алгоритм Левенберга-Марквардта

В настоящее время также широко используется прямой поиск гистограммы распределения методом наименьших квадратов, который в частности реализован в программах McSAS [46] и VOLDIS [A6].

Следует отметить, что программный пакет McSAS основан на поиске гистограммы методом Монте-Карло, который требует много процессорного времени и, следовательно, количество точек в распределении ограничено несколькими десятками.

Основным недостатком прямого подхода являются высокочастотные колебания на кривой распределения. Причиной является сильная корреляция формы кривых рассеяния от частиц с размерами соседних точек в распределении.

Поэтому идеи VOLDIS заключаются в следующем:

1) Поиск минимума невязки осуществляется быстрым методом Левенберга-Марквардта [76], а именно вариант алгоритма NL2SOL с ограничениями на переменные - DN2FB (краткое описание метода представлено ниже).

2) Программа напрямую меняет значения распределения в каждом узле гистограммы. Однако это распределение будет состоять из большого числа узких пиков из-за сильной корреляции форм кривых интенсивности рассеяния соседних (т.е., близких) по размеру частиц. 3) Чтобы преодолеть эту трудность, интенсивность модельного рассеяния рассчитывается по сглаженному контуру распределения, после чего рассчитывается невязка из формы и стабильности интенсивности, экстраполированной к нулевому углу вне области подгонки.

Алгоритм Левенберга-Марквардта является одним из наиболее эффективных методов решения задач нелинейных наименьших квадратов. Он был сформулирован независимо Кеннетом Левенбергом [77] в 1944 году и Дональдом Марквардтом [78] в 1963 году.

В соответствии с алгоритмом, направление поиска p_k определяется как решение системы уравнений:

$$(J_k^T(\boldsymbol{x}_k)J_k(\boldsymbol{x}_k) + \lambda_k I)\boldsymbol{p}_k = -J_k^T(\boldsymbol{x}_k)\boldsymbol{f}_k(\boldsymbol{x}_k), \qquad (1.11)$$

где $J_k(x_k)$ – матрица Якоби вектор-функции $f_k(x_k)$, элементы которой представляют собой значения разностей экспериментальная интенсивность – теоретическая, x_k – вектор параметров модели, λ_k - некоторое неотрицательное число, играющее роль параметра регуляризации, I – единичная матрица, k – номер шага поиска.

Шаг вдоль p_k в начале полагается единичным, то есть очередная точка в пространстве параметров будет

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{p}_k$$

Можно показать, что p_k представляет собой решение задачи на условный минимум вида

найти
$$min\frac{1}{2}\|J_k(x_k)p + f_k(x_k)\|_2^2$$

 $p \in \mathbb{R}^n$

при ограничении $\|\boldsymbol{p}\|_2 \leq \Delta$,

где Δ – параметр, связанный с λ_k . Таким образом, метод Левенберга-Марквардта, реализованный в алгоритмах NL2SOL и DN2FB, относится к классу методов доверительной окрестности. Монотонное убывание минимизируемой функции достигается в нем за счет подбора «хороших» значений λ_k . При $\lambda_k=0$, p_k будет направлением Гаусса-Ньютона, когда $\lambda_k \rightarrow \infty$, норма $||\mathbf{p}_k|| \rightarrow 0$ и вектор \mathbf{p}_k в пределе становится параллельным антиградиенту. Следовательно, неравенство для функции невязки $F(x_k + p_k) < F_k$ всегда можно обеспечить, выбрав λ_k достаточно большим.

1.3.3. Параметрическое задание функции распределения по размерам

При использовании нелинейных методов минимизации для анализа полидисперсных систем функция распределения частиц по размерам задается в аналитическом виде (программа SASFIT [45]) или комбинации аналитических выражений (MIXTURE/POLYMIX [5]).

В последнем случае интенсивность рассеяния многокомпонентной разбавленной системы представляют в виде линейной комбинации вкладов рассеяния от *K* компонент, при этом каждая компонента есть система частиц заданной формы, распределение которых описывается одним аналитическим выражением с учетом, при необходимости, межчастичной интерференции:

$$I(s) = \sum V_k I_k(s), \qquad (1.12)$$

где V_k – объемная доля k-ого компонента,

 $I_k(s)$ – интенсивность рассеяния k-ого компонента,

s – вектор рассеяния, его модуль (1.2)

Для полидисперсной системы слабовзаимодействующих частиц, интенсивность каждого компонента можно представить в локально-монодисперсном приближении [79]:

$$I_k(s) = \int D_k(R) v_k(R) [\Delta \rho_k(R)]^2 i_{0k}(s, R) S_k(s, R) dR, \qquad (1.13)$$

где *R* – размер частицы,

 $D_k(R)$ – нормированное объемное распределение частиц по размеру для *k*-ого компонента,

 $v_k(R)$ – эффективный объем частицы k-ого компонента,

 $\Delta \rho_k(R)$ – рассеивающий контраст *k*-ого компонента,

 $i_{0k}(s, R)$ – нормированная к единичному объему интенсивность рассеяния от частицы в *k*-ом компоненте,

 $S_k(s, R)$ – структурный фактор (ответственный за межчастичную интерференцию). Если концентрация смеси достаточно мала, то можно пренебречь структурным фактором ($S_k(s, R) = 1.0$ для всех *s*).

Для расчета функции объемного распределения *k*-ого компонента использовалось распределение Шульца:

$$D(R) = \frac{1}{R_0} (z+1)^{z+1} \frac{1}{\Gamma(z+1)} \left(\frac{R}{R_0}\right)^z \exp\left[-\frac{(z+1)R}{R_0}\right],$$
(1.14)

где $z = \left(\frac{R_0}{\Delta R}\right)^2 - 1,$

 R_0 – среднее значение размера частицы,

 ΔR – дисперсия распределения,

 $\Gamma(z)$ – гамма-функция.

Алгоритм, использующий формулы (1.12)-(1.14), реализован в программе MIXTURE [5]. Данная программа позволяет моделировать многокомпонентные системы, содержащие различные типы полидисперсных частиц с помощью приближения простыми геометрическими телами, такими как: двухслойные сферы и цилиндры, эллипсоиды, "гантелеобразные" частицы. С помощью программы MIXTURE можно находить средние размеры частиц, степень полидисперсности и объемные доли компонентов, а также учитывать межчастичные взаимодействия для сильноконцентрированных образцов в приближении Перкуса-Йевика. Такой подход применим к исследованию систем различной природы, в частности, он успешно использовался для количественного описания температурного структурного перехода в трехкомпонентных микроэмульсиях [80], анализа ансамбля состояний самособирающихся икосаэдрических капсул фермента люмазин синтазы в нативном состоянии и при наличии мутаций [81], а также изучения олигомерного состава ассоциатов наночастиц оксида железа, внедренных в капсулы вируса мозаики Бром [82].

В программе MIXTURE для поиска решения используется метод переменной метрики в варианте Бройдена–Флетчера–Гольдфарба–Шанно [83] с простыми ограничениями на параметры (более подробно он будет описан в Главе 3). Программа POLYMIX имеет ту же целевую функцию, но работает в автоматическом режиме по алгоритму Левенберга-Марквардта [76], также с учетом простых ограничений на переменные и предназначена для рутинных вычислений.

Дополнительной особенностью программ MIXTURE и POLYMIX является возможность задания толщины слоя (для сферических частиц типа ядрооболочка) как в абсолютных (размерных) единицах, так и в относительных единицах (как отношение толщины к радиусу ядра частицы). Кроме того, в программе имеется возможность использования константы для корректировки сигнала в области больших углов, применения различных схем для проведения численного интегрирования при расчете моделей и задания весовых функций при расчете невязки между экспериментальными и модельными данными.

Данные программы позволяют восстанавливать функции распределения частиц по размерам, однако эффективность работы программ может зависеть от начального приближения и попадания в локальные минимумы в процессе поиска.

1.3.4. Преобразование Титчмарша-Фокса для вычисления распределения частиц по размерам

Одним из подходов к определению распределения частиц по размерам по данным МУРР является использование преобразования Титчмарша-Фокса [84, 85] для интегрируемой функции g(r):

$$\varphi(s) = \pi^{1/2} \frac{d}{ds} \left[\int_0^\infty g(r) qr J_\nu^2(sr) dr \right]$$
(1.15)

$$g(r) = -2\pi^{1/2} \int_0^\infty \varphi(s) s J_{\nu}(sr) Y_{\nu}(sr) ds, \qquad (1.16)$$

где функции J_v and Y_v - функции Бесселя первого и второго рода (порядка v), соответственно.

Если ввести функцию $\Psi(s)$, которая обращается в ноль при *s*=0:

$$\psi(s) = \int_0^s \varphi(s') ds', \tag{1.17}$$

то уравнения (1.16-1.17) можно переписать в виде пары новых соотношений:

$$\psi(s) = \frac{2}{\pi^{1/2}} \int_0^\infty g(r) R_\nu(sr) dr$$
(1.18)

$$g(r) = \frac{2}{\pi^{1/2}} \int_0^\infty [\psi(s) - \psi(\infty)] B_\nu(sr) ds, \qquad (1.19)$$

где функции ядра интегральных уравнений определяются следующими формулами:

$$R_{\nu}(u) = (\pi/2)uJ_{\nu}^{2}(u) \tag{1.20}$$

$$B_{\nu}(u) = \pi \frac{d}{du} [uJ_{\nu}(u)Y_{\nu}(u)] = \pi Y_{\nu}(u)[(1-2\nu)J_{\nu}(u) + 2uJ_{\nu-1}(u)] + 2, \quad (1.21)$$
причем $\psi(\infty) = \lim_{s \to \infty} \psi(s)$, поскольку функция $g(r)$ интегрируема.

Полученное соотношение имеет общий характер, но его можно применить к малоугловым данным в случае v=3/2. Действительно, если обозначить $\varphi(s) = s^4 I(s)$ и $g(r) \rightarrow cr^2 f(r)$, то уравнения (1.18) и (1.19) перейдут в следующий вид (так называемое преобразование Рэлея) [86]:

$$s^{4}I(s) = \frac{2}{\pi^{1/2}} \int_{0}^{\infty} cr^{2} f(r) R_{3/2}(sr) dr$$
(1.22)

$$r^{2}f(r) = \frac{2}{c\pi^{1/2}} \int_{0}^{\infty} \left[s^{4}I(s) - \Omega_{3/2} \right] B_{3/2}(sr) ds, \qquad (1.23)$$

где $R_{3/2}(u) \equiv \left(\frac{\sin u - u \cos u}{u}\right)^2$ - представляет собой квадрат форм-фактора сферы,

$$B_{3/2}(u) = 2[(1-2u^2)\sin 2u - u(2-u^2)\cos 2u]/u^3 \lor \Omega_{3/2} = \lim_{q \to \infty} [s^4 I(s)].$$

Таким образом, используя инверсное преобразование Рэлея, можно восстановить функцию распределения частиц по размерам, однако для этого необходимо экстраполировать экспериментальные данные в полном диапазоне *s*, и, кроме того, максимально точно определять константу $\Omega_{3/2}$, что бывает затруднительно на практике при наличии большой популяции маленьких частиц. Поэтому, хотя с математической точки инверсное преобразование Рэлея выглядит достаточно просто и привлекательно, использовать его на практике представляется затруднительным.

В работе [49] была предпринята попытка избавиться от присутствия трудно определимой константы $\Omega_{3/2}$ в формуле (1.23). Для этого были сделаны

следующие предположения: функция распределения искалась в виде средних значений $\langle f(r_i) \rangle$ на N интервалах шириной s с центрами в точках (r_1 , r_2 , ..., r_N = r_{max}). Эти функции можно описать с помощью двух-параметрической взвешивающей функции $w_s(r_i, r)$, обладающей следующими свойствами:

$$\langle f(r_i) \rangle \equiv \int_0^\infty f(r) w_s(r_i; r) dr$$

$$\int_0^\infty w_s(r_i; r) dr = 1$$

$$\int_0^\infty r w_s(r_i; r) dr = r_i$$

$$\lim_{s \to \infty} w_s(r_i; r) = \delta(r - r_i)$$

$$(1.24)$$

Если выбрать распределение Шульца для взвешивающих функций $w_s(r_i, r)$ согласно:

$$w_{s}(r_{i};r) = c_{w}r^{(r_{i}/s)^{2}-1}\exp(-rr_{i}/s^{2})$$

$$c_{w} = \frac{1}{\Gamma[(r_{i}/s)^{2}]} \left(\frac{r_{i}}{s^{2}}\right)^{(r_{i}/s)^{2}},$$
(1.25)

то уравнение (1.23) может быть преобразовано к виду:

$$\langle f(r_i) \rangle = \frac{1}{cr_i^2} \frac{2}{\pi^{1/2}} \int_0^\infty s^4 I(s) \psi_s(sr_i) ds,$$
 (1.26)

где функция $\psi_s(sr_i) = r_i^2 \int_0^\infty B_{3/2}(sr) w_s(r_i;r) / r^2 dr$ может быть вычислена аналитически, ее вид приведен в работе [49].

Тем не менее этот подход также обладает ограничением, так как сильно зависит от выбора шага разбиения интервалов для вычисления $\langle f(r_i) \rangle$. Основываясь на использовании теоремы Шеннона-Котельникова, авторы предлагают использовать для выбора значений $s = \alpha^* \Delta r$, где $\Delta r \simeq \frac{r_{max}}{N_{Sh}} = \frac{\pi}{2(s_{max} - s_{min})}$, а параметр α выбирается вручную, главное, чтобы он превышал значение 1/2. В какой-то степени этот подход близок к поиску прямой гистограммы распределений, но в отличие от подхода, реализованного в программах McSAS и VOLDIS, он не позволяет автоматически выбирать разбиение интервалов гистограммы, которое бы давало наилучшее восстановление функции распределения по размерам. Поэтому, хотя данный подход использует элегантный математический аппарат, широкого применения в настоящее время он еще не находит.

1.4. Методика измерений малоуглового рассеяния и экспериментальные установки

Принципиальная схема установки малоуглового рассеяния (рис. 1.2) состоит из источника излучения, из которого выходит сколимированный пучок (фотонов в случае МУРР и нейтронов в случае МУРН), системы формирования пучка, после прохождения которой пучок попадает на образец (раствор макромолекул). Далее излучение рассеивается на образце под углом 2θ и регистрируется детектором, при этом интенсивность упругого рассеяния записывается как функция угла рассеяния. Получается картина зависимости интенсивности излучения от модуля вектора рассеяния I(s). По этой картине можно определить форму и размер исследуемых объектов, их фазовый состав и внутреннюю структуру. Рассеяние от разбавленных растворов белков, нуклеиновых кислот и других макромолекул изотропно и зависит от модуля передачи импульса *s* (1.2).



Рис. 1.2. Принципиальная схема установки малоуглового рассеяния, где 1- источник излучения, 2- система формирования пучка, 3- исследуемый образец, 4- детектор, 2*θ* - угол рассеяния.
В более ранних экспериментах (и в настоящее время во многих лабораториях) в качестве источников рентгеновского излучения используют рентгеновские трубки. Но интенсивность излучения на таких установках ниже, чем на источниках синхротронного излучения, поэтому, чтобы получить хорошую картину рассеяния, следует подбирать более высокую концентрацию частиц в образцах для исследования, что чревато необходимостью учитывать межчастичную интерференцию.

Например, в институте кристаллографии им. А. В. Шубникова ФНИЦ «Кристаллография и фотоника» РАН в Москве располагается лабораторная установка «АМУР-К» [87], источником излучения которой является острофокусная рентгеновская трубка (рис. 1.3.). Установка оснащена однокоординатным позиционно-чувствительным детектором ОДЗМ при фиксированной длине волны излучения λ , равной 0.1542 нм (СиК_а – характеристическая линия медного анода, монохроматор из пиролитического графита) и коллимационной системой Кратки. Расстояние образец-детектор составляет 700 мм. Сечение рентгеновского пучка составляет 0.2 х 8 мм², область углов рассеяния соответствует диапазону значений модуля вектора рассеяния 0.1 < *s* < 10.0 нм⁻¹.



Рис. 1.3. Установка малоуглового рассеяния «АМУР-К» (ФНИЦ «Кристаллография и фотоника» РАН, Москва, Россия).

Конечно, лабораторные установки имеют преимущество в силу своей доступности, однако их недостатком является необходимость использования характеристического излучения, что не дает возможности оперативно изменять длину волны. Источники синхротронного излучения обладают несравнимо большей интенсивностью непрерывного рентгеновского спектра, что позволило совершить прорыв в МУРР. В качестве основных достоинств синхротронного излучения можно перечислить следующее: высокая интенсивность, линейная поляризация пучка в плоскости орбиты, периодическое прерывание его во времени и широкий диапазон энергий. Таким образом, можно сказать, что синхротронное излучения требует более сложной оптической системы, чем для лабораторных установок ввиду того, что необходимо использовать составные монохроматоры для выбора подходящего диапазона длин волн, зеркал полного внешнего отражения, чтобы устранять высшие гармоники и / или сфокусировать пучок.

Синхротрон предназначен для ускорения заряженных частиц (электронов, позитронов или протонов) до высоких энергий, вследствие отклонения от прямолинейного движения в магнитном поле заряженные частицы испускают фотоны, а вот уже поток фотонов используется в качестве источника излучения.

На рис 1.4 представлена линия Европейской Молекулярной Биологической Лаборатории P12 (EMBL-P12) [88] на синхротронном источнике PET-RAIII (DESY (Deutsches Elektronen-Synchrotron), Гамбург, Германия).

На этой станции используется сфокусированный монохроматический пучок рентгеновских лучей из ондуляторного источника. Энергию можно менять в диапазоне от 4 до 20 кэВ (длина волны $\lambda = 0.06$ -0.3 нм), впрочем, для большинства измерений используется энергия 10 кэВ (на длине волны $\lambda=0.124$ нм). Установка оснащена специальными зеркалами в геометрии Киркпатрик-Баеца. Эти зеркала в сочетании с тремя парами щелевой коллимации обеспечивают низкую расходимость пучка, обеспечивая при этом интенсивность 10^{13} фотон/сек. Диапазон значений волновых векторов от 0.08 до 4.5 нм⁻¹. Также имеется оборудование для автоматической смены образцов и двумерный детектор Pilatus 2M (фирма DECTRIS, производство Швейцария). Расстояние образец - детектор составляет 0.5 - 5 м. Под вакуумом находится весь путь прохождения пучка от источника до кюветы с образцом. Образец помещается в кварцевый капилляр и находится при установленной температуре, значения которой можно менять в пределах 7-45°C (до 50°C на короткий промежуток времени).



Рис. 1.4. Малоугловая станция P12 накопительного кольца Petra III, синхротрон DESY (Гамбург, Германия).

1.5. Возможности метода малоуглового рассеяния. Преимущества и ограничения.

В этом разделе выделены основные преимущества и ограничения метода.

<u>Достоинства</u>:

1. Главным достоинством метода является его общность: МУР может быть использовано для исследования упорядоченных, частично

упорядоченных и неупорядоченных объектов в разных состояниях (растворов, гелей, порошков, пористых материалов и т.д.).

2. Метод МУР применяют к структурам различных размеров, как к маленьким белкам и полипептидам (с молекулярной массой от нескольких кДа), так и к макромолекулярным комплексам (с молекулярной массой до десятков МДа), которые могут быть изучены с помощью современного оборудования МУР в условиях, максимально приближенных к естественным.

3. МУР не требует специальной подготовки образцов. Однако, есть некоторые параметры, которые стоит учитывать при планировании эксперимента: чистота и полидисперсность, стабильность и количество образца.

4. С помощью МУР можно исследовать объекты в широком диапазоне размеров частиц: от 1 до 200 нм.

5. Анализ методом МУР не занимает длительного времени (не более 1-2 дней). Сам эксперимент занимает короткое время: секунды, минуты – на синхротронах, а на дифрактометрах - часы.

6. По последним данным, минимальный объем исследуемого образца может достигать 3-5 микролитров (в случае использования микрофлюидных ячеек).

7. МУР позволяет оценивать гибкость и подвижность некоторых структур или частей структур (программа ЕОМ [56]) и гибких белков.

8. В случае МУРН не происходит разрушения образца. Что касается МУРР, то случается радиационное повреждение, но это устранимо с помощью добавления специальных реагентов (таких как DTT, TCEP, глицерин).

Отдельного рассмотрения требуют уникальные возможности МУРН с управлением поляризацией нейтронов в пучке и поляризацией ядер изотопов, обладающих магнитным моментом.

На современных установках, как лабораторных, так и синхротронных, имеется возможность автоматизировать анализ больших серий образцов с помощью роботизированных кассет. Нередко есть возможность исследовать монодисперсный образец, предварительно пропустив исходный полидисперсный раствор его через хроматографическую колонку.

Помимо неоспоримых достоинств метода, можно выделить ряд ограничений:

- Неоднозначность решения, связанная с математической неоднозначностью в случае поиска структуры частиц в монодисперсных системах, с многоэкстремальностью целевой функции при наличии ограничений на искомые параметры и/или с плохой обусловленностью задачи. Плохая обусловленность приводит к численной мультимодальности функции из-за ошибок округлений в численных расчетах.
- Потеря информации за счет сферического усреднения из-за изотропности картины рассеяния для неупорядоченных систем. Подробнее об этом речь пойдет ниже.
- 3. Ограниченность углового диапазона. Это зависит от размеров детектора и геометрических параметров установки.

1.6. Информативность данных и проблема неоднозначности решения в малоугловом рассеянии

Одной из главных проблем малоуглового рассеяния является проблема неоднозначности решений. В случае поиска формы частиц в монодисперсных системах и требует адекватного выбора анизометрии формы частиц при анализе полидисперсных систем. Она заключается в наличии существенно различающихся трехмерных структурных моделей, которые при этом имеют одинаковые функции распределений по расстояниям, и таким образом соответствуют одинаковым кривым малоуглового рассеяния. Основными причинами возникновения неоднозначности является, во-первых, невозможность прямого измерения амплитуды рассеяния (мы можем зарегистрировать только интенсивность, теряя информацию о фазе рассеянной волны), во-вторых, ограниченный угловой диапазон, доступный в эксперименте, и в-третьих, сферическое усреднение интенсивности рассеяния вследствие хаотической ориентации частиц в растворе, что приводит к потере значительной части структурной информации. Таким образом, строго говоря, однозначное *ab initio* восстановление структуры трехмерного объекта (то есть без каких-либо предварительных предположений о его строении) по одномерной кривой малоуглового рассеяния в общем случае не представляется возможным. Существуют способы уменьшения неоднозначности путем наложения ограничений на искомую структуру, за счет введения симметрии частиц или использования структурной информации, полученной другими методами.

Если имеется некоторая априорная информация о форме и структуре частиц, то анализ экспериментальных данных проводится следующим образом: выбирается теоретическая модель, зависящая от некоторого конечного числа параметров, и используется итерационный процесс оптимизации данного набора параметров, позволяющий минимизировать разницу между экспериментальными данными и расчетными данными от теоретической модели. Моделирование начинается с набора невзаимодействующих рассеивающих частиц, форма которых заранее определена (простые формы такие как сферы, цилиндры, стержни и др.). Полная картина рассеяния модели состоит из совокупности интенсивностей рассеяния от каждой частицы в образце. Однако нелинейная зависимость расчетных данных от параметров структурной модели затрудняет поиск глобального минимума функции. Поэтому при анализе для каждого параметра задается диапазон возможных значений, как можно более узкий, с учетом априорной информации, полученной с помощью других методов или исходя из общих соображений о физико-химических особенностях изучаемого объекта.

Однако, с точки зрения принципа суммирования (1.3 – 1.7) всегда можно по модели структуры однозначно рассчитать интенсивность рассеяния. Затем сравнить ее с экспериментальными данными и проверить достоверность модели, а также оценить величину ошибки в обратном пространстве. Кроме того, в процессе модификации структуры с помощью какой-либо процедуры, при этом минимизируя ошибку, можно найти модель объекта, которая хорошо согласуется с экспериментальными данными, но при этом стоит помнить, что такая модель может не являться единственной.

В общем случае необходимо различать неоднозначность задачи определения структурной модели и неустойчивость решения. Но неоднозначность, и неустойчивость приводят к тому, что находимые модели частиц могут существенно отличаться друг от друга, если проводить поиск с разных стартовых приближений или изменять параметры метода минимизации.

Оценить достоверность рассчитанной модели можно несколькими способами. Ниже описаны некоторые из них.

Во-первых, оценить показатель качества приближения $\chi^2(1.27)$, который получен в результате сопоставления интенсивностей модели с масштабированными измеренными данными (путем минимизации наименьших квадратов).

$$\chi^{2} = \frac{1}{N-1} \sum_{j=1}^{N} \left[\frac{I(s_{j}) - c I_{calc}(s_{j})}{\sigma(s_{j})} \right]^{2}, \qquad (1.27)$$

где *N* –число экспериментальных точек, *с* – масштабирующий коэффициент, совмещающий сравниваемые кривые рассеяния по критерию наименьших квадратов и определяемый согласно выражению

$$c = \left[\sum_{j=1}^{N} \frac{I(s_j)I_{calc}(s_j)}{\sigma(s_j)}\right] \left[\sum_{j=1}^{N} \frac{I(s_j)^2}{\sigma(s_j)}\right]^{-1},$$
(1.28)

где $I(s_j)$ и $I_{calc}(s_j)$ – экспериментальная и рассчитанная интенсивности соответственно, $\sigma(s_j)$ – экспериментальная ошибка, соответствующая модулю вектора рассеяния s_j . Если $\chi^2 \approx 1$, то можно говорить о том, что подгонка прошла успешно и найдено наиболее вероятное приближение.

Во-вторых, достоверность модели зависит от информативности самих малоугловых данных, которую можно оценить, в том числе, с помощью теоремы Котельникова-Шеннона (или теоремы отсчетов) [89], согласно которой, если аналоговый сигнал имеет ограниченный по частоте спектр, то он может быть восстановлен однозначно и без потерь по своим дискретным отсчётам, взятым с частотой, строго большей удвоенной верхней частоты. Если аналоговым сигналом является кривая интенсивности рассеяния от монодисперсной системы частиц с максимальным размером D_{max} , то максимально допустимым шагом ее представления будет $\Delta s = \pi / D_{max}$ (соответствующие точки в угловом диапазоне *s* принято называть шенноновскими каналами). Поэтому при проведении малоуглового эксперимента важно соблюдение условия $s_{min} < \pi / D_{max}$, т.е. чтобы первый шенноновский канал обязательно входил в экспериментальные данные.

На практике диапазон экспериментальных данных рентгеновского малоуглового рассеяния содержит от 4-5 до 20-25 шенноновских каналов $N_s = (s_{max} - s_{min}) \pi / D_{max}$. Ранее предпринимались попытки соотнести число шенноновских каналов N_s с количеством независимых параметров модели. Здесь s_{max} и s_{min} – максимальное и минимальное значение диапазона вектора рассеяния, где еще можно собрать данные с приемлемым отношением сигнал/шум. Однако, вследствие нелинейного характера зависимости параметров модели, точную связь установить затруднительно или даже невозможно. С другой стороны, при значительном (кратном) превышении числа параметров по сравнению с числом шенноновских каналов, устойчивость поиска, например, при восстановлении формы частицы, ухудшается [90].

Также следует учитывать наличие шумовой составляющей в данных, которая часто приводит к потере полезной информации, содержащейся в экспериментальных кривых в области больших углов. В работе [91] был предложен алгоритм на основе шенноновского приближения, реализованный в программе SHANUM и позволяющий оценивать оптимальный угловой диапазон данных. Поэтому, при моделировании теоретических кривых с добавлением шума (см. Глава 3) угловой диапазон данных выбирался согласно оценке программы SHANUM.

В-третьих, несмотря на то, что МУР достаточно самостоятельный и состоявшийся метод для исследования структур и внутреннего состава, для подтверждения достоверности данных малоугловое рассеяние часто сочетают с другими методами, такими как ЯМР, динамическое светорассеяние, рентгеновская кристаллография, аналитическое центрифугирование, эксклюзионная хроматография.

1.7. К понятию устойчивости решений

В данной работе рассматривается устойчивость поиска "правильных" решений относительно диапазонов допустимых стартовых значений параметров моделей.

Под "правильным" решением понимается распределение, попадающее в окрестность известного ответа (задаваемую с определенным малым разбросом значений параметров) в случае моделирования, либо центр области, в которой сосредоточены решения экспериментальной задачи с аналогично заданным разбросом. В обоих случаях решения получают методом мультистарта со стартовых значений, которые задаются на сетке в пределах допустимых разбросов распределений, определенных методами прямого поиска гистограммы.

В этой формулировке критерием устойчивости служит относительная величина области допустимых стартовых значений параметров модели относительно всей области, на которой задана их сетка.

Под "устойчивостью решения" также понимают устойчивость (или "робастность") алгоритма поиска, т.е. чувствительность решения к схеме и параметрам алгоритма. В работе рассматриваются наиболее устойчивые алгоритмы, отобранные по результатам публикаций, посвященным исследованию поведения процедур поиска на большом спектре тестовых задач.

Устойчивость решения непосредственно связана с числом обусловленности задачи, которое оценивают по отношению максимального и минимального сингулярных значений матрицы вторых производных целевой функции как в стартовой точке, так и в точке минимума. Число обусловленности устанавливает верхнюю границу возможного роста относительно ошибки в норме решения по отношению к норме ошибок в исходных данных. Для рассматриваемых в работе задач число обусловленности может достигать сотен тысяч. Однако, на практике, разброс оказывается намного меньше. Влияние экспериментального шума в случае данных рассеяния сильно ослаблено из-за их переопределенности: так как число точек на кривых рассеяния значительно превышает минимально допустимое число Шенноновских каналов.

Большое число обусловленности делает задачу поиска распределений некорректно поставленной. В математике задача считается корректной (или корректно поставленной), если выполняются условия: 1) задача имеет решение при любых допустимых исходных данных (существование решения); 2) каждым исходным данным соответствует только одно решение (однозначность задачи); 3) решение устойчиво.

При этом задачи, не удовлетворяющие хотя бы одному условию корректности, называются некорректными (или некорректно поставленными). К некорректным относят в том числе и задачи, имеющие принципиальное математическое решение, но неразрешимые вследствие слишком большого влияния случайных погрешностей измерений или неточности расчетов.

В нашем случае и ошибки во входных данных, и ограниченная точность машинной арифметики не обеспечивают строгое выполнение условий 2) и 3). Цель работы состояла как раз в том, чтобы невыполнение условия 2) не приводило к недопустимым нарушениям условия 3). Поэтому под "точным" в работе понимается решение, отклоняющееся от "правильного" в заранее заданных пределах. В свою очередь, критерии "правильности" структурной модели могут быть выработаны из совокупности априорных физико-химических сведений об объекте, с одной стороны, и величины разброса решений, полученных методом мультистарта при варьировании начальных параметров модели и параметров алгоритмов поиска решений.

1.8. Методы оптимизации, рассмотренные в исследовании

В работе проводилось исследование на устойчивость алгоритмов четырех модификаций программы MIXTURE. В каждой из которых были использованы различные минимизационные схемы: метод **BFGS** [83], метод моделирования отжига SA (Simulated Annealing) [92], комбинация этих методов BFGS+SA (см. раздел 4.2.1), а также метод многогранника Нелдера-Мида NM (Nelder-Mead) [93].

1.8.1. Градиентный метод Бройдена-Флетчера-Гольдфарба-Шанно

Алгоритм **BFGS** (Бройдена–Флетчера–Гольдфарба–Шанно) является итерационным методом, относится к классу квазиньютоновских методов, в которых матрица вторых производных (гессиан H(x)) функции вычисляется не напрямую, а в некотором приближении, основываясь на сделанных до этого шагах [83].

Пусть требуется решить задачу оптимизации *min* f(x,y), причем в общем случае функция f(x,y) является невыпуклой и имеет непрерывные вторые производные.

С использованием методов второго порядка данная задача решается итерационно, с помощью разложения функции в полином второй степени:

$$f(x_k + p) = f(x_k) + \nabla f^T(x_k)p + \frac{1}{2}p^T H(x_k)p, \qquad (1.29)$$

где H – матрица Гессе или гессиан f(x).

Ввиду трудоемкости вычисления гессиана H(x), алгоритм **BFGS** вычисляет лишь приближенное значение гессиана B_k (см. ниже).

Затем ищется минимум полученной квадратичной задачи:

$$p_k = -B_k^{-1} \nabla f(x_k) \tag{1.30}$$

После ищется решение вдоль данного направления точки, при котором выполняются условия Вольфе:

$$f(x_k + \alpha_k * p_k) \le f(x_k) + c_1 * \alpha_k * \nabla f_k^T * p_k$$

$$\nabla f(x_k + \alpha_k * p_k) * p_k \ge c_2 * \nabla f_k^T * p_k$$
(1.31)

Константы c_1 и c_2 выбираются следующим образом: $0 \le c_1 \le c_2 \le 1$. Причем c_1 выбирается в окрестности 0 (т.е. функция после совершения шага должна

уменьшиться), а *c*² в окрестности 1 (проекция градиента либо уменьшается, либо меняет направление).

В качестве начального приближения гессиана используют единичную матрицу *I* (либо любую хорошо обусловленную невырожденную матрицу).

Затем вычисляется приближенное значение гессиана на следующем шаге:

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k},$$
(1.32)

где $s_k = x_{k+1} - x_k$ – шаг алгоритма на *k*-й итерации, $y_k = \nabla f_{k+1} - \nabla f_k$ – изменение градиента на последовательных шагах.

Но так как вычисление обратной матрицы сложно, то обновляется обратная к B_k матрица $C_k = B_k^{-1}$ (вместо B_k^{-1}):

$$C_{k+1} = (I - \rho_k s_k y_k^T) C_k (I - \rho_k y_k s_k^T) + \rho_k s_k s_k^T,$$
(1.33)

где $\rho_k = \frac{1}{y_k^T s_k}$.

Метод достаточно эффективен, не требует больших временных затрат для поиска решения, но при этом, решение может остановиться в локальном минимуме целевой функции, находящемся далеко от глобального экстремума целевой функции.

1.8.2. Метод многогранника Нелдера-Мида

Метод **NM** [93] - Нелдера-Мида, также известный как метод деформируемого многогранника (часто используемое название «симплекс-метод» неверно – этот термин относится к другому, конфигурационному методу поиска) — простой и в тоже время эффективный метод, позволяющий оптимизировать функции многих переменных без вычисления градиентов. Суть метода заключается в последовательном перемещении вершин многогранника в пространстве параметров модели в сторону точки экстремума. Однако, если метод находит локальный экстремум типа «плоского узкого оврага», то многогранник может сжаться настолько, что после поиск экстремума прекратится из-за ограниченной точности вычислений. В этом случае достаточно эффективным оказывается повторный запуск с новой конфигурации вершин. Метод отличается высокой надежностью и позволяет, после нескольких запусков, найти решение, более точное, чем ранее рассмотренные алгоритмы, хотя и требует длительных расчетов. К минусам данного метода можно отнести отсутствие ограничений на параметры, так как в решении может получиться отрицательное или нулевое значение.

Симплекс представляет из себя геометрическую фигуру, являющуюся *п*-мерным обобщением треугольника. Для одномерного пространства — это отрезок, для двумерного - треугольник. Таким образом *n*-мерный симплекс имеет n+1 вершину. На первом шаге выбираются n+1 случайные точки и формируется симплекс. Вычисляются значения функции в каждой точке. Далее проводится сортировка, в результате которой из вершин симплекса выбираются три точки: x_h с наибольшим (из выбранных) значением функции f_h , x_g со следующим по величине значением f_g и x_l с наименьшим значением функции f_l . Затем осуществляется операция отражения одной вершины x_h относительно \boldsymbol{x}_c с коэффициентом α , получаем новую точку \boldsymbol{x}_r согласно формуле $\boldsymbol{x}_r = (1+\alpha)$ \boldsymbol{x}_c - $\alpha \boldsymbol{x}_h$ и вычисляем в ней функцию: $f_r = f(\boldsymbol{x}_r)$. Далее проводится сравнение значений f_r со значениями функции f_h , f_g $u f_l$. В зависимости от соотношения между этими величинами производится сжатие или растяжение многогранника (симплекса) в итерационном цикле. На последнем шаге осуществляется проверка сходимости. Суть проверки заключается в том, чтобы проверить взаимную близость полученных вершин симплекса, что предполагает и близость их к искомому минимуму. Если требуемая точность ещё не достигнута, можно продолжить итерации с первого шага. Метод NM дает сильное уменьшение значение функции уже при первых нескольких итерациях, не "застревает" в узких оврагах на поверхности целевой функции, что особенно важно в случае плохо обусловленных задач, но медленно достигает необходимой точности. Это крайне важно в тех ситуациях, когда вычисление значений функции требует много времени. В то же время, из-за отсутствия теорем о сходимости, метод **NM** не может служить эталоном для алгоритмов поиска.

1.8.3. Метод моделирования отжига

Метод моделирования отжига **SA** [92] является стохастическим методом поиска глобального экстремума. Он является наиболее подходящим методом для минимизации функций с большим числом переменных, что идеально подходит для многих задач МУР. Алгоритм основан на методе случайных испытаний, запускается с произвольной модели, параметры которой затем изменяются случайным образом в соответствии со схемой Метрополиса [94] до тех пор, пока решение не станет соответствовать экспериментальным данным, например, по критерию хи-квадрат (χ^2).

Условимся различать два состояния системы, определяемой функцией распределения частиц по размерам $D_{v}(r)$: текущее и пробное значения функции $D_{\nu}(r)$. В итерационном процессе метода **SA** текущее состояние, становясь пробным, служит в дальнейшем в качестве нового текущего состояния системы. Вероятность принятия пробной функции в качестве новой текущей определяется значением больцмановской функции $exp(-\delta\chi^2/T)$, где $\delta\chi^2$ – изменение целевой минимизируемой функции, Т- параметр, называемый "температурой отжига". Выбор нового текущего состояния системы существенно зависит от значения Т. Видно, что более высокая температура означает более высокую вероятность $exp(-\delta \chi^2/T)$), что позволяет принять пробную модель в качестве новой текущей даже в случае худшего пробного приближения в сравнении с текущим, когда $\delta \chi^2 > 0$. Если же $\delta \chi^2 < 0$, то пробная модель всегда принимается в качестве новой текущей. В начале процесса минимизации целевой χ^2 -функции "температура отжига" *T* выбирается достаточно высокой, что фактически соответствует случайному характеру изменений системы. Это заставляет метод SA «блуждать» по пространству значений χ^2 -функции и, за счет этого, удается преодолевать локальные минимумы, которые являются одним

из главных препятствий для большинства нелинейных методов минимизации, например, для квазиньютоновского градиентного метода спуска.

В итерационном процессе минимизации целевой χ^2 -функции температура T_{k+1} на следующем после предыдущего k-шага снижается монотонно как $T_{k+1} = F T_k$, где коэффициент "отжига" F равен 0.93-0.95, а значение целевой χ^2 функции уменьшается аналогично тому, как снижается внутренняя энергия системы в процессе понижения температуры. Однако поиск решения с помощью такого алгоритма занимает длительное время, так как требуется проводить достаточно большое количество температурных итераций (не менее 150-200), причем на каждой итерации осуществляется минимум 1000-10000 вариаций модели.

Таким образом, отличие метода моделирования отжига от метода Монте-Карло состоит в том, что наилучшая модель используется только для сравнения с пробной моделью, получаемой путем случайной вариации текущей модели, но текущая модель не обязана быть наилучшей. Наилучшая модель перезаписывается пробной только в случае, если пробная лучше наилучшей, становясь одновременно и пробной.

1.9. Примеры использования функции распределения по размерам для анализа олигомерного состава растворов биомакромолекул

Оценка распределения частиц по размерам является важным элементом анализа и для решения практических задач структурной биологии. В ряде случаев на основе такого анализа, можно сделать предположение об олигомерном составе белкового раствора, которое позволяет в дальнейшем проводить более детальное моделирование с использованием структурных моделей высокого разрешения.

В качестве одного из примеров использования такого подхода можно привести исследования предкристаллизационного состояния белков. Следует отметить, что кристаллизация белка является длительным процессом и зависит от разных условий (значение pH, температура, концентрация белка и тип осадителя). И заранее неизвестно, произойдет ли кристаллизация или нет. В работах [95-99] был предложен новый подход, который может облегчить поиск оптимальных условий кристаллизации. Этот подход основан на анализе структуры кристаллов белка и идентификации возможных олигомеров с использованием данных МУР, последние могут служить в качестве строительных блоков для кристалла белка.

Для этого были проведены исследования лизоцима методом МУРР [97] и МУРН [95]. Были проанализированы кривые МУР для образцов, находящихся в разных условиях: с разной концентрацией белка, с осадителем и без него, при разных температурах, а так же при замене растворителя с H_2O на D_2O [96]. Предварительная обработка данных МУР выполнялась программой PRI-MUS (усреднение сигнала от буфера раствора, вычитание усредненного сигнала от буфера из экспериментальных данных МУР и нормировка на концентрацию белков). Далее с применением программы POLYMIX были найдены объемные распределения частиц по размерам (рис. 1.5). Все распределения, полученные от чистого белка, содержат пик в районе 15-20 Å, соответствующий мономеру лизоцима. Но на рис. 1.5в, помимо пика от мономера, присутствуют вклады от частиц малых размеров с радиусом около 8-10 Å. Эти пики можно отнести к флуктуациям плотности в буферном растворе. Однако более примечательно, что при добавлении к раствору белка осадителя появляются большие частицы с радиусом в диапазоне 30-35 Å, соответствующие олигомерным частицам белка лизоцима. Было установлено, что данным размерам хорошо соответствуют октамеры. Далее с помощью программы OLIGOMER [5] были определены объемные доли мономеров, димеров и октамеров в растворе лизоцима (построенные на базе кристаллографических моделей высокого разрешения) и их зависимости от температуры и концентрации белка.

Лизоцим стал первым объектом, на котором был опробован данный подход. После чего с применением такого подхода стали исследовать другие белки. Так, на основе анализа данных МУРР в растворе протеиназы К [98] обнаружены мономеры и димеры, других олигомеров не обнаружено. Поэтому димеры можно считать строительными блоками будущего кристалла. При этом показано, что при добавлении осадителя появляется существенная доля димеров протеиназы К в растворе, в то время как влияние температуры слабо влияет на увеличение содержания димеров.



Рис. 1.5. Восстановленные программой POLYMIX распределения частиц по размерам $D_V(R)$ по данным МУРР от белка лизоцима для концентраций 20 мг/мл, 40 мг/мл и 60 мг/мл (без осадителя (а), (б), (в) – соответственно; с добавлением 25 мг/мл NaCl (г), (д), (е) – соответственно).

В растворе термолизина [99] были обнаружены мономеры, димеры и гексамеры, поэтому здесь субъединицей роста кристалла служит гексамер. Причем количество гексамеров растет с понижением температуры раствора, а также с увеличением концентрации осадителя.

Также следует отметить исследования по определению трехмерной формы белка фибриногена [100]. Белок фибриноген содержится в плазме крови всех позвоночных животных и является одним из ключевых белков, принимающих участие в процессе свертывания крови. При взаимодействии с ферментом тромбином он расщепляется и образует нерастворимый фибрин, являющийся основой кровяного сгустка. Фибрин является одним из перспективных биоматериалов в тканевой инженерии, создающим благоприятные условия для формирования капилляроподобной сети сосудов. В связи с этим изучение структуры фибриногена в растворе и его взаимодействия с тромбином представляет большой интерес в области медицины. Кроме того, возможность модификации структуры фибриногена с помощью синтетических полимеров (например, полиэтиленгликоля) открывает большие перспективы в управлении их свойствами с целью адресной доставки лекарств и использовании в тканевой инженерии.

Экспериментальные данные МУРР были получены на лабораторной установке «АМУР-К». С помощью программы DAMMIN [29] была найдена компактная взаимосвязанная модель, которая наилучшим образом аппроксимирует экспериментальные данные с использованием метода моделирования отжига. В результате форма белка фибриногена была восстановлена и определен олигомерный состав раствора. Было показано, что молекулы фибриногена образуют удлиненные структуры, их длина составляет около 50 нм, а их поперечное сечение составляет около 10 нм. Объем этих структур хорошо коррелирует с объемом двух димерных молекул фибриногена, поэтому в данном конкретном случае проводить оценку степени полидисперсности частиц цилиндрического типа (с помощью программы POLYMIX) не потребовалось. Дальнейшие исследования модификаций структуры фибриногена с полиэтиленгликолем и с фактором роста эндотелия сосудов [101, 102] показали что они состоят из нитевидных структур, зависящих от размера молекулы ПЭГ-а, молярного соотношения фибрин:ПЭГ и наличия/отсутствия фактора роста эндотелия сосудов.

Заключение к главе 1.

В настоящее время метод МУР продолжает активно развиваться, многие научные группы занимаются разработкой новых методов анализа и интерпретации данных, построения трехмерных моделей по данным МУР и использования информации комплементарными методами (такими как ЯМР, белковая кристаллография, электронная микроскопия и др.) при моделировании малоугловых данных. Развитие экспериментальной базы малоугловых станций синхротронных центров третьего поколения позволило выполнять измерения в автоматизированном режиме, что резко увеличило возможности метода и количество собираемых данных. В связи с этим развитие инструментов оценки достоверности получаемых решений и улучшения надежности восстановления структурных параметров частиц (в т.ч. функции распределения частиц по размерам) является актуальной темой исследований. Проведение таких оценок должно быть частью самой методики исследований структуры наночастиц по данным МУР.

Следует отметить, что анализ многокомпонентных полидисперсных смесей представляет собой нетривиальную задачу, учитывая нелинейную зависимость параметров системы и отсутствие априорной единственности и однозначности восстанавливаемых решений. Определение функции распределения частиц по размерам биомакромолекул и других наноразмерных структурных неоднородностей в неупорядоченных средах является чрезвычайно востребованной областью исследований в области нанотехнологий и структурной биологии.

В разделе 1.3. описаны подходы и методы, используемые в настоящее время при исследовании полидисперсных систем. При этом в диссертационной работе применяются методы линейных наименьших квадратов (в том числе с регуляризацией по Тихонову), постулирование аналитического распределения методами нелинейных наименьших квадратов, а также прямой поиск распределения частиц по размерам в виде гистограммы методами нелинейных наименьших квадратов.

55

Все вышеперечисленные подходы имеют свои преимущества и недостатки, однако, задача оценки единственности и устойчивости восстановления функции распределения по размерам для многокомпонентных полидисперсных систем остается недостаточно проработанной.

Поэтому в данной работе была поставлена задача развития новых подходов, основанных на сочетании и комбинированном использовании разных методов, которые позволили бы более надежно и точно проводить количественный структурный анализ многокомпонентных полидисперсных систем по данным малоуглового рассеяния. Этому посвящены следующие главы.

ГЛАВА 2. ИССЛЕДОВАНИЕ УСТОЙЧИВОСТИ РЕШЕНИЙ ПРИ АНА-ЛИЗЕ ТЕОРЕТИЧЕСКИХ ПОЛИДИСПЕРСНЫХ СИСТЕМ МЕТОДОМ МАЛОУГЛОВОГО РАССЕЯНИЯ. ЗАВИСИМОСТЬ РЕШЕНИЯ ОТ ПАРАМЕТРОВ МОДЕЛИ И ШУМОВ ДАННЫХ

В данной главе в целях исследования на устойчивость восстановления решений для полидисперсным систем использована программа MIXTURE с реализованным в ней алгоритмом Бройдена-Флетчера-Гольдфарба-Шанно (см. раздел 1.8.1). Исследования проведены для теоретических данных МУР от полидисперсных систем двух сферических частиц. Оценка устойчивости показана на примере рассчитанных кривых МУР от теоретических модельных систем сферических частиц. Кроме того, приведены результаты анализа добавления шумовой составляющей к данным МУР. Добавлялся шум двух типов – гауссовский и пуассоновский. Все результаты приведены ниже.

2.1. Системы с неперекрывающимися распределениями сферических частиц по размерам

В качестве объекта исследования были выбраны модели, состоящие из двух систем сфер разного среднего радиуса R_1 и R_2 , полуширинами распределений dR_1 и dR_2 и объемными долями V_1 и V_2 , соответственно. Будем называть такие парциальные системы, распределение в которых описывается одной параметрической функцией, компонентами. Для проведения анализа устойчивости алгоритма программы MIXTURE были рассчитаны кривые МУР от этих модельных систем. Результаты исследований опубликованы в статьях [A1, A2].

Было проведено моделирование с использованием теоретических наборов данных в шести случаях без шума и с добавлением шума, отвечающего распределениям Гаусса и Пуассона (табл. 2.1) со стандартными отклонениями, типичными для экспериментальных данных. Угловой диапазон данных до 2.0 нм⁻¹также был выбран как типичный для малоугловых измерений. Для нормально распределенного шума относительный уровень шума достигал 25%, уровень пуассоновского шума достигал 25% в области малых интенсивностей. Рассчитанные кривые МУР и соответствующие им функции объемного распределения по размерам представлены на рис. 2.1.

Таблица 2.1. Исходные параметры шести теоретических кривых с неперекрывающимися распределениями частиц по размерам, для которых проводится исследование на устойчивость решения

Модель	V_1 ,	V_2 ,	R_1 ,	R_2 ,	dR_1 ,	dR_2 ,
	отн. ед.	отн. ед.	HM	HM	HM	HM
1	0.75	0.25	5.0	15.0	0.1	0.1
2	0.25	0.75	5.0	15.0	0.1	0.1
3	0.75	0.25	5.0	15.0	2.0	0.1
4	0.25	0.75	5.0	15.0	2.0	0.1
5	0.75	0.25	5.0	15.0	0.1	4.0
6	0.25	0.75	5.0	15.0	0.1	4.0

Выбор шумовых распределений объясняется следующими факторами: случай пуассоновского распределения шума соответствует малоугловым измерениям с использованием детекторов – счетчиков квантов. Гауссово распределение иногда используют как предельный случай для моделирования шумов детекторов с большим собственным шумом. На рис. 2.1 видно, что в случае пуассоновского распределения дисперсия шумов данных становится заметной только в области больших углов ($s > 1.0 \text{ нм}^{-1}$), т.е. в области значений интенсивности, соответствующих 10^{1} - 10^{2} регистрируемых детектором квантов, тогда как в случае гауссовского распределения шумы на малоугловой кривой заметны во всем угловом диапазоне. Кроме того, при добавлении шумов было проверено несколько их реализаций. Результаты оказались аналогичны и в диссертации не приведены.



Рис. 2.1. Шесть модельных кривых МУР (слева) и соответствующие им объемные распределения по радиусам сферических частиц $D_V(R)$ (справа): кривые МУР с относительными уровнями шума 0% и 25% (I - гауссовское распределение, II - пуассоновское распределение). Для лучшей визуализации представлены со смещением по вертикали.

На практике случай *I* (распределение Гаусса) может соответствовать измерениям слабо рассеивающих сильно разбавленных образцов (с концентрацией частиц в растворе меньше 0.5–1.0 мг/мл), случай *II* (распределение Пуассона) – умеренно разбавленным образцам (с концентрацией частиц 3.0–5.0 мг/мл) и стандартной схеме измерений с использование детекторов – счетчиков квантов.

Также было проведено моделирование с использованием тех же теоретических данных МУР, но в более узком угловом диапазоне от 0 до 0.97 нм⁻¹ (результаты не приведены), соответствующем более низкому разрешению моделей. Качественно характер диапазонов устойчивости параметров совпал с приведенными ниже результатами для более широких угловых диапазонов.

Все 6 случаев условно разделены на 3 типа с преобладанием малых или больших частиц:

А (модель 1 и 2) – случай двух узких (и неперекрывающихся) распределений сферических частиц,

Б (модель 3 и 4) – случай широкого распределения малых сфер и узкого распределения больших сфер (также неперекрывающихся),

В (модель 5 и 6) – случай узкого распределения малых сфер и широкого распределения больших сфер (неперекрывающихся).

В каждом случае проведено исследование сходимости с помощью программы MIXTURE к правильному решению (соответствующему модели) <u>при</u> <u>поочередном варьировании стартовых значений параметров (V_1 , V_2 , R_1 , R_2 , dR_1 <u>и dR_2)</u>. Другим параметрам при этом присваивались правильные (соответствующие модели) стартовые значения. Результаты приведены в табл.2.2. Показаны диапазоны устойчивого приближения при варьировании значения одного из стартовых параметров.</u>

60

Таблица 2.2. Диапазоны устойчивого приближения при варьировании одного параметра. Первая строка для каждого параметра соответствует шуму 0%, вторая – гауссовскому шуму 25%, третья – пуассоновскому шуму 25%.

Модель	1	2	3	4	5	6
dR_1 , нм	0.1–4.5	0.01–4.0	0.1–5.0	0.1–5.0	0.01–4.7	0.01–5.0
	0.1–5.0	0.01–4.3	0.1–5.0	0.1–5.0	0.1–4.6	0.01–4.6
	0.1–5.0	0.01–5.0	0.1–6.0	0.1–6.0	0.01–5.0	0.01–5.0
dR_2 , нм	0.01–16.0	0.01–15.0	0.01–20.0	0.01–15.0	0.2–18.6	0.2–15.0
	0.01-15.0	0.2–15.0	0.01–17.2	0.01–18.8	6.0–20.0	0.2–15.0
	0.2–20.0	0.01–20.0	0.01-20.0	0.01–20.0	0.1–10.0	0.2–20.0
<i>R</i> ₁ , нм	3.9–6.1	3.0–6.0	1.0-10.0	2.1–10.0	3.0-6.2	2.9–6.1
	4.0–5.9	3.9–6.0	1.0-10.0	1.0-10.0	4.0-6.2	3.3–6.1
	2.4–6.8	2.4-8.0	1.0-10.0	1.0-10.0	2.1-8.0	1.5-8.0
<i>R</i> ₂ , нм	14.0–16.2	13.8–16.2	10.0-30.0	14.0–16.0	10.0-30.0	10.0-30.0
	14.0–16.0	14.0–16.0	10.4–20.0	14.0–16.2	10.0-30.0	10.0-30.0
	10.0-30.0	10.0-30.0	10.0-30.0	10.0-30.0	10.0-30.0	10.0-30.0

Можно выделить следующие результаты моделирования:

Результаты по радиусам обеих сфер:

А – область, позволяющая найти правильное решение ($R_1 = 5$ нм, $R_2 = 15$ нм), находится в пределах 3.9–6.1 нм для R_1 и 14.0–16.2 нм для R_2 при соотношении объемных долей частиц 75:25. В случае, когда соотношение объемных долей меняется на противоположное – 25:75 (преобладают сферы большего радиyca), область допустимых стартовых значений увеличивается для R_1 до 3.0–6.0 нм, а для R_2 – до 13.8–16.2 нм. Кроме того, добавление 25% гауссовского шума также уменьшает область «успешных» восстановлений решений, пределы для R_1 сжимаются до диапазона 4.0–5.9 нм в случае, когда малые сферы преобладают и до диапазона 3.9–6.0 в случае, когда преобладают большие частицы. Пределы для R_2 сужаются до диапазона 14.0–16.0 нм в обоих случаях. Если к теоретическим данным добавляется 25% пуассоновский шум, то граница для *R*₁ вырастает до 2.4–6.8 нм (с долей 75% малых частиц) и до 2.4–8.0 нм (с долей 25%), а для *R*₂ расширяются до диапазона 10.0–30.0 нм.

Б – границы устойчивого приближения при варьировании R_1 растут по сравнению со случаем A, а при варьировании R_2 растут, когда преобладают меньшие частицы и почти не меняются при преобладании больших частиц.

Следует отметить, что введение 25% гауссовского уровня шума не меняет область по R_1 , в то время как область по R_2 растет, когда преобладают меньшие частицы и почти не меняется при преобладании больших частиц. Наличие же 25% пуассоновского шума существенно увеличивает диапазоны устойчивости.

В – Области по R_1 сужаются, а по R_2 расширяются по сравнению со случаями А и Б.

Из вышесказанного следует, что в случае узкого распределения частиц диапазоны стартовых значений параметров распределений, обеспечивающих устойчивое приближение (будем называть такие диапазоны областями сходи**мости**) при варьировании радиусов сфер R_1 или R_2 уже, чем для случаев с широким распределением. Также видно, что чаще всего при добавлении гауссовского шума области сходимости сужаются, а при добавлении пуассоновского шума – расширяются. Это связано с тем, что добавленный гауссовский шум имел равномерно распределенную дисперсию по угловому диапазону, тогда как шум Пуассона, дисперсия которого равна математическому ожиданию интенсивности, гораздо больше искажает области малых значений интенсивности и относительно меньше – область начальных углов. Начальный участок кривой рассеяния отвечает за форму распределения больших и средних частиц, а наиболее искаженная шумами область слабых интенсивностей в больших углах содержит информацию в основном о малых частиц и атомном фоне. При этом более стабильный поиск распределения неоднородностей больших размеров при относительно невысоком уровне шумов облегчает программе построение распределения частиц меньших радиусов. В каком-то смысле происходит поиск распределения при "разделенных формфакторах". Этот подход

используют при анализе многокомпонентных систем с большими и малыми размерами неоднородностей: модели больших частиц строят по начальному участку кривой рассеяния, информацию о параметрах малых частиц извлекают из среднеуглового участка интенсивности.

Результаты по полидисперсности:

А – Допустимые стартовые отклонения по полидисперсности малых сфер (для случая 1, доля таких сфер 75%) лежат в пределах 0.1-4.5 нм в отсутствии шума, 0.1-5.0 нм – при наличии 25% гауссовского и пуассоновского шума. При понижении объемной доли малых сфер до 25%, область допустимых отклонений по их полидисперсности уменьшается до 0.01-4.0 нм без шума, 0.01-4.3 нм с гауссовским шумом и остается на том же уровне в случае добавления пуассоновского шума. Диапазоны устойчивого приближения по полидисперсности больших сфер (dR_2) составляют 0.01-15.0 нм для теоретических кривых без шума и с гауссовским шумом, а при наличии пуассоновского шума границы расширяются до диапазона 0.01-20.0 нм.

Итак, диапазоны, в которых можно варьировать значения полидисперсностей, достаточно велики. Это означает, что к степени полидисперсности сфер алгоритм более устойчив. Такой же характер зависимости справедлив и для случаев **Б** и **В**.

По объемным долям:

Во всех случаях объемные доли сфер могут варьироваться в полном диапазоне (0.0; 1.0) и не влияют на результаты приближения. Поэтому в табл. 2.2, эти результаты не приведены.

Далее для каждого случая аналогично проводилось исследование сходимости при одновременном изменении значений двух исходных параметров. Результаты моделирования приведены на рис. 2.2, где представлены двумерные контурные графики границ стартовых значений параметров распределений, обеспечивающих сходимость к правильному решению (далее – контурные графики устойчивости), зависящие от следующих пар параметров: (R_1 , R_2), (R_1 , dR_2), (dR_1 , R_2), (dR_1 , dR_2). На рис. 2.2 (a-е) приведены результаты для первой, второй, третьей, четвертой, пятой, шестой модельных систем соответственно. Пунктирной линией отмечены теоретические значения параметров (из табл. 2.1). Если в результате моделирования с помощью программы MIX-TURE, найденные значения параметров отличались от исходных более, чем на 5%, то такому решению присваивался статус «неудача» (области синего цвета), если найденные параметры находились в пределах 5% от заданных, это соответствовало «успешному» нахождению решения (области красного цвета). Уровень отклонений выбран 5%, так как изменения в значениях χ^2 для найденных решению оданных.

Как видно из графиков, характер зависимости устойчивости приближения при варьировании двух параметров аналогичен результатам моделирования при изменении одного из них. Области нахождения правильных решений зависят как от конкретной пары варьируемых параметров, так и от наличия дополнительного шума в данных. Они образуют достаточно сложные границы фрактального вида, с отдельными «очагами» неустойчивости внутри данных диапазонов, что свидетельствует о разной чувствительности метода минимизации к определенным параметрам модели.

Таким образом, были проведены исследования устойчивости восстановления правильных решений для полидисперсной системы, состоящей из сферических частиц двух типов. Были рассмотрены шесть модельных систем (случай узких распределений двух сферических частиц разного радиуса, случай широкого распределения малых сфер и узкого распределения больших сфер, а также случай узкого распределения малых сфер и широкого распределения больших сфер). Объемные распределения для этих систем не

64

перекрывались. В следующем разделе рассмотрены системы сферических частиц с частично перекрывающимися распределениями. При этом важно отметить, что параметры изучаемых систем были известны заранее, и это позволяло понимать, соответствует ли решение глобальному минимуму.

На Рис. 2.2а-е представлены контурные графики границ стартовых значений параметров распределений, обеспечивающих сходимость к правильному решению (далее – контурные графики устойчивости). Показаны зависимости от стартовых значений радиусов сферических частиц и их полидисперсностей $(R_1, R_2), (R_1, dR_2), (dR_1, R_2), (dR_1, dR_2)$. Левая колонка – относительный уровень шума 0%, по центру – гауссовский шум 25%, правая – пуассоновский шум 25%. Пунктирной линией отмечены теоретические значения параметров из табл. 2.1. Синий цвет – области «неудач» (отличие решения от исходных параметров более чем на 5%), красный цвет – «успешное» нахождение решения (найденные параметры – в пределах 5% от заданных). а – первая модельная система, б – вторая, в – третья, г – четвертая, д – пятая, е – шестая. Для лучшей визуализации перекрестия отмечены желтыми кружочками.



Рис. 2.2(а). Модельная система 1.



Рис. 2.2(б). Модельная система 2.



Рис. 2.2(в). Модельная система 3.



Рис. 2.2(г). Модельная система 4.



Рис. 2.2(д). Модельная система 5.



Рис. 2.2(е). Модельная система 6.

2.2. Системы с частично перекрывающимися распределениями сферических частиц по размерам

Исследования, аналогичные представленным в разделе 2.1., также проводились и для теоретических модельных систем с частично перекрывающимися объемными распределениями частиц по размерам. Результаты опубликованы в статье [А3].

В качестве модельных систем были выбраны три полидисперсные системы, состоящие из двух типов сферических частиц, имеющие частично перекрывающиеся распределения частиц по размерам. Исходные параметры частиц приведены в табл. 2.3. Рассчитанные кривые МУР и соответствующие им функции объемного распределения по размерам представлены на рис. 2.3.

Таблица 2.3. Исходные параметры трех теоретических кривых с частично перекрывающимися распределениями частиц по размерам, для которых проводится исследование на устойчивость решения

Mo-	V_1 ,	<i>V</i> ₂ ,	R_1 ,	R_2 ,	dR_1 ,	dR_2 ,
дель	отн.ед.	отн.ед.	HM	НМ	НМ	НМ
1	0.75	0.25	6.0	8.0	0.1	1.0
2	0.75	0.25	7.0	8.0	0.3	0.3
3	0.50	0.50	6.0	8.0	0.7	0.7

Контурные графики диапазонов устойчивости решения при одновременном варьировании пары параметров (*R*₁, *R*₂) представлены на рис. 2.4.

В целом результаты моделирования имеют аналогичные тенденции, что и для теоретических моделей с неперекрывающимися распределениями частиц по размерам. Но на рис. 2.2 и 2.4 видно, что диапазон стартовых значений параметров может быть шире для модельных систем с частично перекрывающимися распределениями.


Рис. 2.3. Три модельные кривые МУР (слева) и соответствующие им объемные распределения по радиусам сферических частиц $D_V(R)$ (справа): кривые МУР с относительными уровнями шума 0% и 25% (І - гауссовское распределение, ІІ - пуассоновское распределение). Для лучшей визуализации представлены со смещением по вертикали.



Рис. 2.4. Контурные графики устойчивости восстановления решения в зависимости от пары параметров (R_1 , R_2) и различных уровней шумов. Левая колонка – относительный уровень шума 0%, по центру – гауссовский шум 25%, правая – пуассоновский шум 25%. Пунктирной линией отмечены теоретические значения параметров из табл. 2.3. Синий цвет – «неудача» (отличие от исходных параметров более чем на 5%), красный цвет – «успешное» нахождение решения (найденные параметры – в пределах 5% от заданных): а – первая модельная система, б – вторая, в – третья.

Заключение к главе 2.

В главе 2 представлены результаты систематического исследования на устойчивость поиска параметризованных распределений с помощью квазиньютоновского градиентного алгоритма Бройдена-Флетчера-Гольдфарба-Шанно, реализованного в программе MIXTURE, для теоретических данных МУР от полидисперсных систем двух сферических частиц, рассчитанных от шести различных моделей с неперекрывающимися распределениями частиц по размерам [A1, A2], а также от трех моделей с частично перекрывающимися распределениями [A3]. Результаты моделирования показали схожий результат. Для этого были проведены старты программы MIXTURE при варьировании одного параметра, а также двумерные «карты» допустимых отклонений исходных значений параметров (объемных долей, радиусов сфер и степени полидисперсности).

Кроме того, исследовано влияние уровня шумовой составляющей в данных на степень отклонения найденного решения от заданной теоретической модели по данным от теоретических модельных систем. Для этого к наборам теоретических данных добавляли шум двух типов (гауссовский и пуассоновский). Было показано, что диапазон области стартовых значений параметров для восстановления правильного решения зависит от многих факторов, в частности, от относительного вклада параметров компонентов смеси в общую кривую рассеяния и от уровня и вида шума, содержащегося в данных. Было выяснено, что пуассоновский шум в данных МУР с относительным уровнем в области малых значений интенсивностей 25% способен увеличить область устойчивости (диапазон сходимости) минимум на 10%, что способствует лучшему восстановлению решения. Наличие гауссовского шума в данных, наоборот, приводит к сужению границ устойчивости.

На основе полученных результатов исследований, описанных в главе 2, была доработана программа POLYMIX (модификация программы MIXTURE): предусмотрен факт того, что распределения средних размеров частиц могут оказываться перекрывающимися, а программа такое перекрывание ограничивает. Кроме того, после этого появилась идея создания программы прямого поиска гистограммы VOLDIS.

ГЛАВА 3. РАЗРАБОТКА ЭФФЕКТИВНОГО АЛГОРИТМА ПОИСКА РЕШЕНИЙ ПРИ АНАЛИЗЕ ДАННЫХ МАЛОУГЛОВОГО РАССЕЯ-НИЯ

В главе 2 представлены результаты исследований для теоретических модельных данных. По построенным двумерным картам видно, что для нахождения истинного решения, нужно иметь достаточно хорошие стартовые значения. В данной главе описан подход, позволяющий добиться хорошего начального приближения.

Для этого были выбраны алгоритмы (описанные в разделе 1.3), которые активно используются в МУР при исследовании полидисперсных объектов.

(1) прямой поиск гистограммы методом линейных наименьших квадратов с регуляризацией решения по Тихонову (GNOM [4]);

(2) прямой поиск распределения частиц по размерам в виде ограниченно гладкой гистограммы методом нелинейных наименьших квадратов (VOLDIS [A6]);

(3) постулирование определенного вида распределения в аналитическом виде и проведение приближения данных методами нелинейных наименьших квадратов (MIXTURE[5], POLYMIX [A6]).

В программах GNOM и VOLDIS используют только один тип формы частиц, чаще всего – сферы, в то время как в программах MIXTURE и POLYMIX есть возможность проводить приближение различными геометрическими телами (двухслойными сферами, цилиндрами, эллипсоидами, гантелеобразными частицами), а также учитывать влияние межчастичной интерференции, вводя еще одну функцию - структурный фактор.

Каждый из этих алгоритмов обладает своими преимуществами и недостатками, и использование только одного из них не всегда позволяет получать устойчивые решения, свободные от артефактов. В данной главе предложена новая схема совместного использования алгоритмов, позволяющая повысить надежность восстановления функции распределения частиц по размерам. Работа схемы проиллюстрирована на примерах двухкомпонентных и трехкомпонентных систем полидисперсных сферических частиц с разнесенными и частично перекрывающимися распределениями компонент.

3.1. Оценка эффективности работы алгоритмов поиска функции распределения на примере двух- и трехкомпонентных полидисперсных систем

Были рассмотрены случаи двухкомпонентных и трехкомпонентных систем полидисперсных сферических частиц с разнесенными и частично перекрывающимися распределениями компонентов. Для этих систем были симулированы наборы данных малоуглового рассеяния без наличия шума и с добавлением шума, подчиняющегося Пуассоновскому распределению. В последнем случае оптимальный угловой диапазон данных выбирался согласно оценке программы SHANUM, которая использует шенноновское приближение для оценки информативности данных. Структурные параметры систем приведены в табл. 3.1.

Как видно по незашумленным данным, представленных на рис. 3.1, все три метода (линейных наименьших квадратов с регуляризацией по Тихонову, нелинейных наименьших квадратов и прямой непараметрический поиск) восстанавливают решение с удовлетворительной для практики точностью (относительная ошибка не превышает 0.1-0.2% на графиках отклонение не видно, и поэтому рисунки не приведены). Естественно, по зашумленным кривым рассеяния, показанным на рис. 3.1, удается найти только приближенное распределение. При этом форма функции распределения, восстанавливаемая программой GNOM методом линейных наименьших квадратов с регуляризацией (синие пунктирные кривые на рисунках), оказывается наиболее отличающейся от точного решения (сплошные красные линии на рисунках). Если положения пиков распределений определяются достаточно точно (отклонения не превышают 5%), то ширины распределений оказываются намного больше заданных при построении моделей. Кроме того, найденные распределения имеют артефакты в виде осцилляций из-за эффекта обрыва функции интенсивности рассеяния. Регуляризация, используемая в программе GNOM, хотя и позволяет находить достаточно гладкие решения (требование гладкости и приводит к уширению пиков) с малой амплитудой артефактов, тем не менее может неверно передавать форму распределения в области начальных (малых) размеров из-за чрезмерного ограничения первой производной по контуру.

Таблица 3.1. Структурные параметры двух- и трехкомпонентных систем. R_i – средний радиус сфер *i*-компоненты, dR_i – ширина распределения *i*-компоненты, v_i – объемная доля *i*-компоненты (i=1÷3).

Номер системы	Компонента 1	Компонента 2	Компонента 3
1	$R_{I} = 5.0$ нм	$R_2 = 12.0$ нм	
	$dR_1 = 1.5$ нм	$dR_2 = 3.0$ нм	
	$v_1 = 0.70$	$v_2 = 0.30$	
2	$R_1 = 6.0$ нм	$R_2 = 11.0$ нм	
	$dR_1 = 4.0$ нм	$dR_2 = 2.0$ нм	
	$v_1 = 0.10$	$v_2 = 0.90$	
3	$R_1 = 8.0$ нм	$R_2 = 12$ нм	
	$dR_1 = 4.0$ нм	$dR_2 = 3.0$ нм	
	$v_1 = 0.60$	$v_2 = 0.40$	
4	$R_1 = 4.0$ нм	$R_2 = 15$ нм	
	$dR_{1} = 0.7$ нм	$dR_2 = 1.5$ нм	
	$v_1 = 0.20$	$v_2 = 0.80$	
5	$R_1 = 4.0$ нм	$R_2 = 8.0$ нм	$R_3 = 15.0$ нм
	$dR_1 = 0.5$ нм	$dR_2 = 1.5$ нм	$dR_3 = 3.0$ нм
	$v_1 = 0.70$	$v_2 = 0.20$	$v_3 = 0.10$
6	$R_1 = 3.0$ нм	$R_2 = 8.0$ нм	$R_3 = 22.0$ нм
	$dR_1 = 0.4$ нм	$dR_2 = 2.0$ нм	$dR_3 = 3.0$ нм
	$v_1 = 0.005$	$v_2 = 0.078$	<i>v</i> ₃ =0.917
7	$R_{I} = 5.0$ нм	$R_2 = 8.0$ нм	$R_3 = 16.0$ нм
	$dR_1 = 2.0$ нм	$dR_2 = 1.0$ нм	$dR_3 = 5.0$ нм
	$v_1 = 0.10$	$v_2 = 0.70$	$v_3 = 0.20$
8	$R_1 = 37.0$ нм	$R_2 = 65.0$ нм	$R_3 = 130$ нм
	$dR_1 = 7.0$ нм	$dR_2 = 4.0$ нм	$dR_3 = 5.5$ нм
	$v_1 = 0.01$	$v_2 = 0.89$	$v_3 = 0.10$

Результаты программы непараметрического поиска гистограммы распределения VOLDIS для зашумленных данных (кривые на рисунках показаны зеленым пунктиром) также содержат отличия от точного решения (красные сплошные линии), причем особенно неустойчивыми также оказываются области малых размеров (особенно хорошо это видно на примере рис. 3.1(6)). Этот эффект связан с тем, что парциальные кривые рассеяния от частиц близких размеров очень похожи, и программа, увеличивая вклад частиц какоголибо радиуса, компенсирует чрезмерный рост интенсивности рассеяния уменьшением вклада от соседних частиц с близкими размерами. При малых же размерах сходство форм кривых рассеяния (например, по критерию корреляции) намного больше, чем в области больших размеров. Это связано с тем, что рассеяние от малоразмерных частиц в экспериментальной области углов представляет собой монотонные медленно спадающие функции, тогда как при больших размерах парциальные кривые содержат особенности в виде максимумов (по крайней мере, на производных), имеющих разные угловые положения.

Программа POLYMIX (аналогично MIXTURE) позволяет находить решения (методом нелинейных наименьших квадратов), практически совпадающие с точными во всех представленных случаях (сплошные красные кривые на рисунках, визуально не отличимые от точных контуров) при старте с приближений, близких к решениям, полученным методами непараметрического поиска распределения и прямого поиска с регуляризацией. Однако, решения могут содержать и значительные отличия, если стартовое приближение далеко от истинного. Как было показано ранее в работе [A2], диапазон области стартовых значений параметров распределений, с которых возможно восстановление правильного решения, зависит от многих факторов, в частности, от относительного вклада параметров компонентов смеси в общую кривую рассеяния и от уровня и вида шума в данных. Поэтому естественной кажется мысль использовать в качестве стартовых параметры, оцененные по решениям GNOM и MIXTURE/POLYMIX.

































Рис. 3.1. (А) Данные МУР от двух- (1-4) и трехкомпонентных (5-8) систем полидисперсных сферических частиц: красная кривая – зашумленные данные, синяя кривая – приближение методом линейных наименьших квадратов с регуляризацией по Тихонову (GNOM), зеленая кривая – прямой поиск распределения в виде гистограммы (VOLDIS). (Б) Заданная функция распределения по размерам (красная кривая) и восстановленные функции распределения: синяя кривая – GNOM, зеленая кривая – VOLDIS. Рассеяние от точной модели визуально совпадает с решением методом нелинейных наименьших квадратов (POLYMIX). Структурные параметры системы приведены в табл. 3.1.

3.2. Новая схема комбинированного использования алгоритмов поиска функции распределения

На основании сравнения полученных результатов восстановления функции распределения частиц по размерам разными алгоритмами, предложена следующая схема их комбинированного использования:

1) использовать прямой поиск распределения частиц по размерам в виде гистограммы для оценки общей формы распределения, числа компонентов (т.е. вкладов, представленных унимодальными распределениями) и среднего радиуса частиц в каждом компоненте. При этом особое внимание следует обратить на область углов рассеяния от 0 до начального экспериментального. Теоретическая кривая интенсивности должна монотонно экстраполировать экспериментальный контур, в противном случае найденное решение будет содержать артефакты в виде значительных вкладов больших размеров частиц. Для того чтобы этого избежать, необходимо провести повторный поиск, искусственно уменьшая R_{max} ;

2) использовать метод линейных наименьших квадратов с регуляризацией по Тихонову для нахождения предварительной оценки функции распределения, указывая в ней максимальный размер частиц R_{max} , подобранный на предыдущем шаге. Цель данного численного эксперимента является получение гладкого контура, уточнения числа и параметров компонентов;

3) уточнить параметры распределения с помощью нелинейного метода наименьших квадратов, используя в качестве стартового приближения параметры распределений, оцененные на предыдущих шагах.

Заключение к главе 3.

Для анализа полидисперсных систем в наноразмерном диапазоне существует ряд методов, позволяющих определять профиль распределения частиц по размерам из данных МУРР и МУРН. Соответствующая задача наименьших квадратов плохо обусловлена, что может приводить к сильной зависимости решения от стартового приближения и параметров алгоритмов поиска. Процедура, основанная на решении системы линейных уравнений с регуляризацией, дает стабильное решение, но не свободное от артефактов. Решения, предлагаемые другими методами, которые ищут функцию распределения как в аналитическом виде, так и в виде непараметрической гистограммы, могут зависеть от параметров поиска и начального приближения.

Для решений, найденных методом прямого поиска распределения с регуляризацией и методом непараметрического поиска, свойственны артефакты в виде паразитных осцилляций, причем для прямого поиска в области малых размеров возможны артефакты из-за ограничений на кривизну контура распределения. А для непараметрического поиска в отличие от прямого поиска, паразитные осцилляции связаны только с корреляцией интенсивности рассеяния от частиц с близкими размерами, т.е. имеют "физический" характер, а не обусловлены абстрактными математическими особенностями метода регуляризации. В случае метода нелинейных наименьших квадратов получаются сложные контуры, не описываемые одним аналитическим распределением, ищут в виде суперпозиции нескольких аналитических компонентов (на практике достаточно 2-6 компонентов).

Работа схемы продемонстрирована на ряде двухкомпонентных и трехкомпонентных систем полидисперсных сферических частиц с разнесенными и частично перекрывающимися распределениями компонент, как в случае отсутствия шума в данных, так и с наличием шумовой составляющей.

Итак, в главе 3 описана новая схема комбинированного использования этих методов, включающая в себя анализ методами регуляризации, прямого поиска гистограммы и в виде суперпозиции гладких аналитических функций, которая позволила улучшить устойчивость восстановления функции распределения частиц по размерам и, как следствие, его надежность. Результаты исследований опубликованы в статьях [А6, А7].

84

ГЛАВА 4. ПОДБОР ЭФФЕКТИВНОГО МЕТОДА ОПТИМИЗАЦИ ДЛЯ ПОИСКА РЕШЕНИЙ НА ПРИМЕРЕ ДАННЫХ МАЛОУГЛОВОГО РЕНТГЕНОВСКОГО РАССЕЯНИЯ ОТ РАСТВОРА КРЕМНЕЗОЛЯ

В главе 4 на примере полученных экспериментальных данных МУРР от двухкомпонентного раствора кремнезоля проведены исследования на устойчивость, по аналогии с рассмотренными в главе 2 теоретическими модельными данными. Первоначальный анализ, как для теоретических кривых, так и для раствора кремнезоля, был проведен программой MIXTURE с минимизационным алгоритмом **BFGS**. Было установлено, что даже при известных заранее параметрах не всегда удается получить истинное решение. Поэтому с целью поиска наиболее эффективного алгоритма были проверены другие оптимизационные схемы, реализованные в виде модификации программ.

4.1. Моделирование с использованием экспериментальных данных

В целях оценки устойчивости решений для системы, состоящей из сферических частиц двух типов, по данным МУР было проведено моделирование с использованием экспериментальных данных, полученных от двухкомпонентного раствора кремнезоля. С помощью методов линейных наименьших квадратов с регуляризацией (программа GNOM [4]) и нелинейных наименьших квадратов (программа MIXTURE [5]) независимо было обнаружено наличие полидисперсных компонент двух типов. Были найдены структурные параметры, описывающие систему и оценены диапазоны этих параметров, при которых возможно восстановить правильное решение.

4.1.1. Кремнезоли. Полидисперсная система сферических частиц

Объектом для исследования был выбран раствор кремнезоля, полученный путем смешения узкодисперсных растворов кремнезолей Ludox TM- 50^{TM} и SMTM. Образцы TM-50 (50 мас.% SiO₂ и 0.3 мас.% Na) и SM (30 мас.% SiO₂ и 0.2 мас.% Na) были изготовлены компанией Grace Davidson [103]. Раствор

кремнезоля разводили водой в соотношении 1:100, чтобы избежать эффекта межчастичного взаимодействия частиц, при этом центрифугирование не использовалось, так как растворы изначально стабилизировались с использованием гидроксида натрия 0.06М NaOH для предотвращения агрегации. При разбавлении дистиллированной водой растворы устойчивы во времени без дополнительного добавления NaOH. Такие растворы кремнезолей используются, в частности, при создании ленгмюровских фосфолипидных мультислоев на границе поверхности воды [104], которые представляют большой интерес для медико-биологический приложений. Также растворы кремнезолей используются при прецизионной полировке зеркал, при этом важно, чтобы размер частиц не превышал 300 Å.

Эксперимент по исследованию раствора кремнезоля с помощью малоуглового рассеяния проводили на лабораторной установке «АМУР-К» [87] следующим образом: образец кремнезоля помещали в стеклянный капилляр, а затем в вакуумную камеру. Расстояние образец-детектор составляло 700 мм. Время измерений одного образца составляло 10 минут. Экспериментальные данные были нормированы на интенсивность падающего пучка, после чего в них вводилась поправка на коллимационные искажения и вычитали интенсивность рассеяния от капилляра, заполненного водой.

В результате измерений получены экспериментальные данные МУР для раствора кремнезоля (рис. 4.1а, точки). Из экспериментальной кривой МУР видно, что шум соответствует статистике Пуассона, при этом относительный уровень шума на хвосте кривой составляет 15%.

Предварительная обработка данных МУР методом линейных наименьших квадратов выявила наличие двух полидисперсных компонент. Соответствующая модельная кривая показана на рис. 4.1а (красная пунктирная линия), функция распределения парных расстояний $D_V(R)$ представлена на рис. 4.1б (красная пунктирная линия). Модельная кривая хорошо согласуется с экспериментальными данными. Из восстановленного распределения частиц по размерам видно, что большие частицы имеют радиус $R \sim 140$ Å; кроме того, в растворе присутствуют меньшие частицы с радиусом $R \sim 59$ Å. Полидисперсность каждого из компонент составляет около 15%, таким образом, распределения компонент практически не перекрываются друг с другом.



Рис. 4.1. (а) Экспериментальная кривая МУР с относительным уровнем пуассоновского шума 15% (точки). Наилучшее приближение к экспериментальным данным, рассчитанное нелинейным методом наименьших квадратов (черная сплошная линия). Наилучшее приближение к экспериментальным данным, рассчитанное методом линейных наименьших квадратов с регуляризацией по Тихонову (красная пунктирная линия). (б) Объемное распределение по радиусам сферических частиц, восстановленное с использованием нелинейным методом наименьших квадратов (черная сплошная линия). Функция распределения парных расстояний, рассчитанная с использованием метода линейных наименьших квадратов с регуляризацией по Тихонову (красная пунктирная линия).

Размеры (диаметры) частиц, заявленные производителем (для сухих образцов), составляют 70 Å для SM [105] и 220 Å для TM-50 [106]. Однако стоит отметить, что данные оценки были получены из анализа численных распределений частиц по размерам, тогда как по данным МУРР оценивались объемные распределения частиц по размерам. Если объемные распределения путем деления на объем частиц перевести в численные $D_N(R_i) = D_V(R_i)/v(R_i)$, то получатся следующие значения средних размеров (диаметров) частиц: 104 Å для SM и 261 Å для TM-50. Результаты находятся в хорошем соответствии с данными от производителя, если учесть, что в растворе поверхностный слой частиц кремнезоля из-за гидролиза состоит в значительной степени из поликремниевых кислот, что должно увеличивать их размер по сравнению с негидролизованными частицами. Потеря воды должна происходить в вакууме электронного микроскопа от 5 до 50% в зависимости от pH и размера частиц [107]. Кроме того, полученные значения согласуется с результатами работы [104], где с помощью МУРР было показано, что в растворе размеры частиц увеличиваются до 100 Å для SM и 270 Å для TM-50.

Но в то же время в статье [108] приведены данные, полученные для раствора кремнезоля ТМ-50, на акустическом спектрометре (32.1 нм), методом лазерной дифракции (29.9 нм) и методом динамического светорассеяния (34.1 нм). Такие размеры превышают значения, полученные в работе методом МУРР, и это указывает на то, что размеры частиц имеют большой разброс, находясь в растворе [107].

Распределение, полученное методом прямого поиска с регуляризацией (рис. 4.16, красная пунктирная линия) имеет асимметричный вид при малых значениях размеров частиц, что обусловлено наличием эффектов обрыва ряда при использовании непрямого Фурье преобразования. Чтобы восстановить более адекватное с физической точки зрения распределение частиц по размерам, использован метод нелинейных наименьших квадратов в приближении системы двух типов сферических полидисперсных частиц. Полученное распределение, позволившее приблизить экспериментальные данные МУР (рис. 4.1а, черная сплошная линия), представлено на рис. 4.1б (черная сплошная линия), оно качественно хорошо согласуется с ранее полученным распределением методом линейных наименьших квадратов (в пределах 10% по среднему диаметру и 20% по полуширине распределений), соответствующие положения максимумов и ширины распределений приведены в таблице 4.1.

Таблица 4.1. Найденные параметры системы из двух сферических наночастиц раствора кремнезоля.

<i>V</i> ₁ , отн. ед.	V ₂ , отн. ед.	$R_1, \text{\AA}$	$R_2, \text{\AA}$	$dR_1, \text{\AA}$	$dR_2,$ Å
0.25	0.75	58.8	139.4	10.00	17.9

На основе экспериментальных данных МУР проводилось исследование на устойчивость программы MIXTURE. Исследовали сходимость расчетов к правильному решению <u>при поочередном варьировании</u> стартовых значений параметров (V_1 , V_2 , R_1 , R_2 , dR_1 и dR_2). Другим параметрам при этом присваивали правильные стартовые значения. Результаты приведены в таблице 4.2. Показаны диапазоны устойчивого приближения при варьировании значения одного из стартовых параметров.

Таблица 4.2. Диапазоны устойчивого приближения при варьировании одного параметра

$dR_1, \text{\AA}$	dR_2 , Å	$R_1, \text{\AA}$	<i>R</i> ₂ , Å	<i>V</i> ₁ , отн. ед.	V ₂ , отн. ед.
0.5–50.0	0.5–50.0	24.0–96.0	100.0–216.0	0.0–1.0	0.0–1.0

Далее в каждом случае проводилось аналогичное исследование сходимости при <u>одновременном изменении значений двух исходных параметров</u>. Результаты моделирования приведены на рис. 4.2. Там представлены двумерные контурные графики, зависящие от следующих пар параметров: (R_1, R_2) , (R_1, dR_2) , (dR_1, R_2) , (dR_1, dR_2) , (V_1, R_2) , (R_1, V_2) . Пунктирной линией отмечены восстановленные точные значения параметров, указанные в табл. 4.1. Если в результате моделирования с помощью программы MIXTURE найденные значения параметров отличались от точных более чем на 5%, то такому решению присваивали статус «неудача» (на рис. 4.2. области синего цвета). Если найденные параметры находились в пределах 5% от требуемых значений, это соответствовало «успешному» нахождению решения (на рис. 4.2. области красного цвета).

Применялся и другой способ оценивания стабильности решений, основанный на варьировании углового диапазона данных, используемых при поиске модели распределения. При этом варьировали максимальный угол на -20, -10 и +10% от диапазона, использованного в предыдущем методе. Цель данного способа состоит в определении оптимального диапазона углов путем анализа разброса параметров распределений. Можно выделить следующие результаты моделирования.

Зависимость по полидисперсности (ширине распределений) частиц.

При варьировании только одного параметра dR_1 или dR_2 правильное решение находилось во всем заданном в программе диапазоне. Другими словами, значения параметров полидисперсностей не влияют на результаты приближения при условии успешного нахождения средних значений радиусов частиц. Таким образом, в таких условиях алгоритм программы устойчив к изменениям параметров dR_1 и dR_2 .



Рис. 4.2. Контурные графики устойчивости восстановления решения в зависимости от пар параметров $(R_1, R_2), (R_1, dR_2), (dR_1, R_2), (dR_1, dR_2), (V_1, R_2), (R_1, V_2) для раствора кремнезоля. Пунктирной линией отмечены значения параметров, восстановленные программой MIXTURE из табл. 4.1. Синий цвет – «неудача» (отличие от исходных параметров более чем на 5%), красный цвет – «успешное» нахождение решения (найденные параметры – в пределах 5% от заданных).$

Что касается варьирования двух параметров одновременно, то можно увидеть на графике зависимости (dR_1 , dR_2), что вся область красная. Это значит, что при любых значениях полидисперсности программа MIXTURE найдет правильное решение. В то же время, на графиках (R_1 , dR_2) и (dR_1 , R_2)

для параметров dR_1 и dR_2 видна выраженная граница, в широких пределах которой находится правильное решение.

<u>Зависимость по объемным долям.</u> Найденные объемные доли компонентов могут варьироваться в широком диапазоне. При варьировании одного параметра (табл. 4.2.) этот диапазон полный (0.0;1.0). А при варьировании одновременно двух параметров также полный диапазон, но при допустимых значениях радиусов, что иллюстрируют зависимости (V_1 , R_2), (R_1 , V_2) на рис. 4.2.

В результате анализа обнаружено, что характер зависимости устойчивости приближения при варьировании двух параметров аналогичен результатам моделирования при изменении одного из них.

Полученные результаты можно сравнить с проведенным нами ранее исследованием устойчивости для теоретических модельных систем [A1-A3]. В целом, рассмотренные ранее системы имели узкие распределения по размерам, что приводило к сужению границ устойчивых решений. В случае реальных данных с более широким распределением, диапазоны устойчивости становятся шире. Также положительную роль играет и наличие умеренного уровня пуассоновского шума в данных. Результаты по радиусам для теоретических наборов данных также, как и в рассматриваемой в данной работе экспериментальной кривой, имели выраженные границы устойчивости на двумерных картах, но более узкие. По полидисперсности результаты и в предыдущих, и в этом случае имеют широкий диапазон устойчивого восстановления данных. Что касается объемных долей, то для теоретических моделей значение V_1 или V_2 не влияло на приближение, тогда как для реальных данных появляется граница, но скорее всего из-за того, что был задан слишком большой диапазон по параметру R_2 .

Сравнение решений, полученных путем варьирования углового диапазона данных показало, что уменьшение диапазона на -20% от максимального значения, использованного в предыдущем методе, приводит к значительным, от 10 до 50% изменениям параметров распределений по сравнению с опорным решением, тогда как варьирование на $\pm 10\%$ вызывает отклонения средних

91

радиусов частиц на 2-4% и полуширины распределений на 3-7%, что является типичным для рассматриваемой задачи. Тем самым была подтверждена оптимальность диапазона, использованного в работе.

На основе полученных результатов можно сказать, что алгоритм программы MIXTURE устойчив в широких диапазонах параметров для систем сферических частиц с частично перекрывающимися объёмными распределениями частиц по размерам и описано это в статье [109]. При этом в данной публикации был неправильно назван образец (раствор кремнезоля TM-50 вместо смеси растворов TM-50 и SM), что никак не повлияло на результаты, так как целью было исследование на устойчивость алгоритма программы MIX-TURE. Кроме того, в приложении А приводятся уточненные экспериментальные данные МУРР от отдельных компонент и их смесей с различным весовым соотношением TM-50 и SM, которые подтверждают качественный состав исходных компонент смеси. В связи с этой неточностью публикация [109] в списке работ автора по теме диссертации не указана. В последующих статьях [А4, А5, А8] неточность была устранена, а на примере той же самой кривой МУРР от раствора кремнезоля исследовались другие оптимизационные методы.

4.2. Результаты применения методов оптимизации, рассмотренных в исследовании

В работе проводилось исследование на устойчивость алгоритмов четырех модификаций программы MIXTURE. В каждой из которых были использованы различные минимизационные схемы, описанные в разделе 1.8: метод **BFGS** [83], метод моделирования отжига **SA** (Simulated Annealing) [92], комбинация этих методов **BFGS+SA** (см. раздел 4.2.1), а также метод многогранника Нелдера-Мида **NM** (Nelder-Mead) [93].

4.2.1. Комбинирование методов оптимизации. Градиентный метод Бройдена-Флетчера-Гольдфарба-Шанно и метод моделирования отжига

Комбинация **BFGS** и **SA** представляет из себя поочередное использование этих алгоритмов минимизации. Схема поиска минимума показано на рис. 4.3. В целях поиска первого (возможно, локального) минимума нелинейной целевой функции (1.27), производят старт программы MIXTURE, использующей градиентный метод минимизации, с некоторых значений параметров двухкомпонентной полидисперсной модели (где N - один из структурных параметров R_1, R_2, dR_1 либо dR_2). Далее, с оптимизированных методом **BFGS** значений производится запуск модифицированной версии MIXTURE с использованием метода **SA** для того, чтобы вывести систему «из равновесия», причем со всего лишь одной итерацией по температуре, поэтому вычисления не занимают длительного времени. После этого с новых полученных значений программа запускается вновь с градиентным методом и так повторяется до тех пор, пока параметры модели не перестанут изменяться. За пороговое значение их изменчивости брали 1-5% от величины, полученной на предыдущей итерации.



Рис. 4.3. Схема поиска глобального минимума целевой функции при комбинированном использовании методов **BFGS** и **SA**.

4.2.2. Исследование устойчивости решений при анализе данных малоуглового рентгеновского рассеяния от раствора кремнезоля с использованием ряда оптимизационных схем

Результатам, описанным в разделе, предшествовал ряд работ, которые опубликованы в статьях [А4, А5].

После исследования раствора кремнезоля программой MIXTURE с **BFGS** возникла идея проверить эффективность работы модифицированной версии программы MIXTURE, в алгоритме которой реализован метод **SA**, разработанный для поиска глобального минимума целевой функции. Результаты такого исследования приведены в работе [A4]. В качестве объекта исследования был выбран тот же раствор кремнезоля, что и в предыдущей работе [109]. Анализ сходимости решения проводили при одновременном варьировании только одной пары параметров – средних радиусов частиц кремнезоля R_1 и R_2 . В результате было установлено, что обе модификации программы (с использованием **BFGS** и **SA** процедур) дали схожие результаты, но при этом границы для **BFGS** процедуры оказались шире. Далее была рассмотрена возможность комбинированного использования **BFGS** и **SA** [A5]. Предполагалось, что это может улучшить возможности восстановления распределений частиц по размерам. Действительно, такой подход позволил расширить границы параметров, при старте с которых могут быть найдены истинные значения системы.

Итак, в текущем разделе приведено систематическое сравнение границ устойчивого приближения для программы MIXTURE с различными минимизационными схемами. Для этого проводились исследования на устойчивость вышеупомянутых модификаций программы с **BFGS**, **SA** и **BFGS+SA** для трех пар параметров: (R_1, R_2) , (dR_1, R_2) и (R_1, dR_2) . Помимо этого, исследовались возможности алгоритма **NM**. Исследования проводились на примере экспериментальных данных малоуглового рассеяния от полидисперсного раствора кремнезоля [A8].

Исследована сходимость расчетов к правильному решению при одновременном варьировании стартовых значений следующих пар параметров: (R_1 , R_2), (dR_1 , R_2), (R_1 , dR_2), тогда как остальные параметры фиксировались на истинных значениях, найденных программой MIXTURE (табл. 4.3.) с объемными долями V_1 =0.25 для меньших частиц кремнезоля (SM) и V_2 =0.75 для больших частиц кремнезоля (TM-50). В табл. 4.3. эти значения не приведены, так как алгоритмы всех исследуемых модификаций программы MIXTURE устойчивы к изменению объемных долей во всем диапазоне от 0 до 1.

Таблица 4.3. Найденные структу	рные параметры с	системы, состоящей
из двух типов сферических наночастиц	кремнезоля - ТМ-5	б0 и SM.

	SM		TM-50	
	$R_1, \text{\AA}$	$dR_1, \text{\AA}$	$R_2, \text{\AA}$	dR_2 , Å
Метод Нелдера-Мида (NM)	59.8	10.0	140.0	17.9
Метод моделирования отжига (SA)	58.7	13.1	139.4	17.9
Квазиньютоновский градиентный ме-	58.8	10.0	139.4	17.9
тод (BFGS)				
BFGS+SA	58.8	10.0	139.4	17.9

Диапазоны, в пределах которых выбирались стартовые значения параметров, были следующими: для радиусов R_1 10.0-96.0 Å и R_2 100.0-500.0 Å, для полидисперсностей dR_1 и dR_2 0.5-50.0Å. Диапазоны были выбраны таким образом, чтобы радиусы меньших и больших частиц не коррелировали друг с другом. При этом шаг расчетной сетки по R_1 и R_2 составлял 2Å, шаг сетки по dR_1 и $dR_2 - 0.5$ Å. Таким образом, для пары параметров (R_1, R_2) было посчитано 8844 точек, для $(dR_1, R_2) - 20100$ точек, для $(R_1, dR_2) - 4400$ точек. На основе этих данных построены двумерные карты, представленные на рис. 4.4. Верхняя панель соответствует результатам, полученным модифицированной версией программы MIXTURE с реализованным методом NM, вторая – программе с методом SA, третья – с BFGS, четвертая (нижняя) панель – комбинации методов **BFGS** и **SA**. Первый столбец (слева) – для пары параметров (R_1 , R_2), второй – для (dR_1, R_2), третий – для (R_1, dR_2). Область красного цвета соответствует нахождению истинного решения (в случае, когда значения параметров, полученных в результате моделирования программой MIXTURE, находилось в пределах 5% от найденных ранее оптимальных значений). Область синего цвета соответствует всем остальным решениям, существенно отличающимся от истинного решения. Пунктирной линией отмечены истинные (оптимальные) значения параметров, восстановленные программой MIXTURE, указанные в табл. 4.3.

Анализ двумерных карт позволил установить, что самые широкие границы устойчивости имеются у комбинации методов **BFGS** и **SA** для всех трех пар параметров. При этом очевидно можно выстроить ряд алгоритмов в порядке возрастания их эффективности: **NM**, **SA**, **BFGS**, **BFGS**+**SA**, поэтому на рис. 4.4. они расположены в соответствующем порядке, сверху вниз.

Анализируя двумерные карты, стоит отметить, что метод **NM** оказался наименее эффективным по сравнению с другими использовавшимися методами. Несмотря на то, что его границы устойчивости шире, чем для метода **SA**, но сравнительно легко попасть в область неустойчивости при незначительном отклонении от истинного значения. Аналогичные результаты представлены в статье [110]. Авторами также был рассмотрен ряд алгоритмов, в том числе методы **BFGS**, **NM** и Trust region (это близко к тому, что используется в POLYMIX). В таком ряду алгоритмов наиболее эффективным оказался метод **BFGS**, а наименее - метод **NM**.



Рис. 4.4. Контурные графики устойчивости восстановления решения программой MIXTURE в зависимости от стартовых значений радиусов сферических частиц кремнезоля и их полидисперсностей (R_1 , R_2), (dR_1 , R_2), (R_1 , dR_2). Пересечение пунктирных линий соответствует точным (истинным) значениям параметров из табл. 4.3. Области синего цвета – «неудача» (отличие от истинных параметров более чем на 5%), области красного цвета – «успешное» нахождение решения (найденные параметры находятся в пределах 5% от истинных). Использованные алгоритмы минимизации: (a) NM, (б) SA, (в) BFGS, (г) BFGS+SA.

Для других трех методов уже можно говорить об имеющихся тенденциях. Для параметров R_1 и R_2 прослеживаются следующие закономерности: стартовые значения допустимо варьировать в пределах 50%, 65% и 80% для R_1 и 15%, 65%, 100% для R_2 от истинного значения для методов **SA**, **BFGS**, **BFGS** + **SA** соответственно. Во всех случаях при значении R_2 более 300Å не удалось добиться восстановления данных. То есть стартовое значение данного параметра может не более чем в 2 раза превышать его истинное значение. Вдобавок к вышеперечисленному, решения являются устойчивыми к варьированию параметров dR_1 и dR_2 во всем исследуемом диапазоне.

Также следует отметить, что метод **SA** теоретически мог бы позволить найти более широкие границы устойчивости, если бы мы существенно увеличили число вариаций модели при каждой итерации по температуре, но это на порядок увеличило бы время расчетов, что, как мы полагаем, неприемлемо с практической точки зрения. Кроме того, метод **SA** требует длительной работы по поиску оптимального режима снижения температуры, общего рецепта для этого до сих пор не выработано. В случае же комбинированной схемы **BFGS**+**SA**, метод **SA** используется исключительно с целью выхода из локального минимума и последующего продолжения поиска градиентным методом, для этой цели достаточно использования одной итерации по температуре. Такой подход также успешно работает и для других задач, что позволяет судить о его универсальности [111].

Если использовать предложенную новую комбинированную схему вместе с подходом, описанным в главе 3, то это позволит еще больше расширить диапазон сходимости к правильному решению и, таким образом, повысить достоверность определения распределений по размерам. На третьем шаге схемы (см. раздел 3.2) можно дополнительно проводить комбинирование локальных и глобальных минимизационных методов. Такой дважды комбинированный подход можно назвать новым **методом поиска распределений частиц по размерам с расширенной областью устойчивости.** В текущей главе описано как такой метод применен к экспериментальным данным МУРР от раствора кремнезоля. Подход также опробован на дополнительном реальном объекте (оксиде цинка) и описан в приложении Б.

Заключение к главе 4.

В главе 4 представлено исследование устойчивости восстановления решения задачи определения объемных распределений частиц по размерам программой MIXTURE на примере данных МУР от раствора кремнезоля, состоящего из узкодисперсных растворов ТМ-50 и SM.

Для раствора кремнезоля предварительно были найдены значения радиусов сферических частиц порядка 59 Å и 140 Å двумя независимыми методами (линейных наименьших квадратов с регуляризацией и нелинейных наименьших квадратов с параметрической моделью), хорошо согласующиеся с ранее полученными результатами [104]. Средние размеры (диаметры) из распределений по числу частиц, вычисляемых из найденных объемных распределений как $D_N(R_i) = D_V(R_i)/v(R_i)$, для кремнезолей ТМ-50 (26.1 нм) и SM (10.4 нм) превышают размеры, указываемые производителем: 22 и 7 нм соответственно для TM-50 и SM [105, 106]. Превышение составило для TM-50 16% и 33% для SM. Данное несоответствие является типичным для кремнезолей в растворе и обусловлено гидролизом в поверхностном слое частиц, приводящим к появлению оболочки из поликремниевых кислот [107]. Эта оболочка, стабильная в растворе, имеет более рыхлую структуру, что приводит к увеличению определяемого по данным рассеяния размера. С другой стороны, найденные для ТМ-50 размеры меньше определяемых в работе [108] на акустическом спектрометре (32.1 нм) на 19%, методом лазерной дифракции (29.9 нм) на 13% и методом динамического светорассеяния (34.1 нм) на 24%. В этой публикации раствор ТМ-50 предложен в качестве размерного стандарта. Резюмируя, можно сказать, что метод малоуглового рассеяния позволил уточнить опубликованные данные для наночастиц кремнезолей, находящихся в естественном состоянии в растворе.

Далее также, как и для теоретических модельных систем проведены исследования устойчивости восстановления правильных решений для раствора кремнезоля TM-50 и SM. Получены данные при изменении одного из стартовых параметров, а также при одновременном варьировании двух параметров. Проведенные ранее исследования устойчивости решений для теоретических наборов данных подтверждают справедливость их использования для реальных экспериментальных систем.

С целью расширить границы устойчивости исследован ряд итеративных алгоритмов: квазиньютоновский градиентный метод минимизации в варианте Бройдена-Флетчера-Гольдфарба-Шанно с простыми ограничениями на параметры (BFGS), метод моделирования отжига (SA), их комбинация (BFGS+SA), а также метод многогранника Нелдера-Мида (NM). Несмотря на то, что все эти методы являются достаточно эффективными, в ряде случаев можно только экспериментальным путем определить будет ли используемый метод подходить для конкретной поставленной задачи. Поэтому, чтобы найти наиболее эффективный алгоритм для решения задачи восстановления объемного распределения частиц по размерам для двухкомпонентной полидисперсной системы были протестированы все вышеупомянутые модификации программы MIXTURE.

Анализ сходимости решения проводили при варьировании трех пар параметров: (R_1, R_2) , (dR_1, R_2) и (R_1, dR_2) , где R_1 и R_2 представляют собой средние радиусы частиц кремнезоля, а dR_1 и dR_2 – их полидисперсности. В результате сравнения границ устойчивости для трех пар параметров раствора кремнезоля (радиусов и полидисперсностей частиц) методы выстроены в порядке возрастания их эффективности: **NM**, **SA**, **BFGS**, **BFGS**+**SA**. Таким образом, алгоритм поиска, с последовательным комбинированием методов минимизации **BFGS** и **SA**, для систем сферических частиц с частично перекрывающимися объемными распределениями частиц по размерам, позволяет максимально эффективно восстанавливать функцию $D_v(r)$ и в более широких диапазонах стартовых значений параметров, чем с использованием других предложенных алгоритмов. При этом, можно выделить несколько закономерностей. Для радиусов частиц R_1 и R_2 стартовые значения можно изменять в пределах 50%, 65% и 80% для R_1 и 15%, 65%, 100% для R_2 от истинного значения для методов **SA**, **BFGS**, **BFGS**+**SA** соответственно, одновременно с этим превышение стартового значения параметра R_2 более, чем в 2 раза затруднило поиск истинного решения (т.е. R_2 не должен был превышать 300Å). В то время как к варьированию dR_1 и dR_2 во всем диапазоне решения оказались устойчивы.

Результаты приведенных в главе 4 исследований опубликованы в статьях [A4, A5, A8].

Подход с сочетанием методов **BFGS+SA** в том числе опробован на экспериментальных данных МУРР от наночастиц оксида цинка и описан в приложении Б. Установлено, что для наночастиц ZnO положения основного пика на распределениях частиц по размерам по данным МУРР и просвечивающей электронной микроскопии совпали в пределах 5%, а полуширины распределений – в пределах 10%. При этом по изображениям просвечивающей электронной микроскопии невозможно оценить размеры малых частиц, тогда как данные МУРР предоставляют информацию о полном диапазоне размеров, от долей нанометра до субмикронных образований.

Таким образом, исследования, описанные в главах 3 и 4, привели к созданию нового метода поиска распределений сферических частиц по размерам с расширенной областью. В целом, развитые подходы применимы к исследованию систем различной природы, в том числе и для растворов белковых комплексов.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ

1. Впервые выполнено систематическое исследование устойчивости решений задачи поиска распределений сферических частиц по размерам из данных малоуглового рассеяния с помощью различных методов поиска. Анализ был проведен на ряде модельных и реальных объектов, состоящих из двух типов полидисперсных систем сферических частиц с перекрывающимися и частично перекрывающимися распределениями. Определены диапазоны сходимости решений при изменении одного и двух стартовых параметров модели в виде двумерных карт допустимых начальных значений среднего размера и полуширины распределений. По результатам модельных экспериментов были выработаны рекомендации для модифицирования программ POLYMIX и VOLDIS с целью повышения их эффективности.

2. На примере данных малоуглового рассеяния от раствора кремнезоля в сферическом приближении формы частиц проведено сравнение эффективности четырех минимизационных схем поиска распределений по размерам: метода многогранника Нелдера-Мида, метода моделирования отжига, квазиньютоновского градиентного метода Бройдена-Флетчера-Гольдфарба-Шанно и нового подхода – сочетания градиентного метода и метода моделирования отжига. Анализ полученных данных позволил расположить методы в порядке возрастания их эффективности: а) метод Нелдера-Мида, б) метод моделирования отжига, в) метод Бройдена-Флетчера-Гольдфарба-Шанно, г) комбинация метода Бройдена-Флетчера-Гольдфарба-Шанно, г) комбинация тогжига. Каждый метод был реализован в виде модификации программы MIX-TURE.

3. Предложена новая схема поиска распределений размеров частиц по данным малоуглового рассеяния, сочетающая в себе последовательный анализ методами прямого поиска с помощью регуляризации, непараметрического сглаживания распределения и в виде суперпозиции гладких аналитических функций. При этом результаты, полученные методами прямого поиска, используют для построения стартовых приближений для суперпозиционного метода, который также дает гладкое решение, но свободен от недостатков прямых методов. Такой подход позволил расширить диапазон сходимости к точному решению обратной задачи в пространстве параметров, описывающих структурную модель исследуемой системы во всех рассмотренных случаях с разбросом не хуже 1% по параметрам распределения. Анализ проведен на примере двух- и трехкомпонентных модельных полидисперсных систем сферических частиц.

4. Разработана новая схема комбинирования методов поиска распределений с чередованием двух принципиально разных алгоритмов минимизации – квазиньютоновского градиентного поиска Бройдена-Флетчера-Гольдфарба-Шанно и метода моделирования отжига. Данная схема позволила расширить область допустимых стартовых значений параметров модели, при которых возможно нахождения параметров распределений с разбросом не хуже 5%, не менее, чем на 15% по сравнению с подходами, основанными на использовании только одного из методов поиска. Комбинированный подход может быть использован как завершающий этап последовательного анализа методами прямого поиска с помощью регуляризации, непараметрического сглаживания распределения и в виде суперпозиции гладких аналитических функций. В совокупности, предложенные схемы представляют собой метод поиска распределений сферических частиц по размерам с расширенной областью сходимости. Проделанная работа обеспечивает базис для создания автоматизированного анализа.

5. Исследование влияния уровня шумов в данных малоуглового рассеяния на устойчивость решений для теоретических двухкомпонентных полидисперсных систем сферических частиц показало, что добавление гауссовского шума к данным уменьшает допустимые диапазоны стартовых значений параметров при поиске истинного решения, а наличие пуассоновского шума с относительным уровнем в области малых значений интенсивностей 25% — расширяет, при этом минимум на 10%.

103

6. Эффективность предложенных схем поиска распределений частиц по размерам с расширенной областью сходимости показана на примерах анализа данных рассеяния от растворов кремнезолей и полимерных матриц с наночастицами оксида цинка.

Средние размеры из распределений по числу частиц, вычисляемых из найденных объемных распределений для кремнезолей ТМ-50 (26.1 нм) и SM (10.4 нм) превышают размеры, указываемые производителем, на 16% для ТМ-50 и 33% для SM. При этом найденные размеры ТМ-50 меньше определяемых на акустическом спектрометре на 19%, методом лазерной дифракции на 13% и методом динамического светорассеяния на 24%.

Для оксида цинка положения основного пика на распределениях частиц по размерам по данным малоуглового рентгеновского рассеяния и просвечивающей электронной микроскопии совпали в пределах 5%, а полуширины распределений – в пределах 10%.

Приложение А.

АНАЛИЗ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ МАЛОУГЛОВОГО РЕНТГЕНОВСКОГО РАССЕЯНИЯ ОТ РАСТВОРА КРЕМНЕЗОЛЯ

Для проверки качества исследованных компонент бинарной смеси частиц кремнезоля SM и TM-50 в растворе были проведены дополнительные измерения методом МУРР как отдельных образцов, так и их смесей при двух различных весовых соотношениях исходных растворов концентрированных растворов (отношение SM:TM – 2:1 и 4:1). Измерения были выполнены на малоугловом дифрактометре АМУР-К [87]. Чтобы избежать влияния эффекта межчастичного взаимодействия исходные растворы были разбавлены в отношении 1:100. Экспериментальные кривые рассеяния после проведения процедуры сглаживания и введения коллимационной поправки представлены на рис. А1а, при этом видно, что кривая рассеяния от больших частиц кремнезоля ТМ-50 имеет больший наклон в области малых углов, тогда как на кривой рассеяния от меньших частиц кремнезоля SM наклон меньше, что характерно для частиц с меньшей степенью агрегации. Кривые рассеяния от бинарной смеси занимают промежуточное положение, что вполне ожидаемо, поскольку экспериментально нами регистрируется усредненный сигнал от обеих компонент смеси. Полученные данные рассеяния были обработаны программами GNOM и VOLDIS, в результате были получены распределения частиц по размерам (рис. А1б, в). Было найдено хорошее соответствие между двумя независимыми подходами прямого поиска функции распределения частиц по размерам в произвольной форме. На основе полученных результатов были выбраны стартовые приближения для программы POLYMIX, в которой функция распределения задается в аналитическом виде, которые также позволили приблизить данные и восстановить распределения по размерам (рис. А1г). На рис. А1г видно, что основные пики распределений частиц SM и TM-50 хорошо соответствуют ожидаемым размерам 6 нм для SM и 14 нм для TM-50 (в пределах ошибки совпадают с результатами, представленными в разделе 4.1.1). В бинарных смесях оба пика остаются на тех же значениях, при этом соотношение максимумов меняется пропорционально весовым соотношениям компонент смеси. Для смеси с соотношением SM:TM 2:1 максимумы пиков примерно равны, для смеси с соотношением SM:TM 4:1 максимум, соответствующий частицам SM, становится в 2 раза выше максимума от частиц ТМ-50. Сравнительная таблица А1 показывает положения максимумов и полуширин распределений основных компонент, найденных с помощью разных алгоритмов. Как видно, результаты находятся в хорошем согласии друг с другом. Комбинированный подход с чередованием методов BFGS и SA в программе MIXTURE также применялся, при этом значения параметров распределений совпали со значениями, полученными программой POLYMIX, с точностью до 0.01%. Это вполне ожидаемо, так как POLYMIX является модификацией MIXTURE (см. последний абзац в заключении главы 2) и отличаются они только методом минимизации. В данном случае обе программы стартуют с изначально хороших значений, оцененных на первых двух шагах программами GNOM и VOLDIS (см. раздел 3.2), используя метод поиска распределений частиц по размерам с расширенной областью сходимости, что и приводит к совпадению результатов.

На рис. А1 видно, что распределения, полученные с помощью комбинированного подхода, содержат минимальное количество максимумов, т.е. структурная модель проще других решений при сравнимом уровне параметра качества (разности между экспериментальной и теоретической кривыми рассеяния). Это, при отсутствии другой априорной информации, также может служить критерием качества ответа.

Таблица А1. Средние значения радиусов (< R_i >) частиц кремнезоля SM и TM-50, их дисперсии σ_i и коэффициенты дисперсности (σ_i /< R_i >) для отдельных компонент и бинарных смесей SM и TM-50.

Метод/	фракция -1 (SM)		фракция -2 (ТМ-50)				
программа	< R ₁ >,	$\sigma_{ m l},$ HM	$\sigma_{\rm i}/<\!\!R_{\rm i}>$	<i><r< i="">₂>, нм</r<></i>	σ_2 , HM	$\sigma_{\rm i}/<\!\!R_{\rm i}>$	
	HM						
Образец SM							
GNOM	5.4	1.3	0.24				
VOLDIS	5.5	1.2	0.22				
POLYMIX	5.5	1.2	0.22				
		Об	разец ТМ-5	50			
GNOM				13.2	1.6	0.12	
VOLDIS				13.0	1.5	0.12	
POLYMIX				13.3	1.4	0.11	
		Смесь	SM/TM-50	(2:1)			
GNOM	5.4	1.4	0.26	13.3	1.7	0.13	
VOLDIS	5.5	1.3	0.24	12.9	1.6	0.12	
POLYMIX	5.5	1.3	0.24	12.9	1.6	0.12	
Смесь SM/TM-50 (4: 1)							
GNOM	5.4	1.4	0.26	13.4	1.8	0.13	
VOLDIS	5.5	1.3	0.24	13.1	1.7	0.13	
POLYMIX	5.5	1.3	0.24	13.2	1.7	0.13	



Рис А1. Экспериментальные данные МУРР (а) после процедуры сглаживания и учета коллимационной поправки от частиц кремнезоля TM-50 (синяя кривая) и TM-50 (красная кривая), а также их бинарных смесей в весовом соотношении SM:TM (2:1) (малиновая кривая) и 4:1 (зеленая кривая). Восстановленные функции распределения частиц программами GNOM (б), VOLDIS (в), POLYMIX и MIXTURE (BFGS+SA) (г), цветовая схема соответствия образцов сохраняется, как на рисунке с кривыми МУРР.
Приложение Б.

РАСПРЕДЕЛЕНИЯ РАЗМЕРОВ НАНОЧАСТИЦ ОКСИДА ЦИНКА ПО ДАННЫМ ПРОСВЕЧИВАЮЩЕЙ ЭЛЕКТРОННОЙ МИКРОСКОПИИ И МАЛОУГЛОВОГО РЕНТГЕНОВСКОГО РАССЕЯНИЯ

Для демонстрации эффективной работы алгоритмов восстановления функции распределения частиц по размерам по данным МУРР и согласия с результатами, полученными альтернативными методами, в частности методом просвечивающей электронной микроскопии (ПЭМ), в данном приложении рассмотрена система наночастиц оксида цинка ZnO, распределенных в полиэтиленовой матрице (предназначена как образец стандартов в МУРР). Способ приготовления образцов представлен в работе [112]. Измерения были выполнены на малоугловом дифрактометре АМУР-К [87]. Из данных малоуглового рассеяния образцом было вычтено рассеяние полиэтиленовой матрицей. Полученная экспериментальная кривая МУРР показана на рис. Б1. Для нахождения функции распределения частиц по размерам были использованы программы GNOM и POLYMIX, MIXTURE (BFGS+SA) с помощью которых удалось практически идеально приблизить данные МУРР, при этом первоначальная оценка программой GNOM позволила определить наличие 2 компонент и их средние размеры, после чего получены совпадающие друг с другом распределения частиц по размерам программами POLYMIX и MIXTURE (BFGS+SA) с точностью до 0.01%. Как видно на рис. Б1, найденные функции распределения частиц по размерам содержат два пика, один для частиц с радиусами 1-2 нм и второй для частиц с радиусом 3-5 нм. Второй пик распределения функции $D_{v}(r)$ хорошо согласуется с распределением частиц по данным ПЭМ. Отсутствие первого пика в данных микроскопии объясняется тем, что частицы малого размера из-за их слабого контраста плохо поддаются распознаванию на изображениях и поэтому исключаются из рассмотрения при анализе данных ПЭМ. Представленные результаты подтверждают практическую значимость

рассматриваемых и разработанных в диссертационной работе алгоритмов при исследовании реальных полидисперсных объектов.



Рис Б1. Экспериментальные данные МУРР (а) после процедуры сглаживания и учета коллимационной поправки от наночастиц ZnO (синяя кривая) и наилучшее приближение к данным, полученное программами POLYMIX и MIXTURE (BFGS+SA). Изображения ПЭМ для наночастиц ZnO (б). Восстановленные по данным МУРР функции распределения частиц (в) программами GNOM (красная кривая), POLYMIX и MIXTURE (BFGS+SA) (малиновая кривая), и распределение, рассчитанное по данным ПЭМ (синяя кривая).

СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ

МУР – Малоугловое рассеяние

МУРР – Малоугловое рентгеновское рассеяние

МУРН – малоугловое нейтронное рассеяние

BFGS (от англ. Broyden-Fletcher-Goldfarb-Shanno) – квазиньютоновский

градиентный метод Бройдена–Флетчера–Гольдфарба–Шанно

SA (от англ. Simulated Annealing) – метод моделирования отжига

NM (от англ. Nelder-Mead) – метод многогранника Нелдера-Мида

BFGS+SA – комбинация методов BFGS и SA

СПИСОК РАБОТ АВТОРА ПО ТЕМЕ ДИССЕРТАЦИИ

- A1. Kryukova A. E., Konarev, P. V., Volkov, V. V. Evaluation of solution stability for two-component polydisperse systems by small-angle scattering // J Phys Conf Ser. 2017. V. 941. P. 012069.
- А2. Крюкова А. Е., Конарев, П. В., Волков, В. В. Исследование устойчивости решений при анализе полидисперсных систем методом малоуглового рассеяния // Кристаллография. – 2018. – Т. 63, № 1. – С. 32-37.
- A3. Konarev P. V., Kryukova, A.E., Volkov, V.V., Svergun, D.I. . Modelling of multicomponent polydisperse systems using small-angle scattering data // J Phys Conf Ser. 2019. V. 1238, № 1. P. 012004.
- A4. Kryukova A. E., Konarev, P. V., Volkov, V. V., Asadchikov, V. E. Restoring silicasol structural parameters using gradient and simulation annealing optimization schemes from small-angle X-ray scattering data // Journal of Molecular Liquids. – 2019. – V. 283. – P. 221-224.
- А5. Крюкова А. Е., Конарев, П. В., Волков, В. В., Асадчиков, В. Е. Комбинирование методов минимизации в целях повышения эффективности восстановления распределений частиц по размерам для раствора кремнезоля по данным малоуглового рассеяния // Ядерная физика и инжиниринг. 2019. Т. 10, № 4. С. 393-398.
- А6. Волков В. В., Конарев П. В., Крюкова А. Е. Комбинированная схема восстановления функции распределения частиц по размерам с использованием данных малоуглового рассеяния // Письма в ЖЭТФ. – 2020. – Т. 112, № 9. – С. 632-636.
- A7. Volkov V. V., Kryukova A.E., Konarev P. V. Approaches for improving the quality of particle size distribution reconstructions from small-angle scattering data // J Phys Conf Ser. – 2020. – V. 1686. – P. 012059.
- А8. Крюкова А. Е., Конарев, П.В., Волков, В.В. Поиск эффективного алгоритма для восстановления решений при анализе данных малоуглового рассеяния от раствора кремнезоля // Кристаллография. – 2021. – Т. 66, № 5. – С. 730-737.

СПИСОК ЦИТИРУЕМОЙ ЛИТЕРАТУРЫ

 Small Angle Scattering of X-Rays. / Guinier A., Fournet G. – New York: Wiley, 1955.

2. Glatter O. A new method for the evaluation of small–angle scattering data // J. Appl. Cryst. – 1977. – V. 10, № 5. – P. 415-421.

3. Рентгеновское и нейтронное малоугловое рассеяние. / Свергун Д. И., Фейгин, Л.А. – Москва: "Наука" Главная редакция физико-математической литературы, 1986. – 278 с.

4. Svergun D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria // J. Appl. Cryst. – 1992. – V. 25. – P. 495-503.

5. Konarev P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., Svergun, D.I.
PRIMUS - a Windows-PC based system for small-angle scattering data analysis //
J. Appl. Cryst. – 2003. – V. 36. – P. 1277-1282.

6. Manalastas-Cantos K., Konarev, P. V., Hajizadeh, N. R., Kikhney, A.G., Petoukhov, M. V., Molodenskiy, D. S., Panjkovich, A., Mertens, H. D. T., Gruzinov, A., Borges, C., Jeffries, C. M., Svergun, D. I., Franke, D. ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis // J. Appl. Cryst. -2021. - V.54, No 1. -P.343-355.

7. Petoukhov M. V., Svergun, D. I. . Ambiguity assessment of small-angle scattering curves from monodisperse systems // Acta Cryst. D. – 2015. – V. 71. – P. 1051–1058.

8. Sitzungsberichte der K.B. Akademie der Wissenschaften. / Friedrich W., Knipping, P., Laue, M., 1912. – 302 p.

9. Bragg W. H., Bragg, W.L. The reflection of X-rays by crystals // Proc Roy Soc Lon A. – 1913. – V. 88, № 605. – P. 428.

10. Raman C. V., Krishnamurti, P. A New X-ray Effect // Nature. – 1929. – V. 124.
– P. 53-54.

11. Krishnamurti P. Studies in X-ray diffraction. I. The structure of amorphous carbon // Ind J Phys. -1930. - V. 5. - P. 473.

12. Warren B. E. X-Ray Diffraction Study of Carbon Black // J Chem Phys. – 1934.
– V. 2. – P. 551.

13. Guinier A. Basic discussion of X-ray scattering at small angles by small discrete particles // Ann. Phys. – 1939. – V. 12. – P. 161-237.

14. Chadwick J. Possible Existence of a Neutron // Nature. – 1932. – V. 129. – P.
312.

Mitchell D. P., Powers, P.N. Bragg reflection of slow neutrons // Phys Rev. –
 1936. – V. 50. – P. 486.

16. Halban H., Preiswerk, P. Preuve expérimentale de la diffraction des neutrons // C R Acad Sci. – 1936. – V. 203. – P. 73-75.

17. Shull C. G., Wollan, E.O. X-Ray, Electron, and Neutron Diffraction // Science.
- 1948. - V. 108. - P. 69-75.

 Schelten J., Mayer, A., Schmatz, W., Hossfeld, F. Neutron small-angle scattering of hemoglobin in aqueous solution // Z Physio Chem. – 1969. – V. 350. – P. 851-855.

19. Gabriel A., Dupont, Y. A Position Sensitive Proportional Detector for X-Ray Crystallography // Rev Sci Instrum. – 1972. – V. 43. – P. 1600.

20. Harrison S. C. Structure of tomato bushy stunt virus: I. The spherically averaged electron density // J Mol Biol. – 1969. – V. 42, № 3. – P. 457-464.

21. Stuhrmann H. B. Ein neus Verfahren zur Bestimmung der Oberflachenform und der inneren Struktur von gelosten globularen Proteinen aus Roentgenkleinwinkelmessungen // Z. Phys. Chem. – 1970. - V. 72. - P. 177-184.

22. Stuhrmann H. B. Interpretation of small-angle scattering functions of dilute solutions and gases. A representation of the structures related to a one-particle scattering function // Acta Cryst. Sect. A. – 1970. – V. 26, No 3. – P. 297-306.

23. Stuhrmann H. B. Die Bestimmung der Oberflächenform von gelöstem Myoglobin aus Röntgenkleinwinkelmessungen // Z Phys Chem. – 1970. – V. 72. – P. 185-198.

24. Stuhrmann H. B., Kirste, R.G. Elimination der intrapartikulären Untergrundstreuung bei der Röntgenkleinwinkelstreuung an kompakten Teilchen (Proteinen) // Z Phys Chem. – 1965. – V. 46. – P. 247-250.

25. Ibel K., Stuhrmann, H.B. Comparison of neutron and X–ray scattering of dilute myoglobin solutions // J. Mol. Biol. – 1975. – V. 93. – P. 255-265.

26. Svergun D. I., Stuhrmann, H.B. . New developments in direct shape determination from small–angle scattering. 1. Theory and model calculations // Acta Cryst. – 1991. – V. A47. – P. 736-744.

27. Svergun D. I. Restoring three–dimensional structure of biopolymers from solution scattering // J. Appl. Cryst. – 1997. – V. 30. – P. 792-797.

28. Chacón P., Morán F., Díaz, J.F., Pantos, E., Andreu, J.M. Low-resolution structures of proteins in solution retrieved from X-ray scattering with a genetic algorithm // Biophys. J. – 1998. – V. 74. – P. 2760-2775.

29. Svergun D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing // Biophys J. – 1999. – V. 76. – P. 2879-2886.

30. Bada M., Walther, D., Arcangioli, B., Doniach, S., Delarue, M. Solution structural studies and low-resolution model of the Schizosaccharomyces pombe sap1 protein // J Mol Biol. $-2000. - V. 300, N_{\odot} 3. - P. 563-574.$

31. Chacón P., Díaz, J.F., Morán, F., Andreu, J.M. Reconstruction of protein form with X-ray solution scattering and a genetic algorithm // J Mol Biol. -2000. - V.299, No 5. - P. 1289-1302.

32. Walther D., Cohen, F.E., Doniach, S. Reconstruction of low-resolution threedimensional density maps from one-dimensional small-angle X-ray solution scattering data for biomolecules // J. Appl. Cryst. -2000. - V. 33. - P. 350-363.

33. Franke D., Svergun, D.I. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering // J. Appl. Cryst. – 2009. – V. 42. – P. 342-346.

34. Hura G. L., Menon, A.L., Hammel, M., Rambo, R.P., Poole II, F.L., Tsutakawa, S.E., Jenney, F.E. Jr., Classen, S., Frankel, K.A., Hopkins, R.C., Yang, S.-J., Scott,

J.W., Dillard, B.D., Adams, M.W.W., Tainer, J.A. Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS) // Nat Methods. – 2009. – V. 6. – P. 606–612.

35. Martel A., Liu, P., Weiss, T.M., Niebuhr, M., Tsuruta, H. An integrated high-throughput data acquisition system for biological solution X-ray scattering studies // J Synchrotron Radiat. – 2012. – V. 19. – P. 431-434.

36. Nielsen S. S., Moller, M., Gillilan, R.E. High-throughput biological small-angle X-ray scattering with a robotically loaded capillary cell // J. Appl. Cryst. – 2012. – V. 45. – P. 213-223.

37. Round A., Felisaz, F, Gobbo, A, Huet, J, Fodinger, L, Villard, C, Pernot, P, Blanchet, C, Roessle, M, Svergun, D, Cipriani, F. BioSAXS Sample Changer: a robotic sample changer for rapid and reliable high-throughput X-ray solution scattering experiments // Acta Cryst D. -2015. -V. 71. -P. 67-75.

38. Alber F., Forster, F., Korkin, D., Topf, M., Sali, A. Integrating diverse data for structure determination of macromolecular assemblies // Annu Rev Biochem. – 2008. – V. 77. – P. 443-477.

39. Russel D., Lasker, K., Webb, B., Vela'zquez-Muriel, J., Tjioe, E., Schneidman-Duhovny, D., Sali, A. Putting the pieces together: integrative structure determination of macromolecular assemblies // PLoS Biol. – 2012. – V. 10, N_{2} 1. – P. e1001244.

40. Goertz V., Dingenouts, N., Nirschl, H. . Comparison of Nanometric Particle Size Distributions as Determined by SAXS, TEM and Analytical Ultracentrifuge // Part. Part. Syst. Caract. . – 2009. – V. 26. – P. 17-24.

41. Pedersen J. S. Determination of size distribution from Small Angle Scattering data for systems with effective hard-sphere interaction // J. Appl. Cryst. -1994. – V. 27. – P. 595-608.

42. Rami M.-L., Meireles, M., Cabane, B., Guizard, C. Colloidal stability for concentrated Zirconia aqueous suspensions // J. Am. Ceram. Soc. – 2009. – V. 92. – P. S50–S56.

43. Alina G., Butler, P., Cho, J., Doucet, M., Gervaise, A., Kienzle, P. SasView for small-angle scattering analysis: [Электронный ресурс] // URL: http://www.sasview.org/.

44. Neutron, X-rays and Light. Scattering Methods Applied to Soft Condensed Matter. / Pedersen J. S.; eds. Lindner P., Zemb Th. – Amsterdam: North Holland, 2002. – 552 p.

45. Bressler I., Kohlbrecher, J., Thunemann, A. F. SASfit: a tool for small-angle scattering data analysis using a library of analytical expressions // J. Appl. Cryst. – 2015. – V. 48. – P. 1587-1598.

46. Bressler I., Kohlbrecher, J., Thunemann, A. F. McSAS: Software for the retrieval of model parameter distributions from scattering patterns // J. Appl. Cryst. – 2015.
– V. 48. – P. 962-969.

47. Glatter O. Interpretation of Real-Space Information from Small-Angle Scattering Experiments // J. Appl. Cryst. – 1979. – V. 12, № Apr. – P. 166-175.

48. Schulz G. V. Ueber die Beziehung zwischen Reaktiongeschwindigkeit und Zusammensetzung des Reaktionproduktes Macropolymerisationsvorgaemgen // Z. Phys. Chem. Abt. B. – 1935. – V. 30. – P. 379.

49. Botet R., Cabane, B. Simple inversion formula for the small-angle X-ray scattering intensity from polydisperse systems of spheres // J. Appl. Cryst. – 2012. – V. 45. - P. 406-416.

50. Panjkovich A., Svergun, D.I. CHROMIXS: automatic and interactive analysis of chromatography-coupled small angle X-ray scattering data // Bioinformatics. – 2017.10.1093/bioinformatics/btx846. – P. 1944-1946.

51. Svergun D. I., Petoukhov, M.V., Koch, M.H.J. Determination of domain structure of proteins from X-ray solution scattering // Biophys. J. – 2001. – V. 80. – P. 2946-2953.

52. Konarev P. V., Petoukhov, M.V., Svergun, D.I. MASSHA - a graphic system for rigid body modelling of macromolecular complexes against solution scattering data // J. Appl. Cryst. – 2001. – V. 34. – P. 527-532. 53. Petoukhov M. V., Svergun, D.I. Global rigid body modelling of macromolecular complexes against small-angle scattering data. // Biophys. J. – 2005. – V. 89. – P. 1237-1250.

54. Petoukhov M. V., Franke, D., Shkumatov, A.V., Tria, G., Kikhney, A.G., Gajda, M., Gorba, C., Mertens, H.D.T., Konarev, P.V. and Svergun, D.I. . New developments in the ATSAS program package for small-angle scattering data analysis // J. Appl. Cryst. -2012. -V. 45, No 2. -P. 342-350.

55. Bernado P., Mylonas, E., Petoukhov, M.V., Blackledge, M., Svergun, D.I. Structural Characterization of Flexible Proteins Using Small-Angle X-ray Scattering // J. Am. Chem. Soc. – 2007. – V. 129, № 17. – P. 5656-5664.

56. Tria G., Mertens, H. D. T., Kachala, M., Svergun, D. I. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering // IUCrJ. – 2015. – V. 2. – P. 207-217.

57. Panjkovich A., Svergun, D.I. . Deciphering conformational transitions of proteins by small angle X-ray scattering and normal mode analysis. // Phys. Chem. Chem. Phys. -2016. -V. 18. -P. 5707-5719.

58. Konarev P. V., Svergun, D.I. . Direct Shape Determination of Intermediates in Evolving Macromolecular Solu-tions from Small-Angle Scattering Data // IUCr Journal. – 2018. – V. 5. – P. 402-409.

59. Svergun D. I., Barberato, C., Koch, M.H.J. CRYSOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates.
// J. Appl. Cryst. – 1995. – V. 28. – P. 768-773.

60. Svergun D. I., Richard, S., Koch, M.H.J., Sayers, Z., Kuprin, S., Zaccai, G. Protein hydration in solution: experimental observation by X-ray and neutron scattering // Proc. Natl. Acad. Sci. USA. – 1998. – V. 95. – P. 768-773.

61. Volkov V. V., Svergun, D.I. Uniqueness of ab-initio shape determination in small-angle scattering // J. Appl. Cryst. – 2003. – V. 36. – P. 860-864.

62. Franke D., Petoukhov, M. V., Konarev, P. V., Panjkovich, A., Tuukkanen, A., Mertens, H. D. T., Kikhney, A. G., Hajizadeh, N. R., Franklin, J. M., Jeffries, C. M., Svergun, D. I. ATSAS 2.8: a comprehensive data analysis suite for small-angle

scattering from macromolecular solutions // J. Appl. Cryst. – 2017. – V. 50. – P. 1212-1225.

63. P. V. Konarev M. V. P., L. A. Dadinova, N. V. Fedorova, P. E. Volyn-sky, D. I. Svergun, O. V. Batishchev, E. V. Shtykova. BILMIX: a new approach to restore the size polydispersity and electron density profiles of lipid bilayers from liposomes using small-angle X-ray scattering data // Journal of Applied Crystallography. – 2020. – V. 53. – P. 236-243.

64. Konarev P. V., Gruzinov, A. Yu., Mertens, H. D. T., Svergun D. I. Restoring structural parameters of lipid mixtures from small-angle X-ray scattering data // J. Appl. Cryst. – 2021. – V. 54. – P. 169-179.

65. Konarev P. V., Graewert, M.A., Jeffries, C.M., Fukuda, M., Cheremnykh, T.A., Volkov, V.V., Svergun, D.I. EFAMIX, a tool to decompose inline chromatography SAXS data from partially overlapping components // Protein Science. – 2022. – V. 31. – P. 269-282.

66. Adinolfi S., Puglisi, R., Crack, J.C., Iannuzzi, C., Dal Piaz, F., Konarev, P.V., Svergun, D.I., Martin, S., Le Brun, N.E., Pastore, A. The Molecular Bases of the Dual Regulation of Bacterial Iron Sulfur Cluster Bio-genesis by CyaY and IscX // Front. Mol. Biosci. -2018. - V. 4. - P. 97-109.

67. Yan R., Yalinca, H., Paoletti, F., Gobbo, F., Marchetti, L., Kuzmanic, A., Lamba, D., Gervasio, F.L., Konarev, P.V., Cattaneo, A., Pastore, A. The Structure of the Pro-domain of Mouse proNGF in Contact with the NGF Domain // Structure. – 2019. – V. 27. – P. 78-89.

68. Sicorello A., Rozycki, B., Konarev, P. V., Svergun, D. I., Pastore, A. . Capturing the conformational ensemble of the mixed folded polyglutamine protein ataxin-3 // Structure. -2021. - V. 29. - P. 1-12.

69. Dadinova L. A., Chesnokov, Y.M., Kamyshinsky, R.A., Orlov, I.A., Petoukhov, M.V., Mozhaev, A.A., Soshinskaya, E.Y., Lazarev, V.N., Manuvera, V.A., Orekhov, A.S., Vasiliev, A.L., Shtykova, E.V. Protective Dps-DNA cocrystallization in stressed cells: an in vitro structural study by small-angle X-ray scattering and cryo-electron tomography // FEBS Lett. – 2019. – V. 593. – P. 1360-1371.

70. Kamyshinsky R., Chesnokov, Y., Dadinova, L., Mozhaev, A., Orlov, I., Petoukhov, M., Orekhov, A., Shtykova, E., Vasiliev, A. Polymorphic Protective Dps-DNA Co-Crystals by Cryo Electron Tomography and Small Angle X-Ray Scattering // Biomolecules. -2019. - V. 10. - P. 39.

71. Dadinova L., Kamyshinsky, R., Chesnokov, Y., Mozhaev, A., Matveev, V., Gruzinov, A., Vasiliev, A., Shtykova, E. Structural Rearrangement of Dps-DNA Complex Caused by Divalent Mg and Fe Cations // Int J Mol Sci. – 2021. – V. 22. – P. 6056.

72. King-Scott J., Konarev, P.V., Panjikar, S., Jordanova, R., Svergun, D.I., Tucker,
P.A. Structural characterization of the multidomain regulatory protein Rv1364c
from Mycobacterium tuberculosis // Structure. – 2011. – V. 19. – P. 56-69.

73. Castelmur E., Strümpfer, J., Franke, B., Bogomolovas, J., Barbieri, S., Qadota, H., Konarev, P.V., Svergun, D.I., Labeit, S., Benian, G.M., Schulten, K., Mayans, O. Identification of an N-terminal inhibitory extension as the primary mechanosensory regulator of twitchin kinase // Proc Natl Acad Sci. – 2012. – V. 109. – P. 13608-13613.

74. Nowak E., Potrzebowski, W., Konarev, P.V., Rausch, J.W., Bona, M.K., Svergun, D.I., Bujnicki, J.M., Le Grice, S.F., Nowotny, M. . Structural analysis of monomeric retroviral reverse transcriptase in complex with an RNA/DNA hybrid // Nucleic Acids Res. -2013. -V. 41. -P. 3874-3887.

75. Solution of ill Posed Problems. / Tikhonov A. N., Arsenin, V. Ya. – New York: Wiley, 1977. – 258 p.

76. The Levenberg-Marquardt Algorithm, Implementation and Theory: Springer-Verlag G. A. W. – Berlin.

77. Levenberg K. A Method for the Solution of Certain non-linear Problems in Last Squares // Quart. Appl. Math. – 1944. – V. 2. – P. 164-168.

78. Marquardt D. An Algorithm for Least-Squares Estimation of Nonlinear Parameters // SIAM Journal on Applied Mathematics. – 1963. – V. 11, № 2. – P. 431-441.

79. Pedersen J. S. Analysis of small-angle scattering data from colloids and polymer solutions: modeling and least-squares fitting // Adv. Colloid Interf. Sci. – 1997. – V.
70. – P. 171-210.

80. Svergun D. I., Konarev, P. V., Volkov, V. V., Koch, M. H. J., Sager, W. F. C., Smeets, J., Blokhuis, E. M. A small angle X-ray scattering study of the dropletcylinder transition in AOT microemulsions // J. Chem. Phys. – 2000. – V. 113. – P. 1651-1665.

81. Zhang X., Konarev P. V., Petoukhov M. V., Svergun D. I., Xing L., Cheng R. H., Haase I., Fischer M., Bacher A., Ladenstein R., Meining W. Multiple assembly States of lumazine synthase: a model relating catalytic function and molecular assembly // J Mol Biol. – 2006. – V. 362, No 4. – P. 753-70.

82. Malyutin A. G., Cheng, H., Sanchez-Felix, O. R., Carlson, K., Stein, B. D., Konarev, P. V., Svergun, D. I., Dragnea, B., Bronstein, L. M. Coat Protein-Dependent Behavior of Poly(ethylene glycol) Tails in Iron Oxide Core Virus-like Nanoparticles // Acs Applied Materials & Interfaces. – 2015. – V. 7, № 22. – P. 12089-12098.

83. Practical Optimisation. / Gill P. E., Murray, W., Wright, M.H.: London:Academic Press, 1981.

84. Titchmarsh E. C. An Inversion Formula involving Bessel Functions // Proc.
London Math. Soc. – 1924. – V. 23. – P. XXII–XXIV.

85. Fox C. A generalization of the Fourier Bessel integral transform // Proc. London Math. Soc. . – 1929. – V. 29. – P. 401-452.

86. Letcher J. H., Schmidt, P. W. . Small-Angle X-Ray Scattering Determination of Particle-Diameter Distributions in Polydisperse Suspensions of Spherical Particles // J. Appl. Phys. – 1966. – V. 37. – P. 649-655. 87. Mogilevsky L. Y., Dembo, A. T., Svergun, D. I., Feigin, L. A. Small-angle scattering diffractometer with a coordinate detector // Crystallography Reports. – 1984. – V. 29. – P. 587-591.

88. Blanchet C. E., Spilotros, A., Schwemmer, F., Graewert, M.A., Kikhney, A., Jeffries, C. M., Franke, D., Mark, D., Zengerle, R., Cipriani, F., Fiedler, S., Roessle, M., Svergun, D.I. Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY) // J. Appl. Cryst. – 2015. – V. 48. – P. 431-443.

89. The Mathematical Theory of Communication. / Shannon C. E. W., W. : Urbana: University of Illinois Press, 1949.

90. Волков В. В. Спектроскопия и малоугловое рассеяние в решении обратных задач исследования многокомпонентных систем: дис. докт. х. наук: 01.04.18. -ИК РАН, Москва, 2013 - 347 с.

91. Konarev P. V., Svergun, D.I. A posteriori determination of the useful data range for small-angle scattering experiments on dilute monodisperse systems // IUCr Journal. – 2015. – V. 2. – P. 352-360.

92. Kirkpatrick S., Gelatt, C. D., Jr., Vecci, M. P. Optimization by simulated annealing // Science. – 1983. – V. 220. – P. 671-680.

93. Nelder J. A., Mead, R. A simplex for function minimization // Computer Journal.
- 1965. - V. 7. - P. 308-313.

94. Metropolis N., Rosenbluth A. W., Rosenbluth M. N., Teller A. H., Teller E. Equations of State Calculations by fast Computing Machines // Journal Chemical Physics. -1953. - V. 21, No 6. -P. 1087-1091.

95. Boikova A. S., Dyakova, Yu.A., Ilina, K.B., Konarev, P.V., Kryukova, A.E., Kuklin, A.I., Marchenkova, M.A., Nabatov, B.V., Blagov, A.E., Pisarevsky, Y.V., Kovalchuk, M.V. Octamer formation in lysozyme solutions at the initial crystallization stage detected by small-angle neutron scattering // Acta Cryst Section D. - 2017. - V. D73. - P. 591-599.

96. Бойкова А. С., Дьякова, Ю.А., Ильина, К.Б., Конарев, П.В., Крюкова, А.Е., Марченкова, М.А., Благов, А.Е., Писаревский, Ю.В., Ковальчук, М.В.

Исследование влияния замены растворителя – H2O на D2O – на начальную стадию кристаллизации лизоцима тетрагональной сингонии методом малоуглового рентгеновского рассеяния // Кристаллография. – 2017. – Т. 62, № 6. – С. 876-881.

97. Дьякова Ю. А., Ильина К.Б., Конарев, П.В., Крюкова, А.Е., Марченкова, М.А., Благов, А.Е., Волков, В.В., Писаревский, Ю.В., Ковальчук, М.В. Исследование условий образования единиц роста белкового кристалла в растворах лизоцима методом малоуглового рассеяния рентгеновских лучей // Кристаллография. – 2017. – Т. 62, № 3. – С. 364-369.

98. Бойкова А. С., Дьякова, Ю.А., Ильина, К.Б., Конарев, П.В., Крюкова, А.Е., Марченкова, М.А., Писаревский, Ю.В., Ковальчук, М.В. Исследование предкристаллизационной стадии раствора (влияния температуры и типа осадителя) протеиназы К методом малоуглового рассеяния рентгеновского излучения // Кристаллография. – 2018. – Т. 63, № 6. – С. 857-862.

99. Kovalchuk M. V., Boikova, A. S., Dyakova, Yu. A., Ilina, K. B., Konarev, P. V., Kryukova, A. E., Marchenkova, M. A., Pisarevsky, Yu. V., Timofeev V. I. Precrystallization phase formation of thermolysin hexamers in solution close to crystallization conditions // Journal of Biomolecular Structure and Dynamics. – 2018. – P. 1538-0254.

100. Крюкова А. Е., Шпичка, А.И., Конарев, П.В., Волков, В.В., Тимашев, П.С., Асадчиков, В.Е. Восстановление формы бычьего фибриногена в растворе по данным малоуглового рассеяния // Кристаллография. – 2018. – Т. 63, № 6. – С. 863-865.

101. Shpichka A. I., Konarev, P. V., Efremov, Yu. M., Kryukova, A. E., Aksenova, N. A., Kotova, S. L., Frolova, A. A., Kosheleva, N. V., Zhigalina, O. M., Yusupov, V. I., Khmelenin, D. N., Koroleva, A., Volkov, V. V., Asadchikov, V. E., Timashev, P. S. Digging deeper: structural background of PEGylated fibrin gels in cell migration and lumenogenesis // RSC Advances. – 2020. – V. 10. – P. 4190–4200.
102. Konarev P. V., Grigorev, V. A., Bikmulina, P. Yu., Presnyakova, V. S., Kryukova, A. E., Volkov, V. V., Shpichka, A. I., Asadchikov, V. E., Timashev, P.

S. The Structural Features of Native Fibrin and Its Conjugates with Polyethylene Glycol and Vascular Endothelial Growth Factor according to Small-Angle X-Ray Scattering // Review Journal of Chemistry. – 2020. – V. 10, № 3-4. – P. 158–163.

103. Dumont F., Warlus, J., Watillon, A. Influence of the Point of Zero Charge of Titanium-Dioxide Hydrosols on the Ionic Adsorption Sequences // Journal of Colloid and Interface Science. -1990. - V. 138, No 2. -P. 543-554.

104. Asadchikov V. E., Volkov, V. V., Volkov, Y. O., Dembo, K. A., Kozhevnikov, I. V., Roshchin, B. S., Frolov, D. A., Tikhonov, A. M. Condensation of silica nanoparticles on a phospholipid membrane // Jetp Letters. – 2011. – V. 94, № 7. – P. 585-587.

105.LudoxSM:[Электронный ресурс]//URL:https://www.chempoint.com/products/grace/ludox-monodispersed-colloidal-
silica/ludox-colloidal-silica/ludox-sm.//URL:

106. Ludox TM-50: [Электронный pecypc] // URL: https://www.chempoint.com/products/grace/ludox-monodispersed-colloidal-silica/ludox-colloidal-silica/ludox-tm-50.

107. Химия кремнезема. / Айлер Р.: в 2-х т. – Москва: «Мир», 1982. – 1128 с. 108. Colloidal silica as a particle size and charge reference material: [Электронный pecypc] // URL: https://www.horiba.com/fileadmin/uploads/Scientific/Documents/PSA/TN158.pdf.

109. Крюкова А. Е., Козлова, А. С., Конарев, П. В., Волков, В. В., Асадчиков, В. Е. Оценка устойчивости решений при восстановлении функции распределения частиц в растворе кремнезоля по размерам по данным малоуглового рассеяния // Кристаллография. – 2018. – Т. 63, № 4. – С. 524-529. 110. Sheta B., Elhabiby, M., El-Sheimy, N. Comparison and Analysis of Nonlinear Least Squares Methods for Vision Based Navigation (Vbn) Algorithms // Xxii Isprs Congress, Technical Commission I. – 2012. – V. 39-B1. – P. 453-456.

111. Конарев П. В., Чуховский, Ф.Н., Волков, В.В. К решению обратной задачи дифракционной рентгеновской топо-томографии. Компьютерные алгоритмы

и 3d реконструкция на примере кристалла с точечным дефектом кулоновского типа // Кристаллография. – 2019. – Т. 64. – С. 173-183.

112. Запорожец М. А., Волков, В.В., Сульянов, С.Н., Рустамова, Е.Г., Губин, С.П., Митюхляев, В.Б., Кузин, А.Ю., Тодуа, П.А., Авилов, А.С. Стандартные образцы наночастиц Аu и ZnO для калибровки малоугловых рентгеновских дифрактометров // Измерительная техника. – 2013. – Т. 56, № 4. – С. 26-30.